University of South Bohemia in České Budějovice, Faculty of Science

GENOMIC DISTRIBUTION OF MAJOR FAMILIES OF SATELLITE DNA IN VICIA FABA AND LATHYRUS SATIVUS

BACHELOR THESIS

Böttinger Katharina

Biological Chemistry

Supervisor: RNDr. Jiri Macas, Ph.D.

Laboratory of Molecular Cytogenetics, Institute of Plant Molecular Biology, Biology Centre CAS

and

Faculty of Science, University of South Bohemia

České Budějovice, 2016

Böttinger, K., 2016: Genomic distribution of major families of satellite DNA in *Vicia faba* and *Lathyrus sativus* [Bc. Thesis, in English] – 29 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

The aim of this thesis was to investigate the genomic distribution of satellite DNA repeats in *Vicia faba* and *Lathyrus sativus* and the association of some satellite repeats with centromeric chromatin.

Sworn declaration:

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography.

I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages.

Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated y the National Registry of University Theses and a plagiarism detection system.

Date …….

Signature……..

Acknowledgements:

I would like to express my thanks to RNDr. Jiri Macas, Ph.D. for giving me the opportunity to work in his laboratory and the great working atmosphere I experienced. I also would like to thank Iva Fukova, Ph.D. for the extraordinary support in the thesis process, her patience, understanding, guidance and time. Last but not least I express many thanks to all lab colleagues and study colleagues for their assistance and help.

This project would have been impossible without the financial support from the Czech Science Foundation grant GBP501/12/G090.

Abbreviations:

aa	amino acid
APM	AminProphosMethyl
ChIP-seq	Chromatin ImmunoPrecipitation-sequencing
DAPI	4',6-DiAmidino-2-PhenylIndole
dATP	desoxyAdenosine TriPhosphate
dCTP	desoxyCytidine TriPhosphate
dGTP	desoxyGuanosine TriPhosphate
DNA	DeoxyriboNucleic Acid
dTTP	desoxyThymidineTriPhosphate
dUTP	dexoxyUridine TrisPhosphate
FISH	Fluorescence *in situ* Hybridization
IGS	InterGenic Spacer
NGS	Next Generation Sequencing
NT	Nick Translation
ON	Over Night
PCR	Polymerase Chain Reaction
RT	Room Temperature
satDNA	satellite DeoxyriboNucleic Acid
TE	Transposable Element

CONTENT:

1. INTRODUCTION

1.1 SATELLITE DNA

Satellite DNA (satDNA) is a class of repetitive DNA characterized by highly abundant, tandemly arranged repeated units (monomers) being tens to thousands of nucleotides long and often forming continuous arrays spanning up to 100 Mbp [1,2,3,4,5]. These arrays make up heterochromatin which is located in pericentromeric, intercalary or subtelomeric regions of plant chromosomes [6]. SatDNA has been identified in most higher plant genomes investigated so far [7]. However, the total amounts of satellite repeats vary greatly between plant genomes and thus contribute significantly to genome size variation. In the extreme case, individual satellite repeats can make up to 20% of nuclear DNA [8].

Rapid evolutionary changes of satDNA lead to the formation of families of satDNA, which differ in nucleotide sequence, sequence complexity, repeat unit length, location, and abundance, and are, therefore, species- or genus-specific [9]. SatDNA sequences may vary enormously between closely related species and between different chromosomes within one species, which allows us to deduce that satDNA arrays experience high rates of evolution through genome expansion and shrinkage [10]. This great sequence variability suggests that nucleotide sequence is of little or no importance for establishment or maintenance of satDNA in the genome. However, some short, A/T-rich sequence motifs have been found enriched in satellite repeats [2]. It was proposed that the A-T abundance and alternating A-T dispersion lead to a bending of the DNA into a super-helical tertiary structure [8]. Palindrome sequences, which are frequent parts of satDNA, build dyad, twofold structures which may be associated with satDNA amplification mechanism [11]. In addition, they could work as targets for enzyme recognition during recombination and gene-conversion processes [11] and may also be targets for transcription factors [8].

1.1.1 SATDNA AND PLANT CENTROMERES

The centromere is a key region in chromosome segregation during mitosis and meiosis. Centromeres are thought to be determined by both epigenetic and genomic factors [12,13]. They are epigenetically marked by the presence of centromere-specific histone H3 variant called CENH3 [9]. The underlying DNA sequences are often composed of the arrays of satDNA interrupted by transposable elements (TEs) [9]. The fact that satDNA families are a frequent component of centromeres suggests their importance for chromosome structure or function. Centromeric repetitive DNA sequences differ from one another in their primary sequence, size of repeating units, and abundance in the genome [14], thus it could be their tandem structure itself that is important. It has also been proposed that satDNA transcripts operate as epigenetic signals needed for organization of pericentromeric heterochromatin during embryogenesis and epigenetic regulation of heterochromatin establishment [6,13].

On the other hand, newly established centromeres (neocentromeres) often lack satellite repeats and are mostly based on single-copy sequences. The evolutionary progress from neocentromeres to mature centromeres may be accompanied by the stabilizing of CENH3 chromatin by accumulation of satDNA repeats. This can be achieved by tandem duplication of a single-copy sequence or through transposition of existing satDNA repeats into a new position [6,9,15,16]. Contrary to canonical H3 which is conserved in all eukaryotes, CENH3 shows significant variability between species [17]. In grass species the total size of CENH3 correlates with genome size and chromosomes of different sizes in the same species tend to have CENH3 domains of a similar size, although they have differing contents of satellite repeats [15]. However, Neumann et al. (2015) found that there is no correlation of centromere, chromosome, or genome size and arrangement of CENH3 in *Pisum* and *Lathyrus* species [18].

1.2 LEGUME TRIBE *FABEAE* AS A MODEL FOR REPETITIVE DNA AND CENTROMERE FUNCTION

The tribe of *Fabeae*, which includes genera of *Lathyrus* L., *Lens* Mill., *Pisum* L., *Vavilovia* Fed., and *Vicia* L. is characterized by a large variation of genome sizes between the species which is mostly caused by differential accumulation of repetitive DNA. SatDNA makes up substantial part of these repeats and it was found to vary substantially both in abundance and sequence composition [19]. The tribe has also been established as a new model for investigating centromere evolution, because it includes genera possessing monocentric chromosomes (*Vicia* and *Lens*) along with those containing chromosomes with extremely enlarged primary constrictions containing multiple CenH3 domains (*Lathyrus* and *Pisum*). While the former centromere type is prevailing in most plant species, the latter, also called meta-polycentric, is a newly discovered type of chromosome organization that may represent a transition stage to holocentric chromosomes [18,20].

1.2.1 *LATHYRUS SATIVUS* AND META-POLYCENTRIC CHROMOSOMES

L. sativus has a genome size of 6.52 Gbp/1C, where 45.46% are made up by Ogre elements, 6.85% consist of Maximus/SIRE repeats, and satDNA repeats span up 10.73%. With its characteristic primary constriction *L. sativus* is a representative of the species with a meta-polycentric chromosome. Contrary to most other plants, there are two different copies of the CENH3 protein, *CenH3-1* and *CenH3-2*, in *Pisum* and *Lathyrus* species. These proteins have only 55% identity and differ in sequence and length, 123 aa and 119 aa for *CenH3-1* and *CenH3-2*, respectively [18]. CenH3-1 and CenH3-2 target all 14 *L. sativus* centromeres and co-localize in interphase and mitotic chromosomes [14], although high resolution microscopy revealed that this co-localization is only partial [18,20].

1.2.2 *VICIA FABA* AND PLANT CYTOGENETICS

V. faba serves as a classical model for molecular plant cytogenetics. The first cytogenetic studies of *V. faba* were done in the mid-20th century and it is, therefore, one of the cytogenetically best characterized plants. *V. faba* is the representative of the largest genome for tribe *Fabeae* with a DNA content of 13.41 Gbp/1 C, which is shared between five acrocentric chromosome pairs and one metacentric chromosome pair giving the field bean its characteristic

karyotype [19,21,22]. The five acrocentric chromosomes cannot be differentiated easily. *V. faba* shows characteristic patterns of chromatin differentiation along the chromosomes and a first detailed cytogenetic mapping of different locations of repetitive DNA sequences was first described by J. Fuchs et al. (1998) [23].

More than half (54.32%) of *V. faba*'s genome are constituted of Ogre elements, which are LTR–retrotransposons belonging to the Ty3/gypsy lineage. Ty1/copia elements are less abundant, where one representative – Maximus/SIRE – makes up 6.77% of the genome. *V. faba* has a total amount of satDNA of 935 Mbp/1C (6.97% of genome) [19]. The previously identified satellite *Fok-I* occurs in high copy numbers in *V. faba* and is one of the most abundant plant repeats with 2.5 x 10^7 copies per haploid genome. Its distribution on chromosomes provides useful marker for FISH-based discrimination of acrocentric chromosomes [24,25]. A number of novel satellite repeats were identified in frame of a bioinformatics analysis of Next-Generation-Sequencing (NGS) data conducted by Macas et al. (2015) [19], which provided sequence data and predictions for the existence of satDNA repeats investigated in this study.

1.3 PREDICTION OF SATELLITE REPEATS

The introduction of next generation sequencing technologies into genetic research allows a detailed characterization of large and complex genomes, which has not been possible previously. One of the most efficient bioinformatics methods for characterization of satellite repeats from genomic data employs graph-based clustering of NGS reads [26,27] and this approach was also used to investigate repetitive DNA in 23 *Fabeae* species including *V. faba* and *L. sativus* [19]. This study revealed a number of putative novel satDNA families in both species. Moreover, ChIP-seq experiments using CENH3 antibodies performed in Macas' laboratory suggested that some of these novel repeats are potentially associated with centromeric chromatin (unpublished results). A method of choice for confirmation of these results and further investigation of corresponding satDNA repeats is their detection on mitotic chromosomes using *in situ* hybridization. Due to their arrangement in a limited number of long arrays, satellite repeats produce a typical FISH pattern, consisting from a small number of chromosome bands or bright spots. Moreover, position of these signals within primary constrictions of chromosomes indicates association of a repeat with centromeric chromatin, which can be further confirmed by performing combined FISH/CENH3-immunodetection experiments [14].

2. AIMS OF THE THESIS

The aims of this work were to verify the existence and investigate genomic distribution of selected families of satellite repeats in *Vicia faba* and *Lathyrus sativus* which were previously identified by bioinformatics analysis of genome sequencing data. Specific objectives included labeling corresponding sequences to prepare hybridization probes, optimizing squash preparations of root tip meristems and performing fluorescence *in situ* hybridization (FISH) on mitotic chromosomes. In addition to characterization of genomic distribution of the most abundant satDNA families, the work focused on satellite repeats that were predicted to be associated with centromeric chromatin based on CENH3 ChIP-seq (Chromatin ImmunoPrecipitation-sequencing) data. Due to the multi-domain structure of centromeres in *L. sativus*, a combination of FISH with CENH3 immunolabeling had to be adapted to achieve this task.

3. MATERIALS AND METHODS

3.1 PLANT SPECIES

All experiments were performed using the broad bean (*Vicia faba)* and grass pea (*Lathyrus sativus*). Seeds of *V. faba* were obtained from Osiva Boršov, Boršov nad Vltavou, Czech Republic and seeds of *L. sativus* were obtained from Fratelli Ingegnoli, Milano, Italy (catalogue no. 455) and both were stored at 4°C in darkness.

3.2 CELL CYCLE SYNCHRONIZATION AND CHROMOSOME ACCUMULATION

The procedure for root meristem synchronization and accumulation of *V. faba* cells in metaphase was adapted from J. Dolezel et al. [28]: *V. faba* seeds were aerated in distilled water for 18 h at 25°C and then transferred to perlit. The substrate was moistened with 1x Hoagland's solution and seeds were incubated for 72 h at 25°C in darkness. Then, the seedlings were transferred into an aerated container with 1x Hoagland's solution and 2.5 mM hydroxyurea and incubated for 18 h at 25°C in darkness. The solution was decanted and replaced with fresh 1x Hoagland and incubated for another 8 h. 2.5 μM amiprophos-methyl (APM) was added to the *V. faba* roots, which were aerated for additional 4 h/25°C in darkness.

L. sativus seeds were frozen for two weeks at -20°C. Then, they were transferred to perlite and moistened with 1x Hoagland's solution and incubated for 72 h at 25°C in darkness. The seedlings were placed into an aerated container containing 1x Hoagland's solution and 1.18 mM hydroxyurea. After 18 h the solution was decanted and replaced by fresh 1x Hoagland's solution and incubated for 4 h/25°C in darkness. Then 15 μM oryzalin was added and the roots were incubated for 2 h. Before use the roots were placed on ice water in order to receive well separated chromosomes.

3.3 SLIDE PREPARATION

3.3.1 SQUASH PREPARATION OF ROOT TIPS

The fresh *V. faba* roots were harvested to get 1-3 cm long root tips from plantlets and fixed in a 3:1 ethanol:glacial acetic acid fixative at 4°C over night (ON). For the enzymatic

digestion the root tips were rehydrated in distilled water. Then, the meristems were cut off and washed with 1x Tris buffer for 15-30 min at room temperature (RT). The meristems were digested in an enzyme solution CP/KCl (2.5% cellulase Onozuka R-10, 2.5% pectolyase Y-23; 7.5 mM KCl pH 7.0, 7.5 mM EDTA pH 7.0) and incubated for 25-30 min at 28°C. The enzyme solution was discarded and root tips were washed with 75 mM KCl. For the slide preparation the root tips were macerated in 45% acetic acid, squashed with a cover slip, and placed in liquid nitrogen. The cover slip was removed and the slides were dehydrated in ethanol series and stored at -20°C.

3.3.2 Flame Dry Method

The roots were harvested at 1-3 cm length from plantlets, washed with cold ddH_2O, and immediately fixed in 3:1 methanol:glacial acetic acid solution. Roots were left cooled at 4°C for two days and then enzymatically digested. The roots were washed with distilled water, the meristems were cut off, and then placed in the enzyme solution CPP (4% cellulase Onozuka R-10, 2% pectinase, 0.4% pectolyase Y-23 in 10 mM citrate buffer, pH 4.5). Incubation for 1 h at 37°C followed. Then, the digested root tips were washed and stored in fixative solution (3:1 methanol:glacial acetic acid) for at least one day.

3.3.3 Formaldehyde Squash Preparation of Root Tips

L. sativus root tips were cut at 1-2 cm length and put into ice cold miliQ water, transferred into 3% formaldehyde in 1x Tris buffer and incubated at RT for 25 min, then washed in 1x Tris buffer at RT for 25 min, and finally put into cold 1x PBS (120 mM NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO_4, 2.7 mM KCl). The roots were put in a drop of 1x PBS and the meristems were cut under a binocular microscope. The meristems were enzymatically digested in an enzymatic solution (2% pectolyase Y-23, 2% cellulase Onozuka R-10 in 1x PBS) at 28°C for 90 min until the material was soft, and washed in ice cold 1x PBS for at least 15 min. The digested material was transferred on a slide, LB01 solution (15 mM Tris-HCl, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100, 15 mM mercaptoenthanol, pH 7.5) [29] was added, and the meristems were chopped. Then, the material was squashed and the slide was put into liquid nitrogen to remove the cover slip. The slides were put immediately in 1x PBS and were stored in 1x PBS/0.1% Tween 20 at 4 °C until the use (ON).

Formaldehyde squash preparation was only used for immunodetection in combination with FISH.

3.4 FLUORESCENCE *IN SITU* HYBRIDIZATION

3.4.1 SATELLITE PROBES

The probes were previously designed according to the satellite sequences obtained by a low-coverage next-generation-sequencing of the *V. faba* and *L. sativus* genome (performed in Macas' laboratory; unpublished data). According to some of the satellite sequences oligonucleotide probes were designed. 5'-end-labeled probes were synthesized by Integrated DNA Technologies (Leuven, Belgium). The labels used were either biotin, DIG, or Rhodamine-Red-X (RR-X) (see Tab. 1. and Tab. 2). The remaining probes were amplified from cloned fragments and labelled either by nick translation (NT) or by PCR (see Tab. 3, Tab. 4, and chapter 3.4.2).

The following satellite repeats were localized with FISH in *V. faba* and *L. sativus* chromosomes. The tables below give an overview of the most important information about the repeats mentioning their ID, probe length [nt], %GC content, and sequence.

Tab. 1 – *V. faba* oligonucleotide probes

ID	Labeling	Length [nt]	% GC	sequence [5'->3']
VFBm2H1	Biotin	50	30	CTTTTAACTTAGTTTCTTAGTGATGAAGTAACCTAAAATTTGGTGATGGA
VFBm3_Fok_H1	Fluorescein	50	40	CTACCTTCCATAATGACAAGGCTACCATCCATTGGAGTAACAAAAATCTC
VFBm15H1	Biotin	48	33	CAACAACAACAACAACAACGTCAAATAAACAACAACAACAACAACAAC
VFBm127H1	Biotin	50	40	ATCAAAGAAAGGTTTAACACGRACGAGTGTTTGAATCAATACGGACGAGT

Tab. 2 – *L. sativus* oligonucleotide probes

ID	Labeling	Length [nt]	% GC	sequence [5'->3']
LASm1H1	biotin	40	50	TTCGGGTTCGATGCCTGTTTGTTTAACAGATGCCTAGACGG
LASm2H1	DIG	42	50	CGCTCAACGTAATAACCGGAGTCTGGATACYCAACGAAACTG
LASm3H1	RR-X	50	28	GTKAAAAAACTCACCAATTTCACTATAAARACCATWACAAAAGTTCAAAG

Tab. 3 – *V. faba* cloned FISH probes

Name	ID	Insert (monomer) length [bp]	Labeling	PCR primers	seqence [5'->3']
VFBm102c102	c1637	702 (irregular)	PCR	F	GCAGAAAATCTGATGAAAAATGATG
				R	TTGTTCACTTCAAATTTCGTCAG
VFBm144c57	c1650	1015 (44)	NT	F	TACCATAATGAATGGACCTTTATACT
				R	CGTTACATATTTTGACTAAGTACTTTTAATATG
VFBm164c16	c1655	2048 (irregular)	NT	F	CTAATCATGTTATGTCTCATGTAGTTTC
				R	GAAATGTTAATATTTGTTAATCAAAGACT
VFBm186C4	c1656	1763 (1762)	NT	F	AGACATAGATTTAGGTTCAAATTCGT
				R	TAGACTTTAGCATATCCAATAGATGAA
VFBm197C23	c1661	1719 (47)	NT	F	GGTTATAAAACAACAAGCAAAGTAAG
				R	CCTTGCATGTTTCCCTTTAT
VFBm200C23	c1666, c1664	2011, 147 (80)	NT	F	ATCAAATTAGTTGGGGCTTG
				R	TTCGGCAATCGTAATCAAC
VFBm200C1	c1667	309 (irregular)	PCR	F	GTTGTTTTGGTTGGTTCCAT
				R	AGCAAAGTGGGCTAGTCTTCT

Tab. 4 - *L. sativus* cloned FISH probes

Name	ID	insert (monomer) length [bp]	Labeling	PCR primers	sequence [5'->3']
LASm3c355	c1641	221 (50)	PCR	F	TGGTTTTTATGGTGAAATTAGTGAG
				R	CACCATAAAAACCATTACAAAAGTTG
LASm7c476	c1644	729 (370), 646	NT	F	GTTTCTTCGTCAGTAAGCCACAG
				R	TGGTGATGGAGAAGAAACATATTG

3.4.2 PREPARATION OF CLONED FISH PROBES

3.4.2.1 PLASMID DNA ISOLATION

First, bacteria (*Escherichia coli*) containing the plasmid *pCR 4-TOPO* (*InvitrogenTMlife technologies*, Carlsbad, CA) bearing the satellite DNA fragment were cultivated on LB agar plates with 50 µg/mL carbenicillin at 37°C ON, then transferred to liquid LB medium containing the same antibiotic, and incubated at 37°C ON.

Following the manufacturers instructions, *Wizard Plus SV Minipreps DNA Purification System* kit was used to extract plasmid DNA from bacteria.

Gel extraction and PCR product purification was done according to the protocol of *Wizard SV gel and PCR clean-up System* with the modification of elution with 20 µL nuclease free water instead of 50 µL.

After the purification the DNA concentration was measured using *NanoDrop* Spectrophotometer at 260 nm.

3.4.2.2 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION

Each 20 µL reaction contained 1x Colorless GO Taq Flexi buffer (*Promega*, Madison, WI), 1.5 mM $MgCl_2$, 0.2 mM each nucleotide, 0.5 µM each primer (M13F 5'-TGTAAAACGACGGCCAGT-3' and M13R 5'-AGGAAACAGCTATGACCATG-3'), 0.5 unit of GO Taq Flexi Polymerase (*Promega*, Madison, WI), and 1 ng of plasmid DNA. The amplification regime consisted of an initial denaturation step at 94°C /2 min, followed by 29 cycles of 94°C /20 sec, 51°C /20 sec, 72°C/2 min and a final extension step at 72°C for 5 min.

Gel extraction and PCR product purification was done according to the protocol of *Wizard SV gel and PCR clean-up System* with the modification of elution with 20 µL nuclease free water instead of 50 µL.

3.4.2.3 LABELING BY PCR

Each 20 µL PCR labeling reaction contained 1x Colorless GO Taq Flexi buffer, 1.5 mM $MgCl_2$, 35 µM biotin labeled dUTP, 0.1 mM each dATP, dCTP, and dGTP, 65 µM dTTP, 0.5 µM each primer, M13F and M13R, 1 ng of purified PCR product and 0.5 unit GO Taq Flexi polymerase. The program contained 30 cycles and the annealing temperature was 55°C.

Again, gel extraction and PCR product purification was done according to the protocol of *Wizard SV gel and PCR clean-up System* with the modification of elution with 20 µL nuclease free water instead of 50 µL.

3.4.2.4 NICK TRANSLATION

Nick translation was done with *ROCHE* Nick Translation Kit (Roche Diagnostics GmbH, Mannheim, Germany). A 20 µL mix contained 10 µL dNTP mixture (33 µM biotin-16-dUTP, 67 µM dTTP and 50 µM each dATP, dCTP, and dGTP), 1x buffer, 2 µL enzyme mixture (DNA-

polymerase I and DNase in 50% glycerol), and 1.2 µg purified PCR product and was incubated at 15°C for 30 min.

3.4.2.5 GEL ELECTROPHORESIS

All products of PCR and labeling reactions (usually 1/10 of reaction was analysed) were examined by gel electrophoresis in 1% agarose gel in 1x TAE with ethidium bromide. Constant voltage was applied at 5 V/cm until the bromphenol blue migrated 1-2 cm from the edge of the gel. Separated DNA fragments were visualized under UV-light.

3.4.3 FLUORESCENCE IN SITU HYBRIDIZATION (FISH) PROCEDURE

First, the slides were washed twice in 2x SSC (20x Saline Sodium Citrate: 3 M NaCl, 0.3 M sodium citrate, pH 7.0) at RT for 5 min, post-fixed in 4% formaldehyde/2x SSC at RT for 10 min, washed in 2x SSC again, and denatured in PCR buffer (1xPCR buffer [Promega, Medison, WI; M190A], 4 mM $MgCl_2$; 100 µL/slide) at 94 °C for 3 min. Next, the slides were dehydrated in an ethanol series (70% and 96%, 5 min each). The hybridization mixture (50% formamide, 10% dextran sulphate, 0.125 µg/µL calf thymus DNA, 0.125% SDS, 20 ng labeled DNA probe or 3-30 nM oligoprobe in 27 µL volume) was denatured at 76°C for 15 min. After denaturation 20x SSC was added to reach 2x final concentration. Then, the hybridization followed, where the denatured hybridization mixture was applied on the slide and incubated at 25°C ON. The slides were washed in 2x SSC at 35°C for 5 min, in 50% formamide/2x SSC at 35°C for 10 min, in 2x SSC twice for 5min first at 35°C and then at RT, and finally in 1x BT at RT for 10 min. Slides hybridized with directly labeled probes were mounted in 0.5 µg/µL DAPI in Vectashield mounting medium (Vector Laboratories Bulingame, CA). The preparations hybridized with probes labeled with biotin or DIG were blocked with 3% BSA/1x BT at RT for 10 min. All probes for *V. faba* were detected using Streptavidin Alexa 568 (*Molecular probes*; cat. no. S11226, red). Streptavidin Alexa 488 (*Jackson* Immunoresearch; cat. No. 016-540-084, green) was used to detect the probes *LAS_m1_H1*, *LAS_m3_c355* and *LAS_m7_c476* and for detection of *LAS_m2_H1* Rhodamine Red-X-IgG (Jackson *Immunoresearch*; cat. no. 200-292-156; anti-DIG mouse monoclonal, red) was applied. Streptavidin Alexa 568 and 488 as well as Rhodamine Red-X-IgG were diluted in 3% BSA/1x BT to a total volume of 50 µL, applied and incubated at 37°C for 1h. Finally, the slides were mounted in 0.5 µg/mL DAPI in Vectashield.

3.5 IMMUNODETECTION IN COMBINATION WITH FISH

The following primary antibodies were designed by P. Neumann and custom made in *GenScript* (Piscataway, NJ, USA) and diluted in 1x PBS/0.1% Tween20 were applied on formaldehyde squash preparations of *L. sativus* root tips and incubated at 4°C ON:

- anti- CenH3-2, *Pisum sativum*, produced in chicken, diluted 1:500
- anti- CenH3-1, *L. sativus*, produced in rabbit, diluted 1:1000
- anti- CenH3-2, *L. sativus*, produced in rabbit, diluted 1:1000

Next, the slides were washed twice in 1x PBS at RT for 5 min, rinsed in 1x PBS/0.1% Tween20, and the secondary antibodies (goat anti-chicken IgG antibody, DyLight 488, *Jackson Immunoresearch*, cat. no. 103-485-155 and goat anti-rabbit IgG antibody, DyLight 488, *Jackson Immunoresearch*, cat. no. 111-485-144, both diluted 1:500 in 1x PBS/0.1% Tween 20) were applied and incubated at RT for 1 h. Slides were postfixed in 4% formaldehyde in 1xPBS at RT for 10 min, washed twice in 1x PBS at RT for 5 min, and then FISH followed with the only modification that all washing steps prior to the slide denaturation were done in 1x PBS.

3.6 MICROSCOPY

All preparations were observed using a fluorescence microscope (Nikon Eclipse 600) which was equipped with a DS-Qi1Mc cooled camera. Signals were captured separately using UV-2A (blue), G-2A (red), and B-2A (green) filter sets, and processed using NIS Elements 3.0 software (Laboratory Imaging, Praha, Czech Republic). Afterwards, the pictures were adjusted and formatted using GIMP2.8.

4. RESULTS

4.1 LOCALIZATION OF SATDNA REPEATS FOR *V. FABA* WITH FISH

Ten satellite repeats were selected for FISH experiments: four of them were chosen because of their high abundance in the *V. faba* genome, while the remaining six represented potential centromeric satellites based on ChIP-seq experiments (Tab. 1 and Tab. 3).

The repeat *VFB_m3_Fok*, short *Fok*, is situated on the long arms of all acrocentric chromosomes and was employed as a reference in most experiments, since it shows different positions and patterns on each chromosome [2,3,23,30,31].

All novel satellite repeats except for one (*VFB_m197_C23*) produced FISH signals characteristic for satellite repeats. However, they differed in number, intensity and chromosome distribution of the signals. These signals are summarized in Fig. 1 and Supplementary Tab. S 1 (see Attachment) and examples of FISH images are given in Supplementary Fig. S 1 and Supplementary Fig. S 2 (see Attachment).

The repeat *VFB_m102* was located on the short arm of chromosome 2, 3, and 5. Additionally, it appeared on chromosome 1 and 6 (Supplementary Fig. S 2f). There are two repeats which were positioned on all six chromosomes. *VFB_m2* occurs on every chromosome as minor dot-like signals. On chromosome 1, however, it appeared as a large, broad band in the interstitial region (Supplementary Fig. S 1b). The second repeat existing on every chromosome was *VFB_m15*. On chromosome 4 and 6 it was located in between the signals of *Fok* (Supplementary Fig. S 1a).

Centromeric signals were revealed for all repeats which were found to be associated with CENH3 chromatin by ChIP-seq. However, there was no repeat occupying centromeres of all chromosomes. *VFB_m144* was a centromeric satellite of chromosome 1 (Supplementary Fig. S 2b), as well as *VFB_m186* (Supplementary Fig. S 2a). *VFB_m127* was situated in the centromere of chromosome 1 (Supplementary Fig. S 2e) and gave a weak signal on the lower part of the long arm of chromosome 6 (*or* 4) (not visible). The position on either of the two chromosomes could not be established as only six complements were scored and these provided insufficient data to

distinguish the chromosomes. The centromeric repeat *VFB_m200* was investigated using two probes, *VFB_m200_C23* and *VFB_m200_C1* which both produced signals on chromosome 6. It was not clear whether these two hybridized at the exact same positions as the probes were hybridized separately (Supplementary Fig. S 2c, d).

VFB_m164 was situated in centromere of either chromosome 3 *or* 4. Again it was not possible to assign the repeat to a chromosome due to an insufficient number of three captured pictures showing a signal (not shown). Hybridization of the probe *VFB_m197_C23* gave no clear results at all (not shown). It might be that the repeat lies on one of the acrocentric chromosomes, however, the signal might have also arisen from background, since it only occurred a few times (not significant) on different slides. Consequently, neither of them is considered in the karyotype.

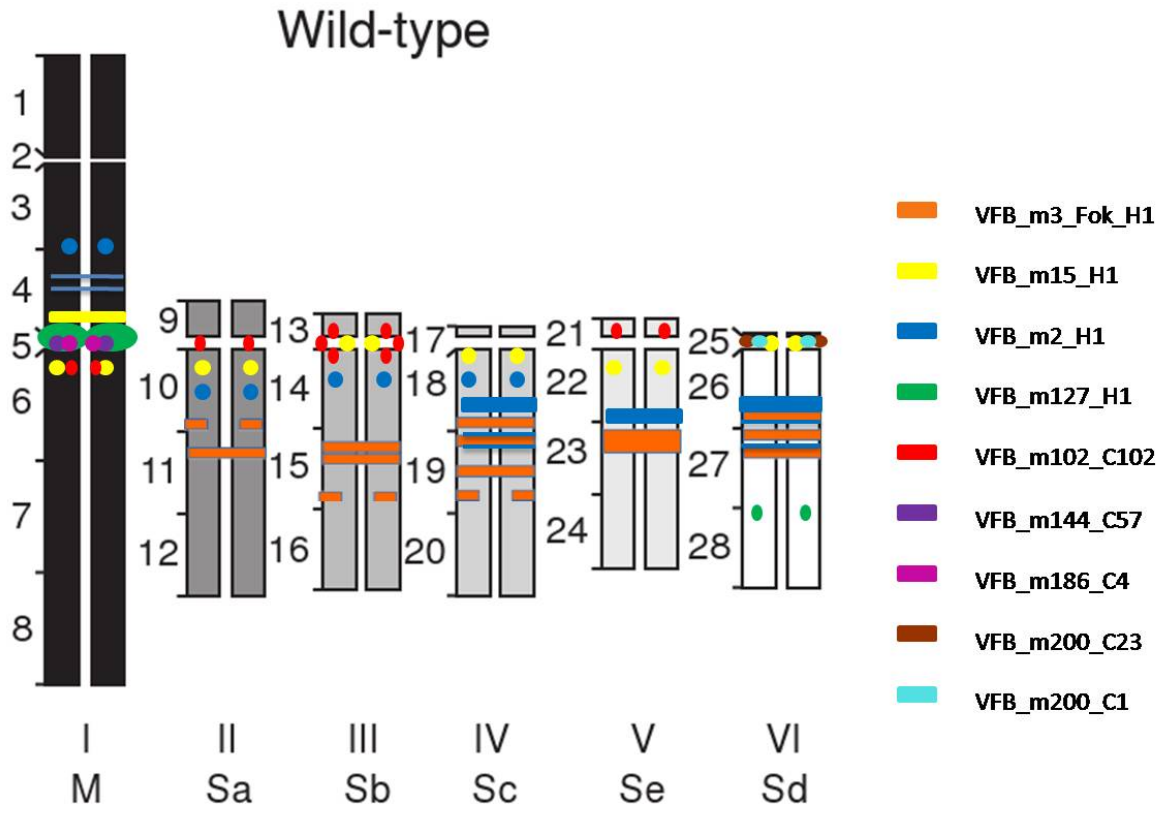


Fig. 1 – Schematic representation of FISH patterns of satellite repeat probes on *V. faba* chromosomes. The karyotype is based on Fuchs et al. (1998) [23], where the chromosomes are subdivided into sections numbered 1-28. The nomenclature was adopted by Michaelis, Rieger (1959) (roman numbers) [32] and Evans (1961) (letters: M, Sa-Sd) [33]. The legend on the right states all satellite repeats and marks them with a corresponding color.

4.2 LOCALIZATION OF SATDNA REPEATS IN *L. SATIVUS* WITH FISH AND IMMUNODETECTION

Four satellite repeats were investigated in *L. sativus*, all of them representing the most abundant families in this species, with one (*LAS_m3*) predicted to be associated with CenH3 chromatin (Supplementary Tab. S 2). The satellite *LAS_m1* was located on two chromosome pairs, once stretching along the whole primary constriction and once as a narrow band (Supplementary Fig. S 3b). *LAS_m2* was a satellite lying on four chromosome pairs, where it covered the whole primary constriction of one chromosome pair, just appeared as two dots in the middle of another chromosome pairs' primary constriction, and as two dots at the pericentromeric region of two more chromosome pairs (Supplementary Fig. S 3c, d). Plant telomeres often terminate in satDNA repeats [34]. *LAS_m7_C476* appeared as telomeric signals on the short arms on four chromosome pairs and was the only repeat localized outside primary constrictions. A high exposure time was necessary in order to make it clear and significant (Supplementary Fig. S 3f).

The *centromeric* repeat *LAS_m3* occurred on all seven chromosome pairs where it appeared along the edges of the primary constriction as evenly distributed dots. This pattern was obtained using oligo-probe *LAS_m3_H1* derived from the most conserved repeat region (see Sequence in Tab. 2). However, the cloned FISH probe C1641 with PCR primers *LAS_m3_C355* produced dots along the edges of primary constriction of one chromosome pair only. The signal appeared to be very weak and therefore an exposure time of twelve seconds was required to make it visible (Supplementary Fig. S 3a, e).

The dot-like patterns along the edges of primary constrictions of all chromosomes produced by *LAS_m3_H1* probe strongly resembled distribution of CENH3 domains described for *L. sativus* by Neumann et al. (2015, 2016) [20,18]. In order to prove that *LAS_m3* is indeed located in CENH3 domains, both CENH3 proteins (*CENH3-1* and *CENH3-2*) were visualized using specific antibodies along with the *LAS_m3* satellite detected using FISH. The antibodies raised against *CENH3-2* produced clear signals that fully co-localized with *LAS_m3*, proving that this satellite is present in all CENH3 chromatin domains (Supplementary Fig. S 4b, c). However,

the other CENH3 variant, *CENH3-1* gave only a weak signal in the immunodetection (Supplementary Fig. S 4a).

Characteristic marks for chromosomes are their size, length of the primary constriction and position of satellite repeats. It is so far not possible to distinguish the whole chromosome set for *L. sativus*. There are two chromosomes which can be distinguished according to their size and unique occurrence and position for *LAS_m1* and *LAS_m2* satellites, in both cases along the whole primary constriction. They are declared as chromosome 1 and 2 in Fig. 2. The rest of the chromosomes cannot be identified. The signals of the remaining probes are therefore symbolically depicted in one chromosome representing chromosome 3 to 7 (Fig. 2).

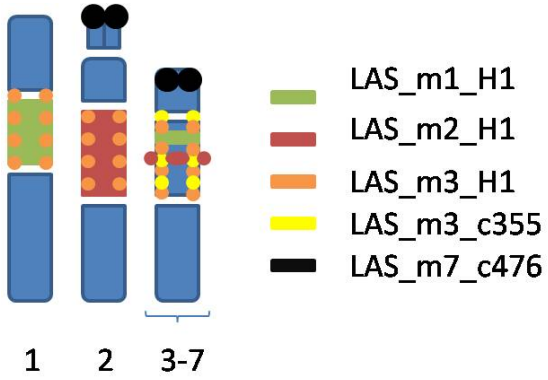


Fig. 2 – Karyotype of *L. sativus* based on FISH signals. The chromosomes are just schematically drawn. Further, only two of the seven chromosomes can be distinguished according to size, length of primary constriction, and mainly location of satellite repeats. These are named 1 and 2. The position of the other probes is summarized in one chromosome representing chromosome 3 to 7. The legend on the right states all satellite repeats and marks them with a corresponding color.

5. DISCUSSION

5.1 ADAPTION OF METHODOLOGY

The meristematic regions of root tips provided cells for metaphase chromosome preparations needed for FISH. However, these chromosomes were often extremely clustered after squashing and hence, were limited in their use for satellite mapping. Flame dry method for slide preparation gave cleaner preparations and better distinguishable and structured chromosomes for *V. faba* compared to squash slide preparations. Consequently, the flame dry method was used for all slide preparations of both *V. faba* and *L. sativus* chromosome as it proved to yield more useful preparations of slides.

5.2 EXISTENCE AND GENOMIC DISTRIBUTION OF SATDNA

V. faba has with 13.4 Gbp/1C the largest genome in tribe *Fabeae*, of which 935 Mbp/1C are satDNA repeats and in *L. sativus* more than ten percent (699 Mbp/1C) of its genome account for satDNA repeats [19]. FISH with two different fluorescent dyes allowed concurrent detection of two different repetitive sequences. This cytogenetic mapping of satDNA, which were previously identified with bioinformatics tools, helped to identify their global genomic distribution and to find their location on the chromosomes. SatDNA repeats occurred in patterns as band-like structures and at concrete positions as dots.

The existence of most of the cloned FISH probes was verified as they did hybridize well with *V. faba* and *L. sativus* chromosomes, however, the methodology for the hybridization of the probes *VFB_m164_C16* and *VFB_m127_C1* requires improvement as the signal appeared to be very weak and long exposure times were necessary to visualize them and hybridization only occurred on a low number of preparations. The existence of *VFB_m197_C23* could not be verified as hybridization did not succeed. This could be due to less efficient probe labeling combined with relatively low genomic abundance of this repeat. On the other hand, signals of highly abundant *Fok* repeat labeled with fluorescein (green) were so strong in some experiments that it passed through the G-2A (red) filter in the microscope, especially when exposure times were set to high values, which caused problems in detecting weakly hybridized probes. Adjustment of the probe concentration helped to eliminate this problem.

5.2.1 CENTROMERIC REPEATS

Distribution of centromeric satellites revealed unusual patterns in both species. In *V. faba*, the species with monocentric chromosomes which are in many plants associated with a single satellite repeat [9,35,36], there was no such satellite found. Instead, some centromeres were found to contain up to three different satellites, while no repeat was detected in several other chromosomes. The former was the case of chromosome 1, where the three satellites appeared to occupy distinct regions of the centromere region. However, as this observation was made in experiments using individual probes, it needs to be verified by simultaneous hybridization in multi-color FISH. Although not proven, it is likely that the probes *VFB_m200_C1* and *VFB_m200_C23* could be located on the same position in centromere of chromosome 6 as they originated from the same repeat.

The cases of two *V. faba* chromosomes (2 and 5) lacking signals of centromeric satellites can be explained in two ways. They could represent satellite DNA-free centromeres similar to those described from several *Solanum* species [37,38]. On the other hand, it is also likely that they contain less abundant centromeric repeats that were not tested in the experiments.

In *L. sativus*, only one of the investigated satDNA repeats (*LAS_m3*) was found to be associated with centromeric chromatin. Interestingly, the experiments using the cloned genomic fragment *LAS_m3_c355* as FISH probe resulted in signals on only one chromosome pair, while the oligo-probe *LAS_m3_H1* detected the repeat on all chromosomes. The most likely explanation is that the probe *LAS_m3_c355* represented a sequence variant (repeat subfamily) that was preferentially amplified on one chromosome, while the oligo-probe derived from repeat consensus sequence detected all repeat copies, thus labeling all chromosomes. The full co-localization of *LAS_m3_H1* signals with CENH3 immuno-detected using *CENH3-2* antibody suggests that *LAS_m3* is the major (or the only) centromeric satellite in this species. This is a striking difference to the other species with meta-polycentric chromosomes, *Pisum sativum*, which was found to have at least 13 different satellite repeats association with centromeric chromatin [14]. It should be noted that although the detection of *CENH3-1* protein failed, the results obtained from *CENH3-2* antibody were sufficient to visualize all centromeric domains because the two variants were found to be co-localized [20]. There are several possible reasons why the antibody *CENH3-1* (*L. sativus*, produced in rabbit) did not bind efficiently. First, the

antibody was generated in a polyclonal immune response, where it was raised against many overlapping antibody binding sites (epitopes). In this reaction only a short peptide (length of immunogen 23aa) was used. Therefore, the binding between the epitope and the antigen-combining site (paratope) only involved a small part of the molecule, which might result in a weak binding affinity and binding interaction [39,40]. Another explanation why it did not bind properly might be the quality of chromosome preparation. For example, the cytoplasm, which has not been removed completely in the squash preparation, could have prevented an efficient binding of the antibody and this effect could be more severe for *CENH3-1* antibody compared to *CENH3-2*.

6. SUMMARY AND CONCLUSION

Altogether ten satellite probes for *V. faba* and four satellite repeats for *L. sativus* have been analyzed by fluorescence *in situ* hybridization. Additionally, immunodetection in combination with FISH has been performed for *L. sativus* using CENH3 (*CENH3-1 and CENH3-2*) and the satellite repeat (*LAS_m3_H1*). The results can be summarized as follows:

1. In *V. faba*, two repeats (*VFB_m2* and *VFB_m15*) occurred on every chromosome pair and seven of the tested satellite probes were located on the centromere of *V. faba* chromosomes. The last probe, *VFB_m197_C23*, did not display any significant results. For *V. faba VFB_m3_Fok_H1* has aided as reference probe and has been usually used in combination with the other satellite probes.
2. Most of the tested satDNA repeats of *L. sativus* were located on and along the primary constriction, where one of which was present on every chromosome pair, whereas one of the analyzed probes was positioned on the telomeres.
3. The results for the immunodetection together with FISH revealed that *CENH3-2* did co-localize with the satellite repeat, whereas *CENH3-1* could not be efficiently detected.

In conclusion, the experiments proved that all detected sequences represent satellite repeats, as predicted from the bioinformatics analysis of genomic NGS data. Additionally, localization of putative centromeric satellites confirmed their association with centromeric chromatin and revealed interesting distribution patterns. These findings will be instrumental for further studies focusing on centromere evolution in plants.

7. REFERENCES

1. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371,** 215–220 (1994).

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

3. Macas, J., Požárková, D., Navrátilová, A., Nouzová, M. & Neumann, P. Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. *Mol. Gen. Genet.* **263,** 741–51 (2000).

4. Mehrotra, S. & Goyal, V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics Bioinforma.* **12,** 164–171 (2014).

5. Ingham, L. D., Hanna, W. W., Baier, J. W. & Hannah, L. C. Origin of the main class of repetitive DNA within selected Pennisetum species. *Mol Gen Genet* **238,** 350–356 (1993).

6. Garrido-Ramos, M. a. Satellite DNA in Plants: More than Just Rubbish. *Cytogenet. Genome Res.* (2015). doi:10.1159/000437008

7. Hemleben, V., Kovařík, A., Torres-Ruiz, R. A., Volkov, R. A. & Beridze, T. Plant highly repeated satellite DNA: Molecular evolution, distribution and use for identification of hybrids. *Syst. Biodivers.* **5,** 277–289 (2007).

8. Pezer, Brajkovic, J., Feliciello, I. & Ugarkovic, D. Satellite DNA-mediated effects on genome regulation. *Repetitive DNA* 153–169 (2012). doi:10.1159/000337116

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

10. Melters, D. P. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14,** R10 (2013).

11. Krawinkel, U., Zoebelein, G. & Bothwell, A. L. M. Palindromic sequences are associated with sites of DNA breakage during gene conversion. *Nucleic Acids Res.* **14,** 983–998

(1986).

12. Garrido-Ramos, M. A. Satellite DNA in Plants: More than Just Rubbish. *Cytogenet. Genome Res.* **146,** 153–170 (2015).

13. Graur, D., Zheng, Y. & Azevedo, R. B. R. An evolutionary classification of genomic function. *Genome Biol. Evol.* **7,** 642–645 (2015).

14. Neumann, P. *et al.* Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genet.* **8,** e1002777 (2012).

15. Zhang, T. *et al.* The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. *Proc. Natl. Acad. Sci.* **110,** E4875–E4883 (2013).

16. Gong, Z. *et al.* Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. *Plant Cell* **24,** 3559–3574 (2012).

17. Malik, H. S. & Henikoff, S. Major Evolutionary Transitions in Centromere Complexity. *Cell* **138,** 1067–1082 (2009).

18. Neumann, P. *et al.* Centromeres off the hook: massive changes in centromere size and structure following duplication of CenH3 gene in *Fabeae* species. *Mol. Biol. Evol.* **32,** 1862–1879 (2015).

19. Macas, J. *et al.* In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae. *PLoS One* **10,** e0143424 (2015).

20. Neumann, P. *et al.* Epigenetic Histone Marks of Extended Meta-Polycentric Centromeres of Lathyrus and Pisum Chromosomes. *Front. Plant Sci.* **7,** 1–11 (2016).

21. Bennett, M. D. & Smith, J. B. Nuclear dna amounts in angiosperms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **274,** 227–274 (1995).

22. Nemec, B. Über die Einwirkung des Chloralhydrats auf die Kern und Zellteilung. *Jahrb. für Wissenschaftliche Bot.* **39,** 645–730 (1904).

23. Fuchs, J., Strehl, S., Brandes, A., Schweizer, D. & Schubert, I. Molecular-cytogenetic characterization of the Vicia faba genome--heterochromatin differentiation, replication patterns and sequence localization. *Chromosom. Res.* **6,** 219–30 (1998).

24. Fuchs, J., Pich, U., Meister, a & Schubert, I. Differentiation of field bean heterochromatin byin situ hybridization with a repeatedFokI sequence. *Chromosom. Res.* **2,** 25–28 (1994).

25. Kato, A., Yakural, K. & Tanifuji, S. Sequence analysis of Vicia faba repeated DNA, the Fok1 repeat element. *Nucleic Acids Res.* **12,** 6415–6426 (1984).

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

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

28. Dolezel, J. Cell cycle synchronization in plant root meristems. *Methods Cell Sci.* **21,** 95–107 (1999).

29. Doležel, J., Binarová, P. & Lucretti, S. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* **31,** 113–120 (1989).

30. Navrátilová, A., Neumann, P. & Macas, J. Karyotype analysis of four *Vicia* species using in situ hybridization with repetitive sequences. *Ann. Bot.* **91,** 921–6 (2003).

31. Navrátilová, A., Neumann, P. & Macas, J. Long-range organization of plant satellite repeats investigated using strand-specific FISH. *Cytogenet. Genome Res.* **109,** 58–62 (2005).

32. Michaelis, A. & Rieger, R. Strukturheterozygotie bei Vicia faba. *Züchter* **29,** 354–361 (1959).

33. Evans, H. Chromatid aberrations induced by gamma irradiation. I. The structure and frequency of chromatid interchanges in diploid and tetraploid cells of Vicia faba. *Genetics*

46, 257–275 (1961).

34. Bass, H. & Birchler, J. A. in *Plant cytogenetics* (eds. Bass, H. & Birchler, J. A.) 350 (Springer, 2012).

35. Cheng, Z. *et al.* Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14,** 1691–1704 (2002).

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

37. Gong, Z. *et al.* Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* **24,** 3559–74 (2012).

38. Zhang, H. *et al.* Boom-Bust Turnovers of Megabase-Sized Centromeric DNA in Solanum Species: Rapid Evolution of DNA Sequences Associated with Centromeres. *Plant Cell* **26,** 1436–1447 (2014).

39. Frank, S. a. *Immunology and Evolution of Infectious Disease*. *Health San Francisco* **300,** (2002).

40. Reverberi, R. & Reverberi, L. Factors affecting the antigen-antibody reaction. *Blood Transfus.* **5,** 227–240 (2007).

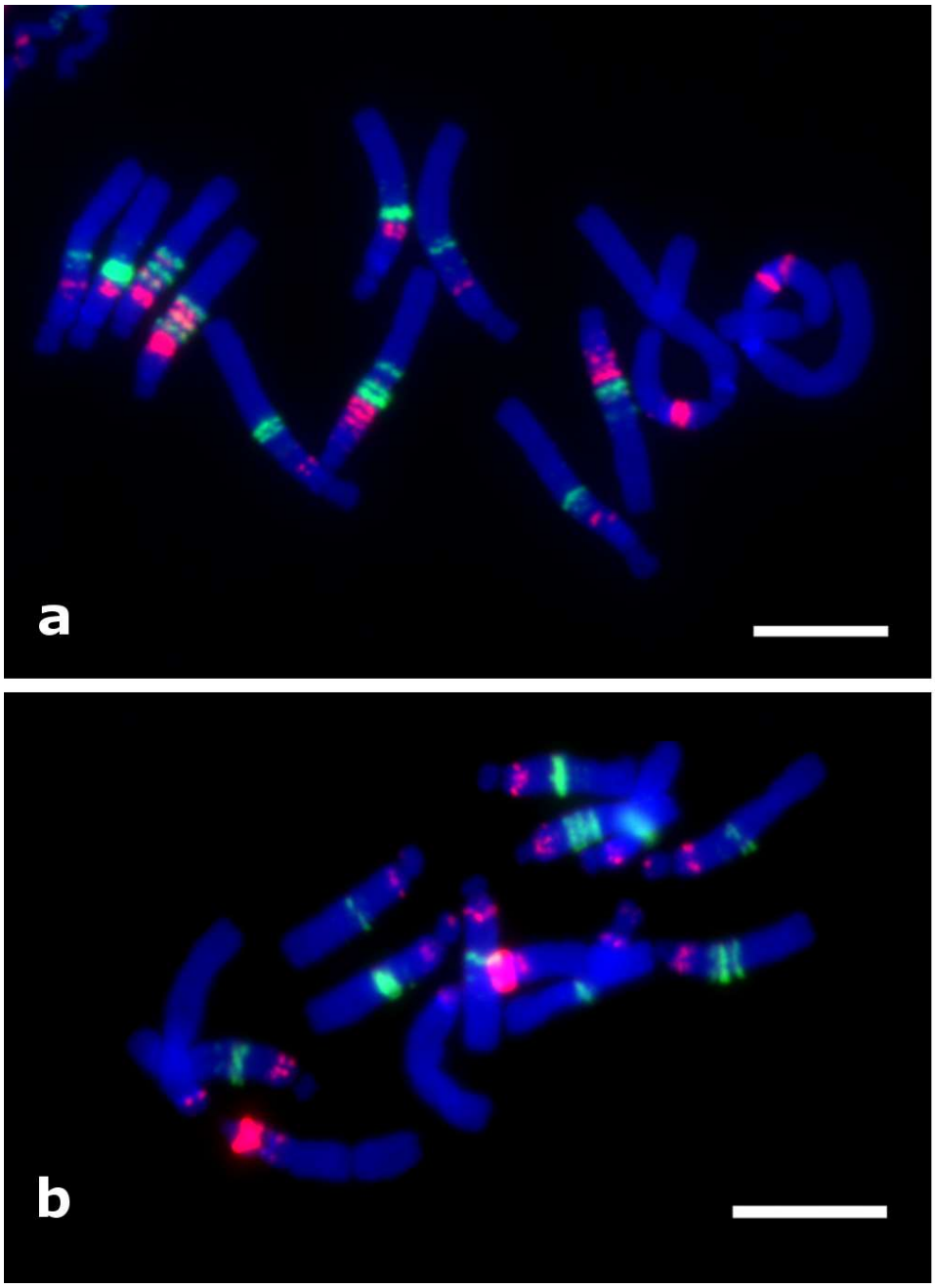
8. ATTACHMENT

Supplementary Tab. S 1 - summarized results for tested *V. faba* satellite repeats.

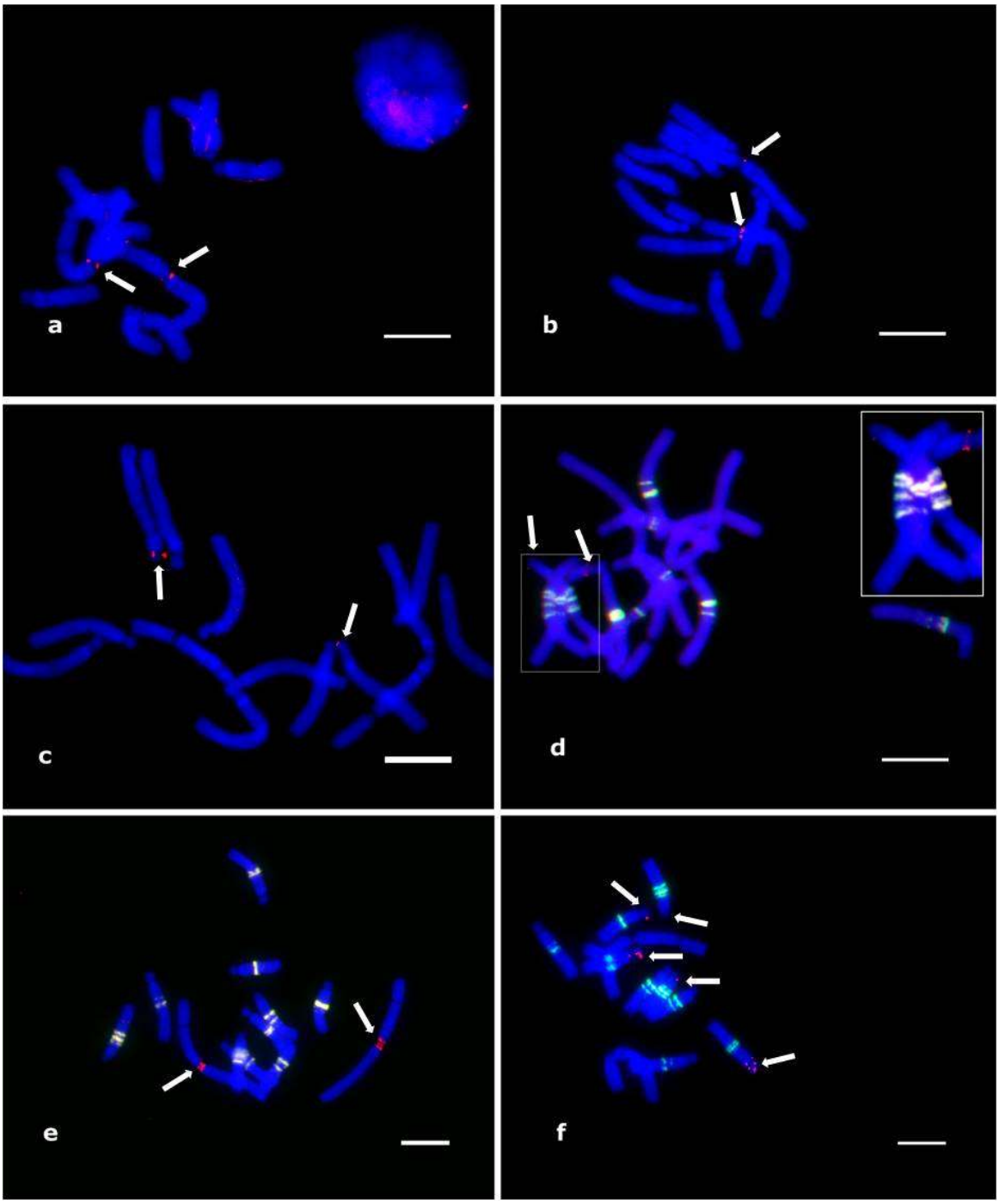
Characteristics of investigated satellite repeats and their localization by FISH									
Satellite	**CenH3/ChIP enriched**	**FISH**							**Notes**
		localization	**chromosome**						
			1	**2**	**3**	**4**	**5**	**6**	
VFB_m2_H1		proximal region	x	x	x	x	x	x	
VFBm3_Fok_H1		long arm		x	x	x	x	x	
VFB_m15_H1		long arm acrocentric on chromosome 2-6, proximal and interstitial region on chromosome 1	x	x	x	x	x	x	
VFB_m127_H1	x	centromere (1) and long arm (4 *or* 6)	x			x		x	4 OR 6
VFB_m102_C102		short arm (2, 3, 5)		x	x		x		
VFB_m144_C57	x	centromere	x						
VFB_m164_C16	x	centromere			x	x			3 OR 4
VFB_m186_C4	x	centromere	x						
VFB_m197_C23	x								no results
VFB_m200_C23	x	centromere						x	
VFB_m200_C1	x	centromere						x	

Supplementary Tab. S 2 - summarized results for tested *L. sativus* satellite repeats.

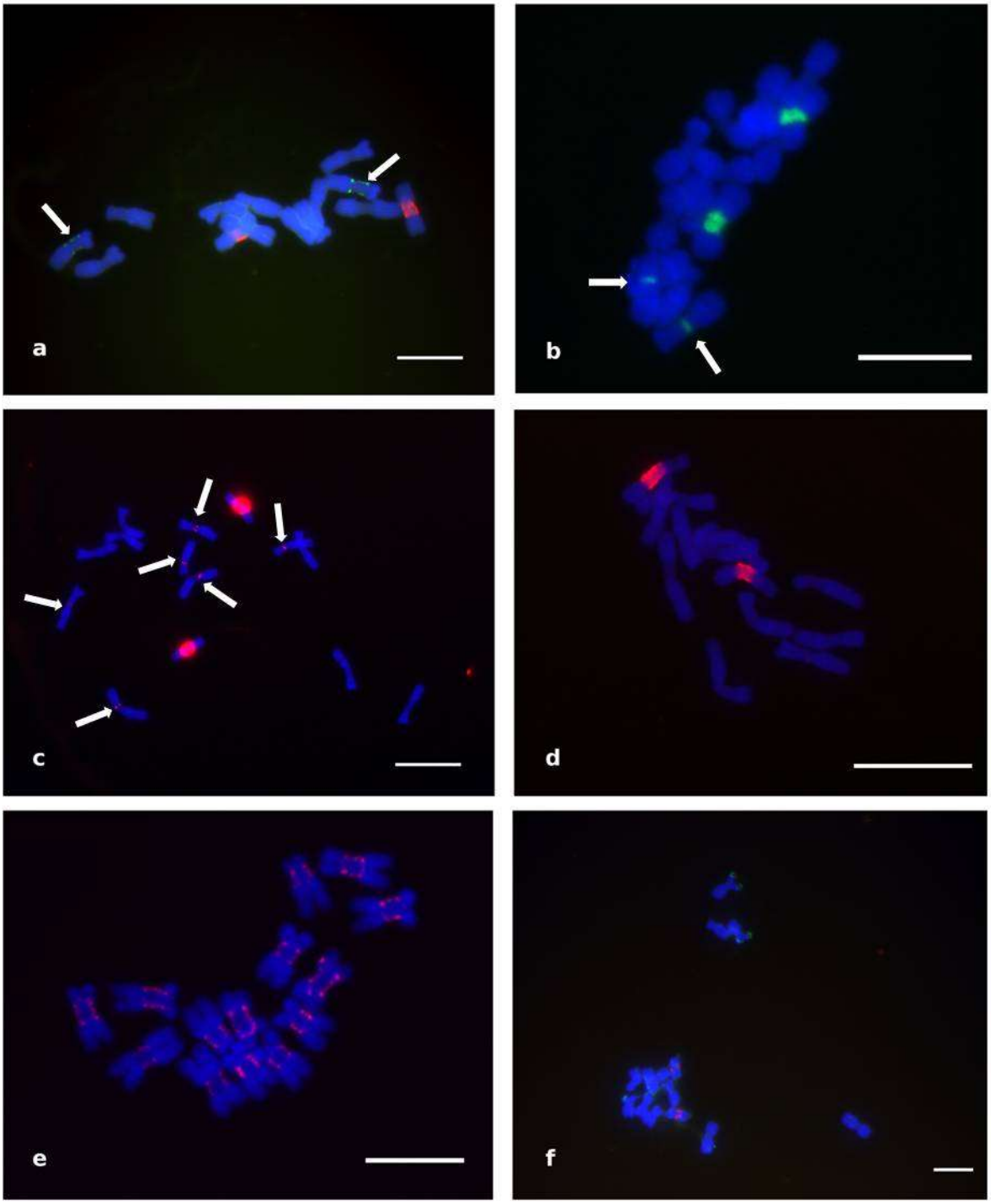
Characteristics of investigated satellite repeats and their localization by FISH			
Satellite	CenH3/ChIP enriched	FISH	
		localization	Notes
LAS_m1_H1		primary constriction	on two chromosome pairs
LAS_m2_H1		primary constriction	on four chromosome pairs
LAS_m3_H1	x	primary constriction	on every chromosome
LAS_m3_C355		primary constriction	on one chromosome pair
LAS_m7_C476		telomere	on four chromosome pairs



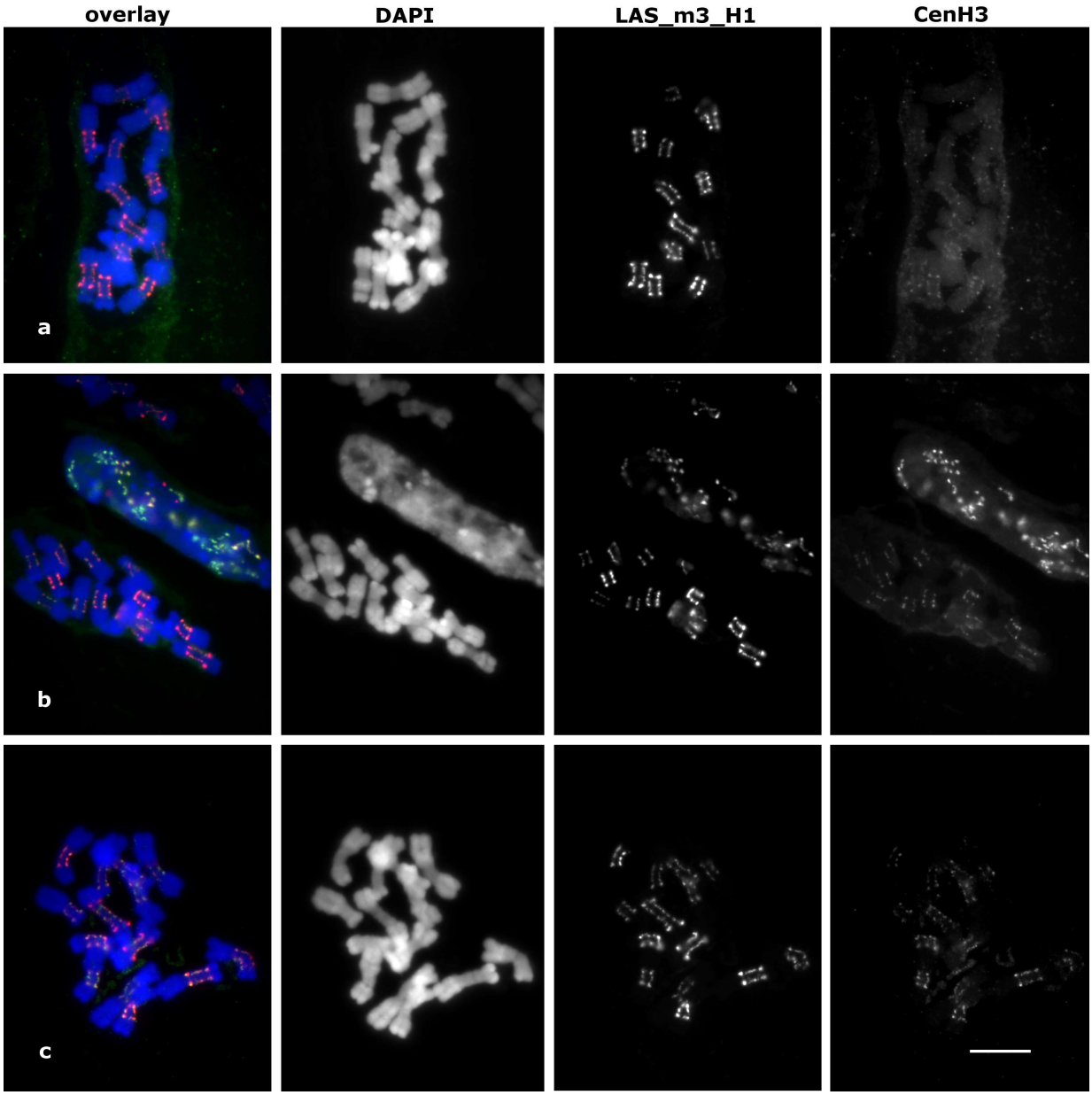
Supplementary Fig. S 1 - Location of *VFB_m2_H1* (red) (b) and *VFB_m15_H1* (red) (a) alongside *VFB_m3_Fok_H1* (green) in metaphase chromosomes of *V. faba*. Chromosomes are counterstained with DAPI (blue). Bar = 10 μm.



Supplementary Fig. S 2 - Detection of centromeric satellite repeats on metaphase chromosomes of *V. faba*. **(a)** *VFB_m186_C4* (red, chromosome 1), **(b)** *VFB_m144_C57* (red, chromosome 1), **(c)** *VFB_m200_C23* (red, chromosome 6), **(d)** *VFB_m200_C1* (red, chromosome 6, enlarged in upper right corner), **(e)** *VFB_m127_H1* (red, chromosome 1), and **(f)** *VFB_m102_C102* (red, chromosome 2, 3, and 5, not all signals visible due to overlay of chromosomes). In **(d) - (f)**: simultaneous detection with *VFB_m3_Fok_H1* (green). Arrows point to weakly visible signals. Chromosomes are counterstained with DAPI (blue). Bar =10 µm.



Supplementary Fig. S 3 - – Mapping of cloned FISH probes and oligonucleotide probes on *L. sativus'* metaphase chromosomes. (a) *LAS_m3_C355* (green) and *LAS_m2_H1* (red), **(b)** *LAS_m1_H1* (green), **(c)** *LAS_m2_H1* (red, long exposure time), **(d)** *LAS_m2_H1* (red, short exposure time), **(e)** *LAS_m3_H1* (red), and **(f)** *LAS_m7_C476* (green) and *LAS_m2_H1* (red). Arrows point to weakly visible signals. Chromosomes are counterstained with DAPI (blue). Bar =10 μm.



Supplementary Fig. S 4 – Simultaneous FISH/immunodetection of satellite *LAS_m3_H1* and CENH3-containing chromatin domains. *LAS_m3_H1* probe (red) was used in combination with three differenct CENH3 antibodies raised against **(a)** *CENH3-1* protein from *L. sativus* (produced in rabbit, diluted 1:1000), **(b)** *CENH3-2* from L. sativus (rabbit, diluted 1:1000), and **(c)** *CENH3-2* from *Pisum sativum* (chicken, diluted 1:500). The columns represent from left to right the merged images, DAPI, the satellite *LAS_m3_H1*, and the immunodetection. Bar =10 μm.