

Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*

Kateřina Havlová¹ · Martina Dvořáčková^{1,3} · Ramon Peiro⁴ · David Abia⁴ · Iva Mozgová² · Lenka Vansáčová² · Crisanto Gutierrez⁴ · Jiří Fajkus^{1,2}

Received: 29 June 2016 / Accepted: 3 August 2016 / Published online: 16 August 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Approximately seven hundred 45S rRNA genes (rDNA) in the *Arabidopsis thaliana* genome are organised in two 4 Mbp-long arrays of tandem repeats arranged in head-to-tail fashion separated by an intergenic spacer (IGS). These arrays make up 5% of the *A. thaliana* genome. IGS are rapidly evolving sequences and frequent rearrangements inside the rDNA loci have generated considerable interspecific and even intra-individual variability which allows to distinguish among otherwise highly conserved rRNA genes. The IGS has not been comprehensively described despite its potential importance in regulation of rDNA transcription and replication. Here we describe the detailed sequence variation in the complete IGS of *A. thaliana* WT plants and provide the reference/consensus IGS sequence, as well as genomic DNA analysis. We further

investigate mutants dysfunctional in chromatin assembly factor-1 (CAF-1) (*fas1* and *fas2* mutants), which are known to have a reduced number of rDNA copies, and plant lines with restored CAF-1 function (segregated from a *fas1xfas2* genetic background) showing major rDNA rearrangements. The systematic rDNA loss in CAF-1 mutants leads to the decreased variability of the IGS and to the occurrence of distinct IGS variants. We present for the first time a comprehensive and representative set of complete IGS sequences, obtained by conventional cloning and by Pacific Biosciences sequencing. Our data expands the knowledge of the *A. thaliana* IGS sequence arrangement and variability, which has not been available in full and in detail until now. This is also the first study combining IGS sequencing data with RFLP analysis of genomic DNA.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-016-0524-1) contains supplementary material, which is available to authorized users.

✉ Martina Dvořáčková
dvorackova.martina@gmail.com

✉ Jiří Fajkus
fajkus@sci.muni.cz

¹ Mendel Centre for Plant Genomics and Proteomics, CEITEC, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

² Faculty of Science, Laboratory of Functional Genomics and Proteomics, National Centre for Biomolecular Research, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic

³ Institute of Biophysics ASCR, v.v.i., Královopolská 135, 61265 Brno, Czech Republic

⁴ Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Nicolas Cabrera 1, Madrid 28049, Spain

Keywords *Arabidopsis thaliana* · Chromatin assembly factor · Nucleolus organizer region · 45S ribosomal DNA · Intergenic spacer · rDNA rearrangements

Introduction

Genome rearrangements play a key role in evolution of species leading to biodiversity, adaptation to particular environmental changes, or tolerance to abiotic and natural stresses (e.g. Geiser et al. 2016; Kelly et al. 2015; Long et al. 2013; Mandakova and Lysak 2008). It is interesting that genome reorganisation often involves 45S rRNA genes (rDNA) (Elliott et al. 2013; Garcia and Kovarik 2013; Long et al. 2013; Weider et al. 2005). Ribosomal genes fall into the group of well-conserved housekeeping genes characterised by repetitiveness, high abundance, and the presence of two distinct fractions, an actively transcribed and an inactive fraction, condensed and often spatially separated from the

active gene copies (Grummt and Pikaard 2003; Pontes et al. 2003; Pontvianne et al. 2013). Examples of ribosomal locus reorganisation can be found in many different organisms. There are plant species containing 45S (or 35S) rDNA associated with 5S rDNA in a single operon, unlike the majority of plants including *Arabidopsis thaliana* where these two units are separated (Garcia and Kovarik 2013; Garcia et al. 2010). Further, in *A. thaliana* from Northern Sweden the amount of rDNA is highly variable, thus having significant impact on the total genome size (Long et al. 2013). The ability of rDNA to expand or contract is well described in budding yeast (e.g. Kobayashi et al. 2004). In marine bacteria, short insertions in 16S rDNA result in structural changes during the adaptation to a high pressure environment (Lauro et al. 2007); in fish, rapid speciation is associated with retrotransposon-driven amplification of rDNA (Symonova et al. 2013); and the phenomenon termed a “jumping” NOR, originally described by (Schubert and Wobus 1985) results in a variable number of rDNA loci in *Allium*. One could conclude that such changes are associated with high sequence variability inside the rDNA genes, but in fact the regions affected the most are the short internal spacers (internal transcribed spacers ITS1 and ITS2) and the spacer between individual gene clusters (intergenic spacer, IGS). The rDNA ITS have become widely used for species identification and even the highly variable IGS can be utilized to distinguish among closely related species in phylogenetic studies (e.g. Cavallero et al. 2015; Han et al. 2016; Konstantinova and Yli-Mattila 2004; Lin et al. 2014; Marcilla et al. 2001).

We focus here on the IGS of rDNA in the plant *A. thaliana*. This model plant contains 570–750 copies of the rRNA gene organised at two chromosomal sites (Pruitt and Meyerowitz 1986) on chromosomes two and four (Copenhaver and Pikaard 1996a). Transcriptionally active rDNA copies are known as Nucleolus Organizer Regions (NORs). Long rDNA arrays consist of ca. 10 kb-long repetitive units of which 5.7 kb represent the coding region (18S–5.8S–25S) separated by a ~4.5 kb long IGS. So far, the *A. thaliana* IGS has been described mainly leaving the intraspecies sequence variability aside (Gruendler et al. 1989), with only fragments of IGS sequences available in public databases which allow to assemble only consensual IGS. The IGS typically contains tandem repetitions enriched with *SalI* restriction sites (*SalI* boxes), spacer and rDNA gene promoters, transcription terminators, and signals for pre-rRNA processing (Abou-Ellail et al. 2011; Gruendler et al. 1989). *SalI* boxes, in particular, are often discussed in connection with regulation of rDNA transcription, which is independent of the well-described epigenetic regulation (Durut et al. 2014; Earley et al. 2010; Pontvianne et al. 2010). *SalI* boxes function as enhancers or terminators of rDNA transcription in *Xenopus* or mammals (Grummt et al. 1986; Pikaard et al. 1990).

The *A. thaliana* IGS can be divided into three parts, the non-transcribed spacer (NTS) and two external transcribed spacers (3'ETS and 5'ETS) which are transcribed by polymerase I as a part of the nascent pre-rRNA. The sequence variability within the 3'ETS has been extensively studied (Abou-Ellail et al. 2011; Pontvianne et al. 2010). Four 3'ETS variants were described from which we can distinguish four rRNA gene subtypes. Three of the variants, var1-3, are abundant, while var4 is relatively rare. rRNA gene subtypes differ in transcriptional activity according to the plant's developmental stage; var2, var3 and var4 are expressed in roots, flowers, leaves and seedlings while var1 is expressed only in 2-day-old seedlings. Recently, the chromosomal position of these variants has been determined (Chandrasekhara et al. 2016). Var1 and a subset of var3 map to NOR2 while var2, var4 and the majority of var3 map to NOR4. Since all the inactive rRNA gene subtypes map to the same rDNA locus, Chandrasekhara et al. suggest that the gene subtypes are selectively silenced based on their chromosomal position rather than in a sequence-dependent manner. Our latest studies further show that this distribution of active and inactive rDNA variants to chromosomes 4 and 2, respectively, can be altered when the ribosomal genes are eliminated and subsequently recovered, as observed in *fasciata* (*fas1* or *fas2*) mutants and plant lines derived from a *fas* background with restored *FAS* function, respectively (Pavlišťová et al. 2016). *FASCIATA 1* and *2* are two of the subunits of the highly conserved trimeric histone H3/H4 chaperone Chromatin Assembly Factor-1, CAF-1 (Smith and Stillman 1989) whose deficiency results in severe phenotypic defects including stem fasciation, abnormal leaf and flower morphology, and disorganization of the apical meristem (Kaya et al. 2001; Leyser and Furner 1992; Reinholz 1966). The *fas* mutation leads to cell cycle-related defects such as increased homologous recombination (Endo et al. 2006; Kirik et al. 2006; Takeda et al. 2004), slower progression of the S-phase, constitutive activation of the G2 checkpoint and triggering the endocycle (Exner et al. 2006; Ramirez-Parra and Gutierrez 2007; Schonrock et al. 2006). In addition, *fas* mutants show progressive telomere shortening and loss of rDNA repeats while other repetitive genomic regions remain unaffected (Mozgova et al. 2010). Our previous results show that both originally active and originally inactive rDNA variants are lost in *fas* mutants (Pontvianne et al. 2013) and that this loss is mitotic and connected with homology-dependent repair (Muchova et al. 2015). In experiments leading to *FAS* gene reintroduction we observed that in contrast to WT plants both rDNA loci, two as well as four, can be activated (Pavlišťová et al. 2016).

As we previously reviewed (Dvorackova et al. 2015), rDNA represents a problematic template to replicate and a hot spot for recombination. The *fas* mutants, as well as plant

lines with restored *FAS* function (here termed revertants), are unique examples of plants where rDNA rearrangements have occurred. Further, the IGS sequence has not been studied extensively in *A. thaliana* plants and so far only fragmentary information has been available. This lack of knowledge of the complete IGS sequence and its naturally existing variants limits interpretation of studies addressing genetic and epigenetic factors involved in regulation of rDNA loci. Here we address the question whether the combination of 3'ETS variants and the juxtaposed IGS regions (mainly *Sall* boxes) is completely random or whether there are some rules for IGS arrangements. Furthermore, we ask how *fas* disruption reflects on IGS sequence and variability, and if plant lines with recovered rDNA arrays contain the same IGS set-up as the original plants. We present a comprehensive set of IGS sequences obtained by conventional cloning and Pacific Biosciences (PacBio) sequencing supported by genomic DNA analysis by RFLP. Finally, we characterise a new type of variability within the NTS region where the repetitive elements and spacer promoters are located, and report a reference IGS sequence summarising all the information obtained.

Material and methods

Plant material and DNA isolation

All *A. thaliana* plants were on a Columbia 0 background (Col 0). The T-DNA insertion mutants *fas1-4* and *fas2-4* (named *fas1* and *fas2* in the following text) are designated as NASC: N828822, SAIL_662_D10 and NASC: N533228, SALK_033228 (Exner et al. 2006). Wild-type (WT) *A. thaliana* plants had no mutant history or were segregated from the progeny of *fas1* heterozygotes and grown for another two or five generations (G2+/, G5+/-). The *fas1* mutants were segregated from heterozygous progeny and used in the first generation or grown for another two, five or seven generations (*fas1-4* G1-/-, G2-/-, G5-/-, G7-/-). The *fas2-4* mutants were segregated in the first generation (*fas2-4* G1-/-). WT *A. thaliana* revertant lines 1, 3, 4, and 6 were segregated from the progeny of *FAS1fas1/FAS2fas2* double heterozygotes (Pavlištová et al. 2016). These lines have a low (line 6), medium (line 3), or high (lines 1 and 4) rDNA content. Genomic DNA was isolated from leaves of 5 weeks-old plants using the protocol according to (Dellaporta et al. 1983) or using a NucleoSpin Plant II Midiprep Kit (Macherey–Nagel). Leaves used for a single sample came from 1 to 3 different individuals of the same genotype and generation (“siblings”). The DNA quality was checked by electrophoresis and the genotype was tested by PCR as described by (Mozgova et al. 2010).

Cloning, plasmid DNA isolation and sequencing

We amplified the IGS using a pair of primers designed in a conserved region of the 25S and 18S rRNA genes (25SFw and 18SR or 25SFw_seq and 18SR_seq, online resource 1). The PCR contained 1.25 U of ExTaq polymerase (TaKaRa), 0.8 mmol l⁻¹ of dNTP mix, 5 µl of 10× ExTaq Buffer, 50 ng of genomic DNA, 0.6 µmol l⁻¹ of primers and water up to 50 µl. The PCR conditions included incubation at 94 °C for 30 s, followed by 30 cycles of 98 °C for 10 s, 55 °C for 30 s, 72 °C for 4 min, with final incubation at 72 °C for 20 min.

The cloning reaction was prepared using 4 µl of fresh PCR product (approx. 100 ng), 1 µl of plasmid pCRTM Invitrogen (10 ng µl⁻¹) and 1 µl of salt solution (from the TOPO[®] TA Cloning[®] Kit, Invitrogen). The reaction was incubated for 30 min at room temperature then mixed with 20 µl of electrocompetent cells (*E. coli* ElectroMAXTM Stbl4TM, Invitrogen) or chemocompetent cells (*E. coli* One Shot[®] TOP10, Invitrogen). In the case of electrocompetent cells, the salt in the cloning reaction was diluted 5×, the bacteria were transformed by electroporation at 2.5 kV, mixed with 1 ml of SOC medium, incubated at 30 °C, 200 rpm for 90 min, then spread on Petri dishes with LB medium and incubated at 30 °C for 16 h. In the case of chemocompetent cells, the bacteria were transformed by heatshock at 42 °C in a water bath for 45 s, mixed with 1 ml of SOC medium, incubated at 37 °C, 200 rpm for 90 min, then spread on Petri dishes with LB medium containing ampicillin (50 µg ml⁻¹), IPTG (10⁻⁴ mol ml⁻¹) and X-gal (0.04 mg ml⁻¹). Plates were incubated at 37 °C for 16 h.

The blue–white test was used to select clones with a plasmid insertion and the length of the insertion was tested by PCR with M13F/M13R primers. The plasmid DNA was isolated using GenEluteTM HP Plasmid Miniprep Kits (Sigma Aldrich) or QIA[®] Spin Miniprep Kits (Qiagen). Sequencing of the clones was performed by Macrogen (South Korea) using the Sanger method with primers covering the whole IGS; for details see the list of primers (online resource 1). We obtained approx. 1.1 kb long reads which were assembled by their overlapping regions.

Single molecule real-time sequencing

Two series of single molecule real-time (SMRT) sequencing were conducted. In the first series, the samples were prepared by PCR with the use of Q5[®] High-Fidelity DNA polymerase (NEB) to amplify the shorter, approx. 3.5 kb long, fragment of the IGS. In the second series, Phusion[®] Hot Start II High-Fidelity polymerase (Thermo Scientific) was used to produce both fragments of the IGS, 3.5 and 4.5 kb long.

In the case of PCR with Phusion polymerase, 25–30 identical 10 µl PCRs were prepared and run in parallel to obtain 5 µg of the product required for SMRT. A single PCR contained 0.2 U of polymerase, 0.3 mmol l⁻¹ of dNTP mix,

2 μl of $5\times\text{GC}$ Reaction Buffer (Thermo Scientific), 20 ng of genomic DNA, $0.25\ \mu\text{mol l}^{-1}$ of primers Fx and Rx (see the list of primers for details), 2.5% DMSO, $1.25\ \text{mmol l}^{-1}$ of MgCl_2 and MilliQ water up to 10 μl . The products were mixed and purified using a QIA[®] PCR purification Kit (Qiagen). The process was repeated with ten different genomic DNAs and different pairs of primers FxRx where x is 0–9. The primers were designed in the conserved region of the 18S and 25S rRNA genes and were distinguished by a unique barcode (online resource 1). The annealing temperature had to be optimized for each pair of primers and was in the ranges 69–71 °C or 49–51 °C, respectively. PCR conditions included incubation at 98 °C for 30 s, followed by 2 cycles of 98 °C for 10 s, 49–51 °C for 30 s, 72 °C for 2 min 30 s, followed by 25 cycles of 98 °C for 10 s, 69–71 °C for 30 s, 72 °C for 2 min 30 s, with final incubation at 72 °C for 5 min. The quality and concentration of PCR products was measured using electrophoresis and a Nanodrop (Thermo Scientific). The PCR products were then mixed equimolarly and sequenced by GATC Biotech (Germany).

In the case of Q5 polymerase, 7–15 identical 25 μl PCRs were prepared and run in parallel. A single PCR contained 0.5 U of polymerase, $0.2\ \text{mmol l}^{-1}$ of dNTP mix, 2 μl of Reaction Buffer $5\times$ (NEB), 2 μl of $5\times$ Enhancer Buffer (NEB), 25 ng of genomic DNA, $0.25\ \mu\text{mol l}^{-1}$ of primers Fx and Rx, and MilliQ water up to 25 μl . The PCR conditions were incubation at 98 °C for 30 s, followed by 2 cycles of 98 °C for 10 s, 49 °C for 30 s, 72 °C for 3 min 20 s, followed by 25 cycles of 98 °C for 10 s, 69 °C for 30 s, 72 °C for 3 min 20 s, with final incubation at 72 °C for 5 min. The rest of the process was identical to that using Phusion polymerase.

Sequence analysis

The Pacific Biosciences sequencing method produces long reads which cover the sequenced region multiple times. The raw data were processed with the RS_ReadsOfInsert protocol in a SMRT analysis pipeline that produced consensus reads. We filtered the acquired data based on these criteria: high quality reads longer than 3 kbp containing the beginning (CCCTCCCCTAA) and end (ATCGATGAATG) of the IGS and at least one transcription initiation site (TIS) (TATATAGGG). The reads were identified by barcodes.

To search the sequences for *SaII* boxes and gene and spacer promoters (GP and SP), we needed a clear definition of these elements. We defined the *SaII* box as a close occurrence of two and more *SaII* restriction sites (GTTCGAC). The borders of a *SaII* box are determined by the position of the first and last *SaII* restriction sites. The promoters were defined as the region from –55 to +6 nucleotides around a TIS sequence (TATATAGGG), where the underlined A is the first transcribed nucleotide +1. To search for these elements as well as for restriction sites, we used Python scripting.

To compare IGS sequences with information available in public databases we extracted two versions of the IGS reference. The first with 294, 294 and 500 bp long *SaII* boxes, here termed var1.294.294.500, was published in (Chandrasekhara et al. 2016) and the second was created as follows. The longest clone was BLAST searched against the official TAIR10 *A. thaliana* genome, and the TAIR10 genome fragments were merged into a unique IGS reference with 314 and 1077 bp long *SaII* boxes (online resource 2).

To summarize the IGS sequences obtained, we constructed a single IGS consensus as follows. We created 19 consensus sequences, one from each of the IGS types detected in the WT. These consensus sequences were aligned and a final WT consensus was extracted from the alignment. The consensus contains 316 and 1278 bp long *SaII* boxes (online resource 3). All multiple sequence alignments were created by ClustalW.

Restriction fragment analysis, probe labelling, and hybridization

The hybridization probes IGS1 and IGS2 were prepared by PCR using plasmid DNA containing a single IGS clone as a template. The PCR with IGS1F/IGS1R or IGS2F/IGS2R primers (online resource 1) contained 1.25 U of *Taq* polymerase (NEB) and was performed according to the manufacturer's instructions. The PCR product was extracted from a 1% agarose gel by a QIAX[®] II Gel Extraction Kit (Qiagen) and labeled with radioactive α -[³²P]dCTP according to the Rediprime II DNA Labeling System protocol (GE Healthcare Life Sciences).

To digest genomic DNA, 900 ng of *fas2* G2–/– or G3–/– and 500 ng of *fas2* G2+/+ was mixed with 30 U of *EcoRI* and 15 U of *HindIII* (NEB); the larger amount of mutant DNA was used to compensate for the loss of rRNA genes in *fas* mutants. The mix was incubated at 37 °C for 16 h, then 15 U *EcoRI* and 7.5 U of *HindIII* was added, the mixture was incubated for another 2 h, lyophilized, and subjected to electrophoresis on a 1.3% agarose gel overnight at 40 V. The digested DNA from the gel was transferred by Southern blot to a Hybond[™]–XL membrane (GE Healthcare) and hybridized with the IGS1 or IGS2 probe in 0.25 M Na-phosphate pH 7.5, 7% SDS, 0.016 M EDTA at 65 °C overnight. The membrane was washed three times in $2\times\text{SSC}$, 0.5% SDS at 65 °C. The hybridization process was repeated with chloroplast probes to check that the DNA was digested completely (Fajkus and Reich 1991).

Pulsed field gel electrophoresis

Cells were embedded in agarose blocks as described in (Fojtova et al. 2002). The DNA concentration and integrity were checked by a preliminary pulsed field gel electrophoresis

(PFGE) using a CHEF MAPPER (Biorad). Pieces of agarose blocks were washed three times in $0.1 \times \text{TE}$ (0.1 mM EDTA, 1 mM Tris-HCl, pH 8.0) and digested overnight with 80 U of *HindIII* (NEB). DNA was separated by PFGE (pulse times: 0.22–17.33 s linear) at 14 °C in a 1% (w/v) agarose gel in $0.5 \times \text{TBE}$ (4.5 mM Tris-HCl, 4.5 mM boric acid, 1.25 mM EDTA). The gel was Southern-blotted onto Hybond™-XL membrane (GE Healthcare) and hybridized with an IGS1 or IGS2 probe as described above.

Results

Sequencing data and nomenclature of IGS variants

To explore the IGS variants in detail, we used PacBio Single Molecule Real-Time sequencing, as well as cloning

followed by Sanger sequencing. The sequences were submitted to GenBank under the accession numbers KU994650–KU994739 for the clones, KU992939–KU994649 for the PacBio reads, and SRP071272 for the raw reads. The number of reads and detected IGS variants obtained in long reads from PacBio SMRT sequencing are given in Table 1 and summarized in Fig. 1. Corresponding results obtained by Sanger sequencing of cloned PCR products are shown in Table 2.

To simplify addressing variants in the text, we named them systematically. A complete IGS variant name consists of an 3'ETS type and a *SalI* box type separated by a dot. There are five types of 3'ETS variants, var1–var5, including the var5 newly described here (see the next Section). If a particular IGS contains two or three *SalI* boxes in a row, they are also represented by their type (length in bp) separated by a dot. For example, we use as a reference the IGS

Table 1 Number of IGS sequences obtained by PacBio sequencing

IGS variant	WT	Revertant lines				<i>fas1</i>			<i>fas2</i>
		6 (low)	3 (medium)	1 (high)	4 (high)	G1	G5	G7	G1
var1.294	7			7	5	6	4		4
var2.294	4								
var3.294									1
var1.294.633		1		1	1	1	1		
var2.294.1005	3								
var3.294.1005	1	1	1				1		
var1.294.1045	6	3	2	1	2		4		
var2.294.1045	24	1	1						
var3.294.1234			1						
var1.294.1254	1		1	2					
var3.314.1005	6	1	2					6	
var1.314.1045				1					
var1.366	4				2	1			
var5.366	54	63	14	7	1	20	26	52	10
var1.458	5		2	1	3	5	2		1
var2.458	3								
var3.458	190	165	27	162	72	137	90	72	51
var1.582	81	15		67	46	47	43	4	15
var2.582	7	3							1
var3.582	7			1	4		1		
var1.705	3			1	1	1			
var2.705			1						
var1.1045									1
var3.1045								2	
var1.1505	1								
Sum	407	253	52	252	137	218	172	136	84

The sequences are classified according to the IGS variant and the source of genomic DNA (WT, revertant lines with low, medium or high rDNA content and *fas* mutants in generations 1, 5 and 7). The sequences are deposited in GenBank under the accession numbers KU992939–KU994649. The names of individual IGS variants consist of the 3'ETS variant and the *SalI* box variant (based on the length in bp) separated by a dot

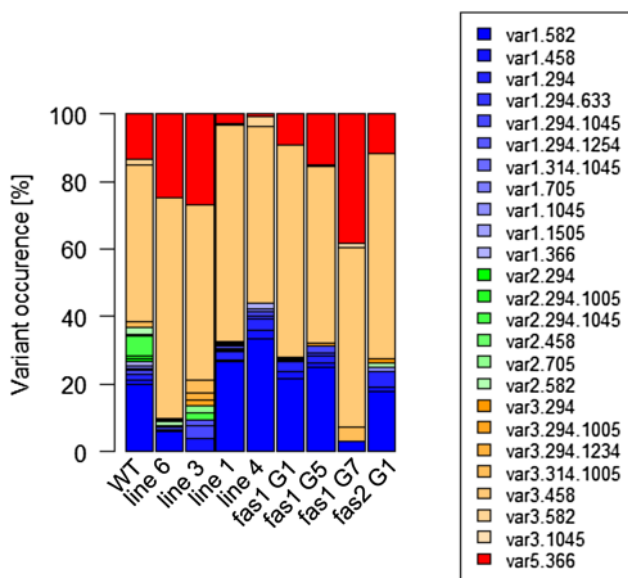


Fig. 1 Graphical summary of PacBio read distribution in all plants: WT, revertant lines with low (6), medium (3) or high (1, 4) rDNA content, and *fas1*, *fas2* mutants. The sequences can be found in GenBank under the accession numbers KU992939–KU994649. The name of individual IGS variants consists of the 3’ETS variant and the *SalI* box variant (based on the length in bp) separated by a *dot*

sequence assembled from sequences in public databases and published in (Chandrasekhara et al. 2016), which is 4725 bp long and contains two 294 bp-long *SalI* boxes, one 500 bp long *SalI* box, two spacer promoters, and one gene promoter (Fig. 2a). As this consists of a var1 3’ETS followed by three *SalI* boxes which are 294, 294 and 500 bp long, it is designated as var1.294.294.500 using our nomenclature.

Two different variants in the 5’ETS region characterised de novo in this paper were named varA and varB.

IGS variants in wild-type plants

Our results show that the IGS sequences within individual sister WT plants are far from being homogeneous. PCR amplification of the IGS, using primers for the conserved region of 18S and 25S genes, revealed two major products 3.5 and 4.5 kb long (online resource 4). Analysing our dataset of WT IGS sequences which consists of 407 IGS PacBio sequences and 16 clones (Tables 1, 2) we found variability in all three parts of the IGS, the 3’ETS, the NTS, and the 5’ETS.

The beginning of the IGS, which is delimited by the end of the 25S rRNA gene at the proximal site and by the first *SalI* restriction site at the distal site, contains the 3’ETS sequence according to which the rRNA genes can be classified into four types, var1–4 (Abou-Ellail et al. 2011). Given the small copy number of var4 in the genome, we did not detect it in WT reads, it was however found in *fas1* clones. In addition, our data shows one yet undescribed although abundant rRNA

Table 2 Numbers of individual IGS types obtained by cloning

IGS variant	<i>FAS</i> G2	<i>fas1</i> G2	<i>fas1</i> G5	<i>fas1</i> G7
var1.294	1			
var3.294.1045		1		
var1.294.1045	1	1		
var1.294.633		1		
var3.314.1005			1	5
var5.366	1	3	4	
var3.1045				1
var3.458	6	11	5	4
var1.458		1		
var1.582	2	4	5	1
var3.582		1		
var1.705		2		
var2.1045	1			
var4.797				1
Other clones which may contain a truncated <i>SalI</i> box region				
var3.1056				2
var3.314.892				1
var3.160.1005				1
var3.1342				1
var3.314.737				1
var3.953				1
var3.829				1
var2.1118				1
var2.294.942				1
var3.1015	1			1
var3.552	1			
var2.294.89.273	1			
var1.643	1			
var1.1020	1			
var1.222.986		1		
var1.664		1		
var3.68.633		1		
var3.779		1		
var3.531		1		
var3.573			1	

PCR products were cloned in *E. coli*. In total, 86 clones were sequenced by Sanger sequencing. The clones are classified according to the genomic DNA used in PCR. The majority of the clones belong to 14 significant IGS variants. Nevertheless, 22 clones appear to be truncated in the *SalI* box region. The sequences can be found in GenBank under the accession numbers KU994650–KU994739. The name of individual IGS variant consists of the 3’ETS variant and the *SalI* box variant (based on the length in bp) separated by a *dot*

gene type which we named var5 which is closely related to var3 but differs mainly in a 100 bp deletion at the distal site of the 3’ETS and can therefore be regarded as a var3 subtype. The first *SalI* restriction site is situated at position 778 bp (var1), 726 bp (var2), 655 bp (var3), 736 bp (var4) or 549 bp

(var5) from the end of the 25S rRNA gene. The alignment of WT 3'ETS variants is presented in Fig. 2b, c.

In the NTS, we detected significant length variability in the *SaII* boxes which is caused by differences in the number of 20 bp repeats, creating short (<1 kb) and long (>1 kb) *SaII* clusters. Altogether, we found ten different types of *SaII* boxes in the WT, and two types specific to *fas1*. We named them according to their length in bp (294, 314, 366, 458, 582, 705, 1005, 1045, 1254, or 1505). The alignment of all 12 *SaII* box types is presented in Fig. 2d. While some IGS have only one *SaII* box, others have a combination of two different types. Such variability is clearly seen in the PCR amplification of the IGS that shows two major products of length ~3.5 and ~4.5 kb (online resource 4). The short PCR product contains only one *SaII* box of these types: 294 (2.9%), 366 (14.1%), 458 (48.7%), 582 (23.2%) or 705 (0.7%). The long PCR product usually contains a combination of two *SaII* boxes separated by a spacer promoter. There are four different combinations of *SaII* box types (named according to the length of the individual *SaII* box in bp separated by a dot) occurring with different frequencies: 294.1005 (1%), 294.1045 (7.4%), 294.1254 (0.2%) and 314.1005 (1%). Rarely, we could see an IGS with only one long *SaII* box of the type 1045 (0.5%) or 1505 (0.2%). Thus within the shorter *SaII* boxes, only types 294 and 314 tend to be combined into longer IGS variants, while other *SaII* box types are isolated. In addition, the variant 314 is always combined into a longer IGS.

Combinations of 3'ETS variants var1–5 with adjoining *SaII* box types creates 19 unique IGS variants (Fig. 3). Some of the *SaII* box types tend to be connected to a specific 3'ETS variant, var1–5 (Table 3). The 366 *SaII* box occurred in 55 cases together with var5, while only in four cases with var1. The 458 *SaII* box is mostly connected with var3 (46.8%) and rarely with var2 (0.7%) or var1 (1.2%). The 582 *SaII* box is typically found with var1 (19.8%) and rarely with var2 (1.7%) or var3 (1.7%). It is interesting that although 3'ETS var1 is the longest variant, its most abundant adjoining IGS is rather short. The highest number of long IGS types, 294.1045, is associated with a 3'ETS of var2. To conclude, in the WT each 3'ETS seems to contain one most common type of adjoining IGS (var1.582, var2.294.1045, var3.458, or var5.366) (see Table 3 for details). Although our methods provide only semi-quantitative data, these four types of IGS are clearly more frequent than the remaining variants (Fig. 1).

It was previously shown by analysing 3'ETS rDNA expression that var1 is the most abundant and the least expressed 3'ETS variant and that it often undergoes epigenetic reprogramming (Abou-Ellail et al. 2011; Earley et al. 2006, 2010). Here we show that this 3'ETS variant is also associated with the highest number of different adjoining regions, forming eight different subtypes of IGS. The least

variability is associated with 3'ETS var5, where only one IGS subtype was found.

The last region of the IGS, the 1841 bp long 5'ETS situated upstream of the 18S rRNA gene, was described in detail (Gruendler et al. 1991). It contains two tandem 310 bp long repeats termed C1 and C2, which differ from each other in 4 SNPs at position 28, 130, 182 and 274 bp (Gruendler et al. 1991). Although the length heterogeneity in the 5'ETS is rare, we detected in 4% of sequences a 310 bp long deletion corresponding to a C repeat (Fig. 4b). We designated the major variant, containing both C repeats, as varA and the deletion variant, containing only one C repeat, as varB (Fig. 2a, online resource 5). Based on the SNPs, we conclude that the C repeat in varB results from various combinations of original C1 and C2 repeat rather than selective deletion of either C1 or C2 (online resource 5). The varB sequences come from different IGS variants: var3.458, var3.582, var1.582, var1.366 and var5.366. Therefore, the varB does not seem to be linked to a specific 3'ETS or *SaII* box variant.

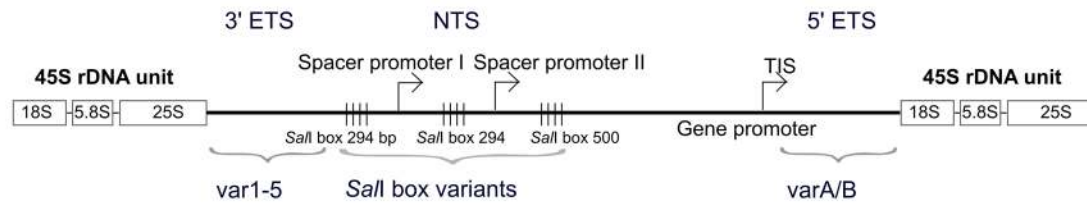
Consensus and reference WT IGS

To compare IGS sequences with information available in public databases we extracted an additional version of the IGS reference by BLAST searching of the longest clone against the official TAIR10 *A. thaliana* genome. The TAIR10 genome fragments have been merged into a unique IGS reference with 314 and 1077 bp long *SaII* boxes (online resource 2). To summarize the IGS sequences, we have constructed a single IGS consensus, which contains 316 and 1278 bp long *SaII* boxes (online resource 3).

IGS variants in *fas1* and *fas2* mutants

We describe here that in the WT the four most common IGS types—var1.582, var2.294.1045, var3.458, and var5.366—are present. In *fas1* and *fas2* mutants rDNA genes are systematically lost (Mozgova et al. 2010) and this is reflected in our data by the decreased variability in different types of IGS sequences obtained (Tables 1, 2). We found 19 IGS variants in the WT, while in both *fas* mutants there are only 6–10 variants for a given generation (Table 4). Some of these variants are apparently exclusive for mutants (resulting from non-allelic homologous recombination (Kirik et al. 2006)) and they do not occur in the WT: var3.294, var1.1045, var3.1045, var1.294.633, var3.294.1045 and var4.797 (Table 4, online resource 6). On the other hand, there are some WT variants which do not occur in *fas1* mutants: var1.1505, var2.294.1005, var2.294, var2.294.1045, var1.294.1254, var2.458, var2.1045 (online resource 6). Distinct occurrence of variants may result from the presumed mechanism of rDNA loss *via* single-strand annealing recombination events (Muchova et al. 2015).

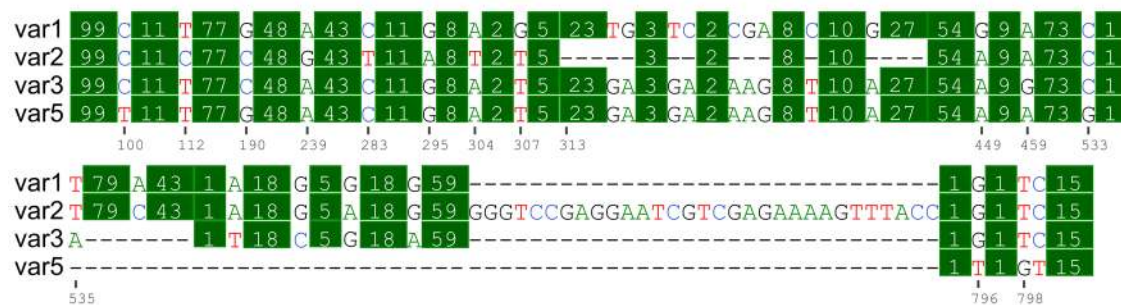
A Reference intergenic spacer (IGS)



B 3' ETS variants (except for var4)



C 3' ETS variants (except for var4)



D SalI box variants

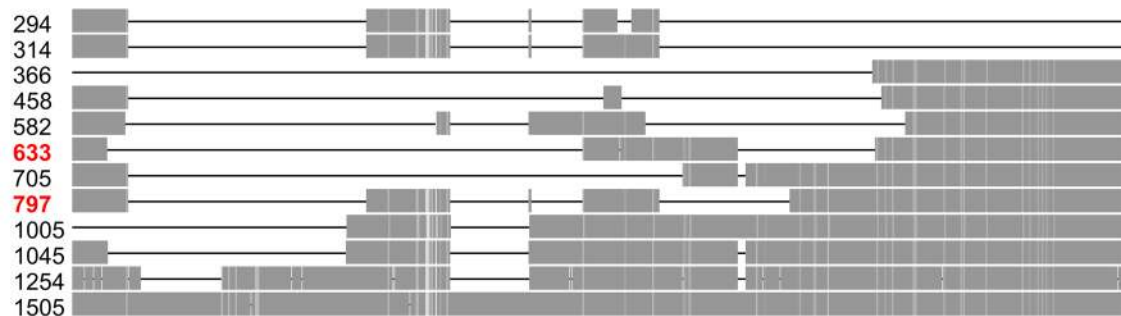


Fig. 2 Alignments of IGS variants. **a** Schematic view of the reference var1.294.294.500 (Chandrasekhara et al. 2016). **b** Alignment of the 3'ETS region found in the WT, var1/2/3/5. Var4 was not detected in our dataset and therefore it is not included in this alignment. Four different 3'ETS variants are shown from the end of the 25S rRNA gene to the first *SalI* restriction site. Lighter shades represent lower sequence identity. **c** Alignment of the 3'ETS region. Positions containing different

sequences are shown in full, while regions containing identical bases are displayed as a box (green) with a number representing the number of bp in each box. **d** Alignment of *SalI* box types. Lighter shades represent lower sequence identity. Numbers represent the length of individual *SalI* box in bp. *SalI* boxes 633 and 797 (red) are found only in *fas1* mutants

Restriction fragment length polymorphism

We further focused on analysis of the IGS at the genomic DNA level, using selected restriction enzymes for RFLP analyses. Using in silico prediction of restriction sites, we created restriction maps of individual IGS variants and

compared them to the signals visible in RFLP analysis (Fig. 4).

Firstly, genomic DNA was digested with *EcoRI* and *HindIII* and hybridized with an IGS1 probe specific to the 3'ETS (Fig. 4a). The most abundant IGS variants could all be mapped to a strong signal. Some of the less abundant

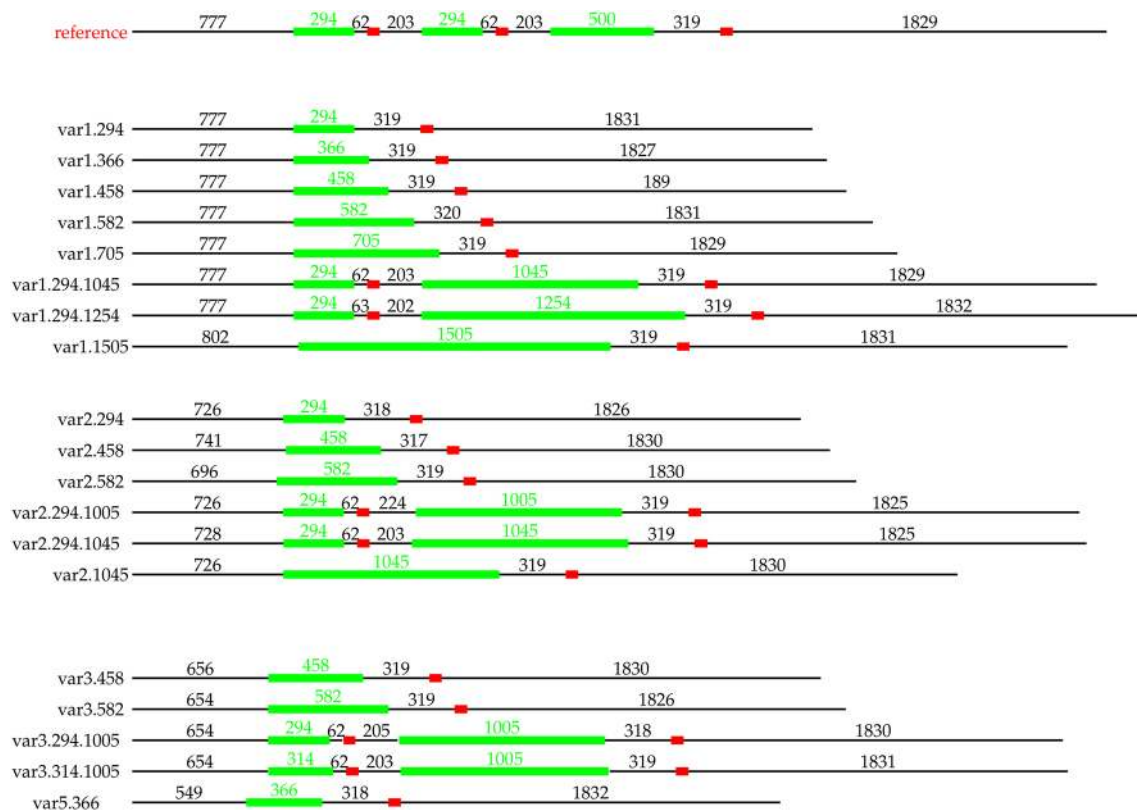


Fig. 3 Distribution of promoters and *SalI* boxes in WT IGS variants. Lines represent individual IGS variants showing schematic length differences between IGS types. *SalI* boxes are in green, promoters in red. The numbers above each line represent the length (in bp) of the

corresponding region. The name of individual IGS variants consists of the 3'ETS variant and the *SalI* box variant (based on the length in bp) separated by a dot

variants mapped to a smeared signal and their existence could not be sufficiently supported by RFLP. There were also three relatively strong signals which could not be assigned to any of our variants, suggesting the existence of other variants including the reference var1.294.294.500 (Chandrasekhara et al. 2016).

Secondly, genomic DNA was digested with *EcoRI* and hybridized with an IGS2 probe specific to the 5'ETS (Fig. 4b). The 5'ETS variants varA/B could be mapped to the signals, the varB signal being considerably weaker than the varA signal, which corresponds to our sequencing data. The RFLP analysis also revealed an inhibition of cleavage at the first *EcoRI* restriction site, situated in the 18S rRNA gene, due to an overlapping CpG methylation in a significant subset of 18S rRNA genes, which led to incomplete DNA digestion and a subsequent 1506 bp long electrophoretic mobility shift of varA/varB signals.

IGS variants in revertant lines

FAS1FAS1/FAS2FAS2 plants segregated from a cross between *fas1* and *fas2* plants (the progeny of *FAS1fas1/FAS2fas2* double heterozygotes) show very uneven and asymmetric

recovery of rDNA genes at the levels of their copy numbers, distribution of rDNA loci, or expression of 3'ETS variants (Pavlišťová et al. 2016). Therefore, we selected four revertant lines as representatives of revertants with a low (line 6), medium (line 3) or high (lines 1 and 4) amount of rDNA (Table 1). Unlike the WT, these lines also express 3'ETS var1; lines 1 and 4 in addition to var3, and lines 4 and 6 with relatively balanced expression of var1-3. The lines 1 and 4 have no detectable var2 in their genome (based on PCR studies). Further, line 4 has acquired a complete change of its NOR set-up; the vast majority of its rDNA is present on chromosome 2, thus forming an active NOR2 associated with the nucleolus (Pavlišťová et al. 2016).

Our sequencing data show that the variability of the IGS in revertant lines is decreased, showing 10–12 variants for a single line compared to the 19 variants in the WT. This is a similar situation to the *fas1* and *fas2* mutants, indicating that rearrangements of the IGS are probably relatively rare during recovery. In mutants, only a subset of IGS types remains present, and thus without major reorganisation the spectrum of IGS types cannot expand to the original WT level. This supports the view that restoration of CAF-1 function occurs preferentially *via* precise (allelic) homologous recombination

Table 3 Numbers of PacBio reads and clones in WT classified according to *SalI* box type and 3'ETS variant

PacBio reads and clones in WT				
<i>SalI</i> box type	3'ETS variant			
	var1	var2	var3	var5
294	8 (1.9%)	4 (1%)		
366	4 (1%)			55 (13.1%)
458	5 (1.2%)	3 (0.7%)	196 (46.8%)	
582	83 (19.8%)	7 (1.7%)	7 (1.7%)	
705	3 (0.7%)			
294.1005		3 (0.7%)	1 (0.2%)	
294.1045	7 (1.7%)	24 (5.7%)		
294.1254	1 (0.2%)			
314.1005			6 (1.4%)	
1045		1 (0.2%)		
1505	1 (0.2%)			

The *SalI* box types of the IGS tend to be joined with specific 3'ETS variants, var1–5. The 366 *SalI* box is typical for var5, the 458 is typical for var3, and the 582 is typical for var1

that prevents further loss of rDNA. Var1.582 is preferentially recovered in lines 1 and 4, and var3.458 in lines 3 and 6 (Fig. 1; Table 1). Both of these variants belong to the most abundant type in the WT and are still present in *fasI* G7 plants.

Regarding the origin of IGS variants, most of the revertant variants were found in the WT as well as in *fasI* mutants (online resource 6). However, there are some exceptions; variant var1.294.633 seems to come from a mutant progenitor since it was found only in revertant lines and *fasI* mutants, with no occurrence in the WT. On the other hand, variants var2.294.1045 and var1.294.1254 were found only in revertant lines and the WT, and thus were probably formed de novo in revertants. Therefore, a small percentage of new variants might have originated from rDNA rearrangements. In revertant lines with medium or high rDNA content, we further detected three rare IGS variants which were not found in either the WT or *fasI* mutants (var2.705, var1.314.1045 and var3.294.1234).

To conclude, revertant lines share specific variants with both WT as well as *fasI* mutants, with only sporadic formation of new variants. It seems that the loss of IGS variability in parental mutants propagates to revertants. Apparently, decreased IGS variability leads to preferential recovery of variants which remain in the genome at the moment of *FAS* restoration, after the extensive rDNA loss in previous mutant generations.

Organisation of rDNA clusters

We next examined whether the newly formed rDNA clusters in revertant lines appeared more similar to the *fasI* mutant

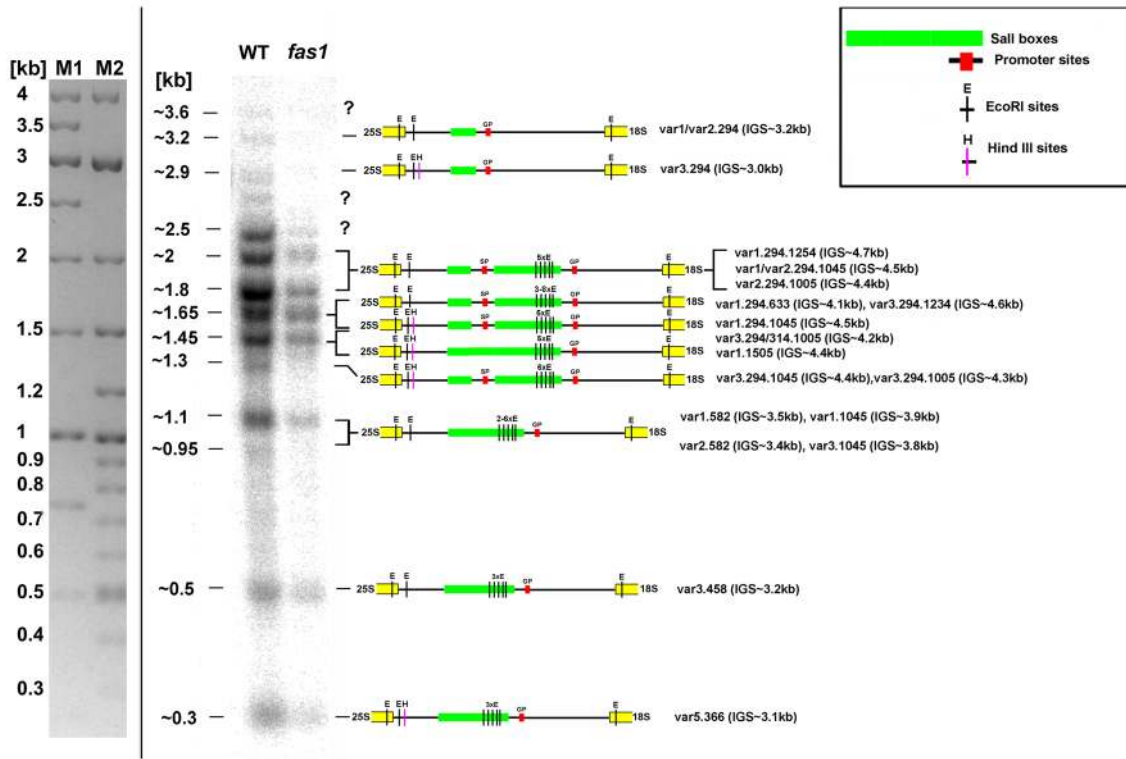
or to the WT. We used a full *HindIII* digest in combination with PFGE to analyse the organisation of rDNA clusters in the WT, *fasI*, and revertant lines. This method was described previously and generates an approximately 10 kb ladder representing mono- and oligomers of rDNA units (Copenhaver et al. 1995). The most recent studies show that rDNA genes with *HindIII* sites are present at chromosome 4 in the IGS type containing mostly var3 3'ETS and rarely also in var1 or var4 IGS types, while rDNA clusters on chromosome 2 are *HindIII* site-free (Chandrasekhara et al. 2016). Looking for *HindIII* restriction in our set of clones and PacBio reads, we can confirm that var2 and most of the var1 sequences do not contain *HindIII* restriction sites while most of the var3 and var5 do (for details see online resource 7).

We detected by PFGE that in late generations (G4 or G6) of *fasI* mutants the higher molecular weight rDNA clusters (presumably lacking the *HindIII* site) are completely lost (follow the asterisks in Fig. 5). Previous data showed that active rDNA copies on chromosome four were lost first, followed by inactive copies (Copenhaver et al. 1995; Chandrasekhara et al. 2016; Muchova et al. 2015; Pontvianne et al. 2013). Here we observed the decrease in high molecular weight DNA fragments in *fasI* (Fig. 5, asterisks), while *HindIII* site-containing rDNA copies (likely on chromosome four) are still maintained in the *fasI* genome even in later generations of mutants (Fig. 5). This result rather suggests that loss of rDNA copies occurs at both chromosomal loci simultaneously. It is interesting that although copies lacking a *HindIII* site are mostly lost in *fasI*, in revertant lines (which come from G4 of mutants) rDNA clusters become even more heterogeneous than in the WT in terms of occurrence of the *HindIII* sites (follow the arrows in Fig. 5). The variation in outcomes of the rDNA accumulation suggests a stochastic character for initiation of the recovery process, which is followed by subsequent amplification. The greatest similarity to the WT was observed in line 4, which contains high molecular weight DNA fragments (thus *HindIII* site-lacking rDNA copies) originating from the WT. However, in this line rDNA accumulates on chromosome 2, while being repressed on chromosome 4. The variability in organisation of the recovered rDNA possibly also reflects the variable 3'ETS expression pattern observed in reverted lines, which does not follow the setting in the WT where NOR4 contains all active variants and NOR2 is inactive (Pavlišťová et al. 2016) (Chandrasekhara et al. 2016).

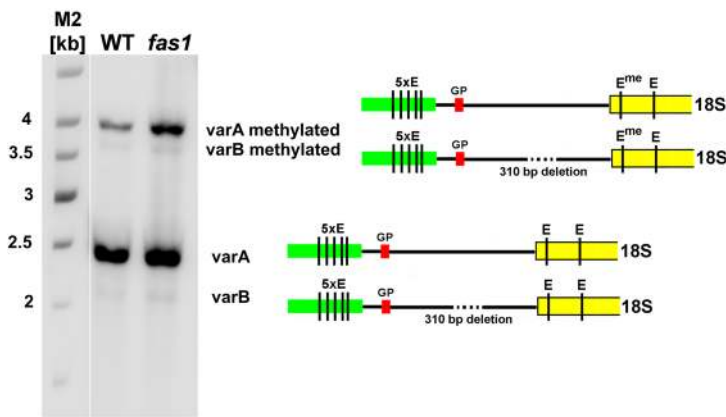
Discussion

Ribosomal RNA gene arrays in various organisms have been the subject of numerous studies concerning their dynamic evolution and regulation of expression. A phenomenon known as concerted evolution has been demonstrated to

A Variability of 3'ETS



B Variability of 5'ETS



(C)



Fig. 4 In silico analysis of IGS restriction fragments. **a** Genomic DNA from *fas1* plants in G2 was digested with *EcoRI* and *HindIII* and hybridized with the IGS1 probe which covers the 3'ETS (see panel c). The IGS variants were matched to the image according to in silico restriction digestion. The observed RFPL is mostly due to a different number and length of *Sall* boxes, in combination with the presence or absence of internal *EcoRI* and *HindIII* sites. Three relatively strong signals (*question marks*) could not be assigned to any of the identified

IGS variants, suggesting that there may be still-unknown IGS variants or that some combinations of overlapping cytosine methylation prevent *EcoRI* from complete DNA digestion at a subset of *EcoRI* sites, resulting in longer fragments. M represents a molecular weight marker (M1–1 kb DNA ladder; M2–2-log ladder). **b** Genomic DNA from *fas1* plants in G3 was digested with *EcoRI* and hybridized with the IGS2 probe, which shows the polymorphism in the 5'ETS. **c** The positions of IGS1 and IGS2 probes

Table 4 IGS variability in *fas* mutants

Genotype	#variants	Variants
<i>fas2</i> G1	8	var1.294, var3.294*, var5.366, var1.458, var3.458, var1.582, var2.582, var1.1045*
<i>fas1</i> G1	8	var1.294, var1.366, var5.366, var1.458, var3.458, var1.582, var1.705, var1.294.633*
<i>fas1</i> G2	9	var5.366, var1.458, var3.458, var1.582, var3.582, var1.705, var1.294.633*, var1.294.1045, var3.294.1045*
<i>fas1</i> G5	10	var1.294, var5.366, var3.458, var1.582, var3.582, var1.294.633*, var1.294.1045, var3.294.1045, var3.294.1005, var3.314.1005
<i>fas1</i> G7	6	var5.366, var3.458, var1.582, var3.1045*, var3.314.1005, var4.797*

There are 6–10 different variants in a single generation of *fas* mutants. Variants with the symbol asterisk are exclusive to mutants and were not found in the WT. The most abundant IGS types are shown in bold font. The name of individual IGS variant consists of the 3'ETS variant and the *SalI* box variant (based on the length in bp) separated by a dot

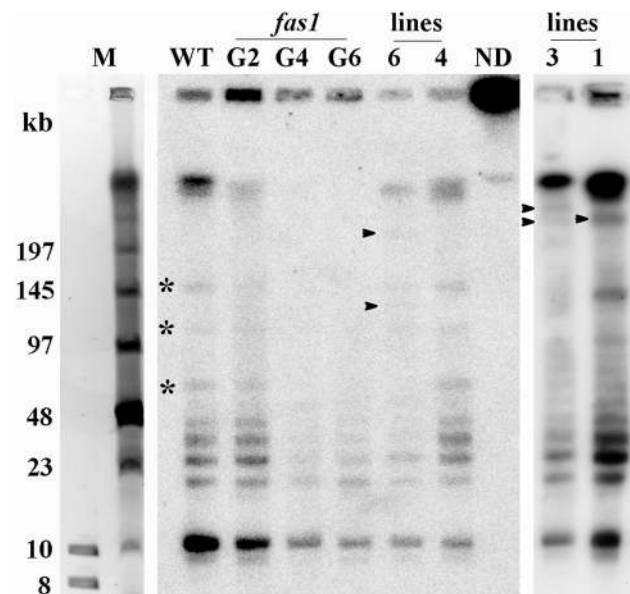


Fig. 5 PFGE analysis of rDNA. High molecular weight DNA digested with *HindIII* was separated by PFGE and hybridized with an IGS2 probe. WT, *fas1* (generations 2, 4 and 6) and revertant lines 1, 3, 4, 6 [WT lines segregated from a cross between *fas1xfas2*, (Pavlišťová et al. 2016)] are shown. ND represents a non-digested DNA control; M is a molecular weight standard. Asterisks mark the fraction of high molecular weight rDNA present in WT plants, lost in *fas1* mutants and recovered in revertant lines with higher copy number. Arrowheads mark newly-formed rDNA fragments present in revertant lines

be responsible for high sequence similarity of rRNA genes within a species. These studies suggested homologous recombination events as the mechanism of rapid sequence

homogenization. The major driving force in the evolution of the rRNA genes is considered to be unequal crossover, with sister chromatid exchange occurring more often than exchange between homologs (Eickbush and Eickbush 2007). This can be demonstrated also in *A. thaliana*, where the rRNA genes within a single NOR are more similar to one another than they are to the rRNA genes on another chromosome (Copenhaver and Pikaard 1996b). Our data based on analysis of rDNA clusters by *HindIII* supports the theory of rDNA loci homogenisation and preferential recombination between homologues. In revertant WT lines, where the amount of rDNA was reduced due to *fas1* or *2* mutations, usually one type of rDNA variant is preferentially recovered after reintroduction of functional *FAS* alleles.

Overall, sequence variation in rDNA is largely limited to the IGS that contains genetic elements controlling RNA polymerase transcription, the gene promoter, enhancers, spacer promoters and terminators. Since rRNA gene activity, both at the level of individual genes and of the whole rDNA loci, is at the same time subject to epigenetic regulations (Pontvianne et al. 2013; Preuss and Pikaard 2007), it is important to know the actual sequence structure of the IGS and its natural variation so that contributions of genetic and epigenetic phenomena can be assessed realistically. Although *A. thaliana* has been used frequently as a model organism in rDNA studies, it is surprising that no contiguous sequences of rDNA units have been available so far, and this knowledge gap was bridged using artificially assembled rDNA fragments.

Interestingly, our data on IGS differ substantially from the reference var1.294.294.500 which was assembled from fragments available in public databases and used in a recent study (Chandrasekhara et al. 2016). The reference contains three *SalI* boxes in a row separated by two spacer promoters. In contrast, we found only one IGS clone (var2.294.89.273) which follows the same arrangement and even this one differs from the reference in the length of the *SalI* box.

Importantly, our results associate previously known variants in the 3'ETS with their actual IGS sequence context, and reveal a novel 3'ETS variant, var5. The distribution of 3'ETS variants on individual NORs was described in detail (Chandrasekhara et al. 2016) showing that var2 IGS types are located the transcriptionally active NOR4, while var1 is on the inactive NOR2. Var3 is distributed on both NORs, but NOR4 contains a var3 with a *HindIII* site. Mutants in the *FAS1* (or *FAS2*) gene lose around 20% of their rDNA repeats per generation, thus representing a unique model to study IGS dynamics. Previous studies showed that transcriptionally active copies at NOR4 are depleted first, then inactive var1-containing copies are activated and subsequently also depleted (Muchova et al. 2015; Pontvianne et al. 2013). Consistent with these studies, when we analysed three different generations of *fas1* mutants (G1,G5,G7) the

actively transcribed var2 (var2.294.942, truncated version of var2.294.1045) was detected only once (G7, conventional cloning, Table 2), while in the WT this IGS variant represents around 10% of SMRT reads (Table 1; Fig. 1). Further, the abundance of the frequently occurring inactive var1.582 (around 20% in the WT) is very low in G7 of *fas1* mutants (Table 1; Fig. 1). Based on SMRT results, the dramatic decrease of this variant occurs between generations 5 and 7 of *fas1* (Table 1; Fig. 1) leaving the remaining copies of var3.458 and var5.366. Aside from the rDNA loss, it is possible that some re-organisation occurs inside the IGS in *fas1* plants due to the increased levels of non-allelic homologous recombination (Kirik et al. 2006). Two additional *SalI* box types, 633 and 797, are found in mutants. Their alignment to the WT *SalI* box types shows that 633 is related to 1045 and might have resulted from a deletion in 1045, while since the *SalI* box 797 is similar to 314 at the proximal site and 705 at the distal site we conclude that 797 might have resulted from their fusion. In the light of the concept of concerted evolution of rDNA, *fas* mutants and plants with a *fas* mutation history and recovered CAF-1 function display acceleration of concerted evolution as demonstrated by a loss, gain and spreading of specific IGS variants within a limited number of generations. This corresponds to generally increased levels of homology-dependent recombination events (Endo et al. 2006; Kirik et al. 2006; Takeda et al. 2004) and direct involvement of these events in loss of rDNA in *fas1* and *fas2* mutants (Muchova et al. 2015).

In conclusion, we describe here the variant arrangements of the IGS in the 45S rDNA in *A. thaliana* WT plants, *fas1* and *fas2* mutants, and plants with restored CAF-1 function. This detailed characterisation allowed the description of a new variant in the 3'ETS region of pre-rRNA, termed var5, and of the preferential association of 3'ETS variants with specific IGS arrangements. Dysfunctional *FAS1* (or *FAS2*) leads to selective loss of some variants and the generation of new IGS variants. Overall, both *fas* mutants show less variability than WT plants. These results correspond to the presumed mechanism of the loss of rDNA copies via the single strand annealing type of homology-dependent repair (Muchova et al. 2015) which, besides causing an absolute reduction of the number of rDNA copies, at the same time simplifies the original WT spectrum of rDNA variants. Plants with restored *FAS1* and *FAS2* function show an IGS spectrum similar to that of their parental mutants, suggesting that rDNA recovery occurs through a relatively precise DNA synthesis-dependent homologous recombination mechanism.

Acknowledgments We thank Veronika Pavlišťová for providing us with the revertant *A. thaliana* lines. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Research, Development, and Innovations

Infrastructures” (CESNET LM2015042) is greatly appreciated. This work was supported by GACR 13-11563P, 16-04166Y and by the Ministry of Education, Youth and Sports of the Czech Republic-projects KONTAKT II no. LH15189 and CEITEC 2020 (LQ1601). Work in the laboratory of C.G. is supported by Grants BFU2012-34821 and BFU2015-68396-R.

Author contributions J.F. and M.D. designed the study, K.H., M.D., I.M. and L.V. performed experiments, all authors performed analysis and interpretation of results, K.H., M.D. and J.F. wrote the manuscript.

References

- Abou-Ellail M, Cooke R, Saez-Vasquez J (2011) Variations on a team major and minor variants of *Arabidopsis thaliana* rDNA genes. *Nucleus* 2:294–299. doi:10.4161/nucl.2.4.16561
- Cavallero S, De Liberato C, Friedrich KG, Di Cave D, Masella V, D’Amelio S, Berrilli F (2015) Genetic heterogeneity and phylogeny of *Trichuris* spp. from captive non-human primates based on ribosomal DNA sequence data. *Infect Genet Evol* 34:450–456. doi:10.1016/j.meegid.2015.06.009
- Chandrasekhara C, Mohannath G, Blevins T, Pontvianne F, Pikaard CS (2016) Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Gene Dev* 30:177–190. doi:10.1101/gad.273755.115
- Copenhaver GP, Pikaard CS (1996a) RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J* 9:259–272. doi:10.1046/j.1365-313X.1996.09020259.x
- Copenhaver GP, Pikaard CS (1996b) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J* 9:273–282. doi:10.1046/j.1365-313X.1996.09020273.x
- Copenhaver GP, Doelling JH, Gens S, Pikaard CS (1995) Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*. *Plant J* 7:273–286
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep: Version II. *Plant Mol Biol Report* 1:19–21
- Durut N et al (2014) A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in *Arabidopsis*. *Plant Cell* 26:1330–1344. doi:10.1105/tpc.114.123893
- Dvorackova M, Fojtova M, Fajkus J (2015) Chromatin dynamics of plant telomeres and ribosomal genes. *Plant J* 83:18–37. doi:10.1111/tpj.12822
- Earley K et al (2006) Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* 20:1283–1293. doi:10.1101/gad.1417706
- Earley KW et al (2010) Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev* 24:1119–1132. doi:10.1101/gad.1914110
- Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175:477–485. doi:10.1534/genetics.107.071399
- Elliott TA, Stage DE, Crease TJ, Eickbush TH (2013) In and out of the rRNA genes: characterization of Pokey elements in the sequenced *Daphnia* genome. *Mobile DNA*. doi:10.1186/1759-8753-4-20
- Endo M et al (2006) Analysis of *Arabidopsis* CAF-1 mutants showing enhanced homologous recombination. *Plant Cell Physiol* 47:S60–S60

- Exner V, Taranto P, Schonrock N, Gruissem W, Hennig L (2006) Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. *Development* 133:4163–4172 doi:[10.1242/dev.02599](https://doi.org/10.1242/dev.02599)
- Fajkus J, Reich J (1991) Evaluation of restriction endonuclease cleavage of plant nuclear-DNA using contaminating chloroplast DNA. *Folia Biol Prague* 37:224–226
- Fojtova M, Fulneckova J, Fajkus J, Kovarik A (2002) Recovery of tobacco cells from cadmium stress is accompanied by DNA repair and increased telomerase activity. *J Exp Bot* 53:2151–2158
- Garcia S, Kovarik A (2013) Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity* 111:23–33. doi:[10.1038/hdy.2013.11](https://doi.org/10.1038/hdy.2013.11)
- Garcia S, Panero JL, Siroky J, Kovarik A (2010) Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC Plant Biol* 10:176. doi:[10.1186/1471-2229-10-176](https://doi.org/10.1186/1471-2229-10-176)
- 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. doi:[10.1105/tpc.15.00791](https://doi.org/10.1105/tpc.15.00791)
- Gruendler P, Unfried I, Pointner R, Schweizer D (1989) Nucleotide-Sequence of the 25S–18S ribosomal gene spacer from *Arabidopsis thaliana*. *Nucleic Acids Res* 17:6395–6396. doi:[10.1093/nar/17.15.6395](https://doi.org/10.1093/nar/17.15.6395)
- Gruendler P, Unfried I, Pascher K, Schweizer D (1991) rDNA intergenic region from *Arabidopsis thaliana*. Structural analysis, intraspecific variation and functional implications. *J Mol Biol* 221:1209–1222
- Grummt I, Pikaard CS (2003) Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Bio* 4:641–649. doi:[10.1038/nrm1171](https://doi.org/10.1038/nrm1171)
- Grummt I, Kuhn A, Bartsch I, Rosenbauer H (1986) A transcription terminator located upstream of the mouse rDNA initiation site affects rRNA synthesis. *Cell* 47:901–911
- Han EH, Cho K, Goo Y, Kim M, Shin YW, Kim YH, Lee SW (2016) Development of molecular markers, based on chloroplast and ribosomal DNA regions, to discriminate three popular medicinal plant species, *Cynanchum wilfordii*, *Cynanchum auriculatum*, and *Polygonum multiflorum*. *Mol Biol Reports*. doi:[10.1007/s11033-016-3959-1](https://doi.org/10.1007/s11033-016-3959-1)
- Kaya H, Shibahara KI, Taoka KI, Iwabuchi M, Stillman B, Araki T (2001) FASCIATA genes for chromatin assembly factor-1 in arabidopsis maintain the cellular organization of apical meristems. *Cell* 104:131–142
- Kelly LJ et al (2015) Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytol* 208:596–607. doi:[10.1111/nph.13471](https://doi.org/10.1111/nph.13471)
- Kirik A, Pecinka A, Wendeler E, Reiss B (2006) The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. *Plant Cell* 18:2431–2442. doi:[10.1105/tpc.106.045088](https://doi.org/10.1105/tpc.106.045088)
- Kobayashi T, Horiuchi T, Tongaonkar P, Vu L, Nomura M (2004) SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. *Cell* 117:441–453
- Konstantinova P, Yli-Mattila T (2004) IGS-RFLP analysis and development of molecular markers for identification of *Fusarium poae*, *Fusarium langsethiae*, *Fusarium sporotrichioides* and *Fusarium kyushuense*. *Int J Food Microbiol* 95:321–331. doi:[10.1016/j.ijfoodmicro.2003.12.010](https://doi.org/10.1016/j.ijfoodmicro.2003.12.010)
- Lauro FM, Chastain RA, Blankenship LE, Yayanos AA, Bartlett DH (2007) The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl Environ Microbiol* 73:838–845. doi:[10.1128/AEM.01726-06](https://doi.org/10.1128/AEM.01726-06)
- Leyser HMO, Furner IJ (1992) Characterization of 3 shoot apical meristem mutants of *Arabidopsis thaliana*. *Development* 116:397–403
- Lin RQ, Shu L, Zhao GH, Cheng T, Zou SS, Zhang Y, Weng YB (2014) Characterization of the intergenic spacer rDNAs of two pig nodule worms, *Oesophagostomum dentatum* and *O. quadripinulatum*. *Sci World J* 2014:147963. doi:[10.1155/2014/147963](https://doi.org/10.1155/2014/147963)
- Long Q et al (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet* 45:884–890. doi:[10.1038/ng.2678](https://doi.org/10.1038/ng.2678)
- Mandakova T, Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in $x=7$ crucifer species (Brassicaceae). *Plant Cell* 20:2559–2570. doi:[10.1105/tpc.108.062166](https://doi.org/10.1105/tpc.108.062166)
- Marcilla A et al (2001) The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Mol Phylogenet Evol* 18:136–142 doi:[10.1006/mpev.2000.0864](https://doi.org/10.1006/mpev.2000.0864)
- Mozgova I, Mokros P, Fajkus J (2010) Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in *Arabidopsis thaliana*. *Plant Cell* 22:2768–2780. doi:[10.1105/tpc.110.076182](https://doi.org/10.1105/tpc.110.076182)
- Muchova V, Amiard S, Mozgova I, Dvorackova M, Gallego ME, White C, Fajkus J (2015) Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants. *Plant J* 81:198–209. doi:[10.1111/tbj.12718](https://doi.org/10.1111/tbj.12718)
- Pavlišťová V, Dvořáčková M, Jež M, Mozgová I, Mokroš P, Fajkus J (2016) Phenotypic reversion in mutants of by reintroduction of genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J*. doi:[10.1111/tbj.13257](https://doi.org/10.1111/tbj.13257)
- Pikaard CS et al (1990) Enhancers for RNA polymerase I in mouse ribosomal DNA. *Mol Cell Biol* 10:4816–4825
- Pontes O et al (2003) Natural variation in nucleolar dominance reveals the relationship between nucleolar organizer chromatin topology and rRNA gene transcription in *Arabidopsis*. *P Natl Acad Sci USA* 100:11418–11423. doi:[10.1073/pnas.193252100](https://doi.org/10.1073/pnas.193252100)
- Pontvianne F et al (2010) Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet* 6:e1001225. doi:[10.1371/journal.pgen.1001225](https://doi.org/10.1371/journal.pgen.1001225)
- Pontvianne F et al (2013) Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Gene Dev* 27:1545–1550. doi:[10.1101/gad.221648.113](https://doi.org/10.1101/gad.221648.113)
- Preuss S, Pikaard CS (2007) rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. *Bba Gene Struct Expr* 1769:383–392 doi:[10.1016/j.bbaexp.2007.02.005](https://doi.org/10.1016/j.bbaexp.2007.02.005)
- Pruitt RE, Meyerowitz EM (1986) Characterization of the genome of *Arabidopsis thaliana*. *J Mol Biol* 187:169–183. doi:[10.1016/0022-2836\(86\)90226-3](https://doi.org/10.1016/0022-2836(86)90226-3)
- Ramirez-Parra E, Gutierrez C (2007) E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. *Plant Physiol* 144:105–120. doi:[10.1104/pp.106.094979](https://doi.org/10.1104/pp.106.094979)
- Reinholz E (1966) Radiation induced mutants showing changed inflorescence characteristics *Arabid Inf Serv* 3:19–20
- Schonrock N, Exner V, Probst A, Gruissem W, Hennig L (2006) Functional genomic analysis of CAF-1 mutants in *Arabidopsis thaliana*. *J Biol Chem* 281:9560–9568. doi:[10.1074/jbc.M513426200](https://doi.org/10.1074/jbc.M513426200)
- Schubert I, Wobus U (1985) In situ hybridisation confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* 92:143–148. doi:[10.1007/BF00328466](https://doi.org/10.1007/BF00328466)

- Smith S, Stillman B (1989) Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell* 58:15–25
- Symonova R et al (2013) Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol Biol* 13:42. doi:[10.1186/1471-2148-13-42](https://doi.org/10.1186/1471-2148-13-42)
- Takeda S et al (2004) BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. *Gene Dev* 18:782–793. doi:[10.1101/gad.295404](https://doi.org/10.1101/gad.295404)
- Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, Markow TA (2005) The functional significance of ribosomal (r)DNA variation: Impacts on the evolutionary ecology of organisms. *Annu Rev Ecol Evol S* 36:219–242. doi:[10.1146/annurev.ecolsys.36.102003.152620](https://doi.org/10.1146/annurev.ecolsys.36.102003.152620)