Molecular Phylogenetics and Evolution 61 (2011) 321–332

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Restless 5S: The re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants

Susann Wicke [a,b,*], Andrea Costa [b], Jesùs Muñoz [c], Dietmar Quandt [b,*]

[a] Institute for Evolution and Biodiversity, University of Muenster, Huefferstr. 1, D-48149 Muenster, Germany
[b] Nees Institute for Biodiversity of Plants, University of Bonn, Meckenheimer Allee 170, D-53115 Bonn, Germany
[c] Real Jardín Botánico Madrid (CSIC), Plaza de Murillo 2, E-28014 Madrid, Spain

ARTICLE INFO

Article history:
Received 14 December 2010
Revised 23 June 2011
Accepted 27 June 2011
Available online 3 July 2011

Keywords:
Ribosomal DNA
Molecular evolution
Land plants
5S rDNA
Concerted evolution
(Retro-)transposition

ABSTRACT

Among eukaryotes two types of nuclear ribosomal DNA (nrDNA) organization have been observed. Either all components, i.e. the small ribosomal subunit, 5.8S, large ribosomal subunit, and 5S occur tandemly arranged or the 5S rDNA forms a separate cluster of its own. Generalizations based on data derived from just a few model organisms have led to a superimposition of structural and evolutionary traits to the entire plant kingdom asserting that plants generally possess separate arrays.

This study reveals that plant nrDNA organization into separate arrays is not a distinctive feature, but rather assignable almost solely to seed plants. We show that early diverging land plants and presumably streptophyte algae share a co-localization of all rRNA genes within one repeat unit. This raises the possibility that the state of rDNA gene co-localization had occurred in their common ancestor. Separate rDNA arrays were identified for all basal seed plants and water ferns, implying at least two independent 5S rDNA transposition events during land plant evolution. Screening for 5S derived *Cassandra* transposable elements which might have played a role during the transposition events, indicated that this retrotransposon is absent in early diverging vascular plants including early fern lineages. Thus, *Cassandra* can be rejected as a primary mechanism for 5S rDNA transposition in water ferns. However, the evolution of *Cassandra* and other eukaryotic 5S derived elements might have been a side effect of the 5S rDNA cluster formation. Structural analysis of the intergenic spacers of the ribosomal clusters revealed that transposition events partially affect spacer regions and suggests a slightly different transcription regulation of 5S rDNA in early land plants. 5S rDNA upstream regulatory elements are highly divergent or absent from the LSU–5S spacers of most early divergent land plant lineages. Several putative scenarios and mechanisms involved in the concerted relocation of hundreds of 5S rRNA gene copies are discussed.

1. Introduction

Nuclear ribosomal DNA (nrDNA) is an essential component in organismal genomes and commonly appears in high copy numbers. The relative amount and size of ribosomal DNA (rDNA) arrays in the nuclear DNA (nDNA) can be highly variable, ranging from ~0.4% in human (with approx. 500 copies scattered on five chromosomes) to nearly 17% in *Drosophila melanogaster* in some cases (80–600 copies on the sex chromosomes). To date, the eukaryotic clade showing the largest number of rDNA repeats is the plant lineage (Streptophytes plus Chlorophytes) with up to 22,000 transcription units (Rogers and Bendich, 1987). On a transcription level, ribosomal RNA (rRNA) is the most abundant making up on average 80% of a cells total RNA (Brown, 2002). Therefore, ribosomal RNA and DNA still are the targets of countless research studies in all areas of life sciences.

Due to its central role in cell metabolism, nrDNA shows an extraordinarily high degree of conservation among organisms. In most organisms, three out of the four rRNA components are encoded as a conserved cluster, referred to as 35S or 45S-array (45S cluster). Each array harbors a variable number of processed tandem transcription units which are separated from each other by intergenic spacers (Fig. 1; Procunier and Tartof, 1975; Long and Dawid, 1980; Rogers and Bendich, 1987). The genes encoding the small ribosomal subunit (SSU, 18S), the 5.8S rRNA and the large ribosomal subunit (LSU, 25S or 26S in plants, 28S in animals) are separated by two internal transcribed spacers (ITS 1 and 2) and constitute one transcription unit. Each unit is flanked by 5′- and 3′-external transcribed spacers (ETS), as part of the intergenic spacer (IGS). The middle part of the IGS between the 3′- and 5′-ETS is not transcribed and occasionally referred to as non-transcribed spacer (NTS). The rDNA arrays often appear as a subtelomeric secondary constriction (Goodpasture and Bloom, 1975; Schubert and Künzel, 1990).

* Corresponding authors. Address: Institute for Evolution and Biodiversity, University of Muenster, Huefferstr. 1, D-48149 Muenster, Germany (S. Wicke).
E-mail addresses: susann.wicke@uni-muenster.de (S. Wicke), quandt@uni-bonn.de (D. Quandt).

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.06.023

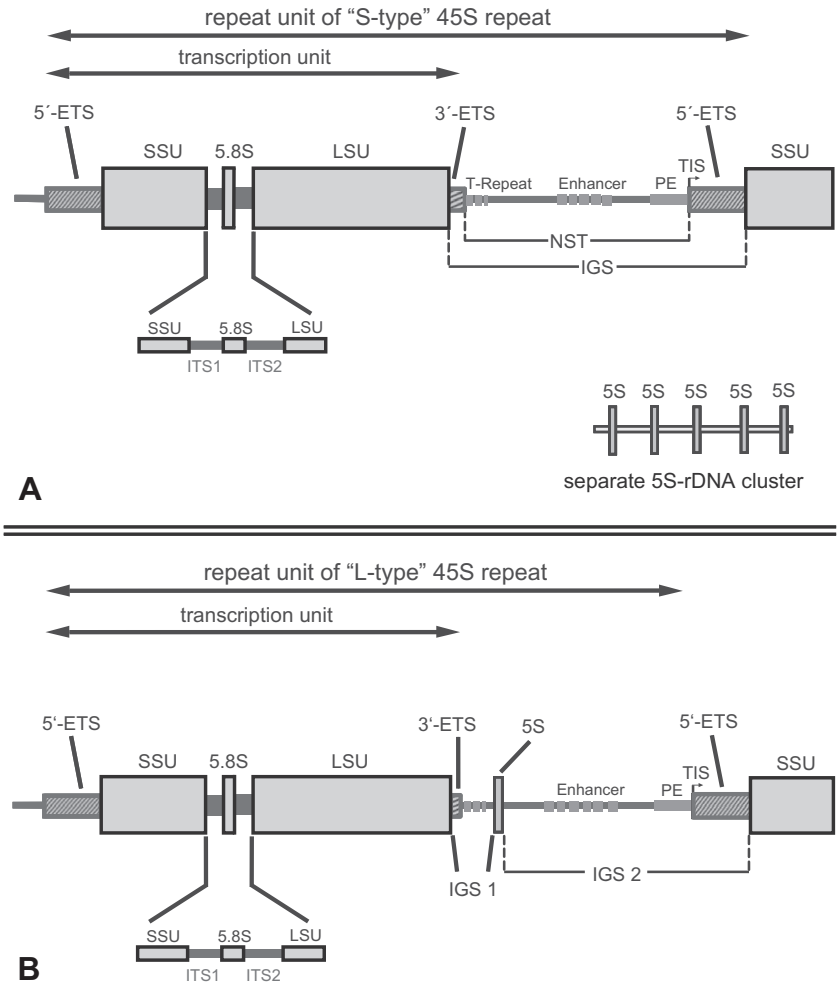


Fig. 1. Organization of the nuclear ribosomal DNA. Typically, in both S- and L-type organization, one transcription unit comprises the 5′-external transcribed spacer (5′-ETS), the gene for the small ribosomal subunit (SSU), the internal transcribed spacers 1 and 2 (ITS1 and ITS2), the gene for the 5.8S rRNA, and the gene for the large ribosomal subunit (LSU) as well as its flanking 3′-external transcribed spacer region (3′-ETS). Together with the non-transcribed intergenic spacer (NTS), these elements constitute one repeat unit. Several transcriptional regulatory elements may be situated in the intergenic spacer (IGS). In S-type lineages the 5S rDNA is localized in a separate cluster. In L-type lineages, the 5S rDNA is localized within the NTS of the 45S repeat, splitting the IGS into IGS1 and 2 [modified after Quandt and Stech (2002)]. PE – promoter element; TIS – transcription initiation site.

The chromosomal location of the region encoding the 5S rRNA is variable in both prokaryotic and eukaryotic organisms. In higher plants and animals, the 5S rRNA genes often form separate clusters. The 5S rDNA array may be found at more than one locus either on the same chromosome as the 45S repeats or scattered across the genome (Pardue et al., 1973; Goldsbrough and Cullis, 1981; Schneeberger et al., 1989 Sastri et al., 1992). In plants, the copy number of 5S RNA genes may vary from 2000 to 75,000 (Appels et al., 1980; Goldsbrough et al., 1982; Vakhitov et al., 1986).

With the availability of genome sequences and an increased interest in the rRNA genes of crops and various model plants (e.g. *Arabidopsis thaliana*, *Zea mays*, *Brassica rapa*), the arrangement and organization of the rRNA genes into a 45S repeat and a separate 5S cluster was widely accepted for angiosperms (Hemleben et al., 1988; Gruendler et al., 1991; Rocha and Bertrand, 1995). Thereafter, this organization type of nrDNA, denominated as "S-type" organization herein ("S" for separation), was generally adopted for all major land plant lineages.

However, a different arrangement has been reported in various distantly related eukaryote lineages such as cryptomonad algae, fungi and protozoans (Rubin and Sulston, 1973; Gerbi, 1986; Gilson et al., 1995). In this "L-type"-organization ("L" for linkage) the 5S rRNA gene is embedded in the NTS of the 45S repeat unit (Fig. 1). In such cases, the 5S rRNA is mostly transcribed on the same strand as the 45S transcription unit. In some organisms, however, 5S rRNA genes are encoded on the opposite strand (Drouin and de Sa, 1995; Garcia et al., 2009).

Both "S" and "L" type nrDNA organization types are randomly scattered across the tree of eukaryotes (Fig. 2; Long and Dawid, 1980; Rogers and Bendich, 1987; Zentgraf et al., 1998; Drouin et al., 1992; Cruces et al., 1989; Kawai et al., 1997), not allowing any conclusion or assumptions concerning evolutionary constraints or fitness. Both types can even coexist in some groups (e.g. arthropods, ascomycetes, alveolates, etc.). However, generally only one or two species were studied on behalf of an entire phylum. For green plants, including the chlorophyte algae lineages, angiosperms are commonly chosen as representatives. A close look at the land plant phylogeny (Fig. 2) strikingly shows that, in contrast to the copious data concerning the nrDNA organization in angiosperms, only limited data is available for early diverging Streptophytes. So far, only two bryophyte species have been studied: the liverwort *Marchantia polymorpha* and the moss *Funaria hygrometrica*. For these taxa, a co-localization of all rRNA genes was reported (Sone et al., 1999). As bryophytes were traditionally considered a monophyletic group

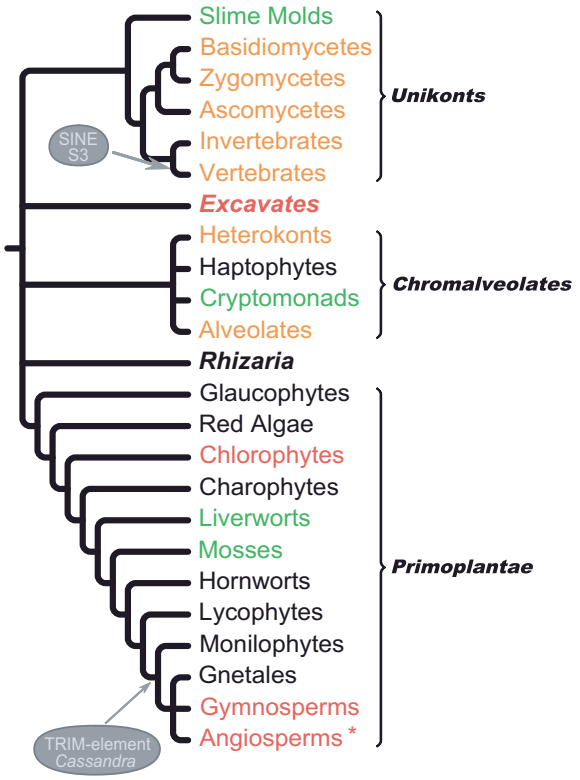


Fig. 2. Overview of the rDNA organization in eukaryotic lineages prior to this study. Co-localization of all nuclear ribosomal genes – disregarding strand localization – appear in fungal lineages as well as cryptomonads, chromists, the liverwort *Marchantia polymorpha* and the moss *Funaria hygrometrica*. Separate clusters of 45S and 5S are validated for vertebrates, some chlorophytes, gymnosperms and angiosperms. In several lineages including invertebrates, excavates and alveolates, both types were detected. Eukaryotic kingdoms are indicated in bold italics. Asterisks mark a secondary re-insertion of 5S rDNA into the IGS of the 45S repeat. The occurrence of 5S-derived transposable elements is indicated by arrows and the name of the respective element [topology based upon the "Tree of Life" by Palmer et al. (2004) maintaining its systematics and nomenclature].

sister to tracheophytes (Mishler and Churchill, 1984; Garbary et al., 1993) a 5S rDNA insertion was postulated that occurred in an ancestral bryophyte lineage after the emergence of vascular plants (Sone et al., 1999). At the same time, morphological as well as advancing molecular studies indicated that this traditional view of land plant evolution is rather controversial, given the fact that almost all possible topologies of the early land plant tree had been published (Shaw and Renzaglia, 2004). However, during the past 5 years evidence for the paraphyly of bryophytes accumulated, i.e. bryophytes constitute a grade with liverworts branching off first and hornworts being sister group to vascular plants (Qiu et al., 2006). Under these circumstances the scenario postulated by Sone et al. (1999) appears quite unlikely as it implies several independent 5S rDNA insertion events.

Thus, we propose that the physical linkage of all rRNA genes within one cluster is a general feature of early land plants. Accordingly, the relocation of 5S rRNA including the formation of a separate and distinct 5S rDNA cluster appears to have occurred later during the evolution of land plants, perhaps in the ancestor of vascular plants. We further assume that the 5S rDNA transposition event caused structural and/or transcriptional alterations in the ribosomal rDNA/rRNA of early land plants. Therefore, we investigated the "jumping 5S"-phenomenon and the molecular evolution of the nuclear ribosomal DNA of Streptophytes based on a representative sampling of land plants and putative green algal ancestors.

2. Material and methods

2.1. Taxon sampling and material

Sampling was guided by the most recent and comprehensive phylogenetic analysis of land plants provided by Qiu and colleagues (2006). A minimum of two representatives for each major lineage sensu Qiu et al. (2006) was chosen. Charophyte and chlorophyte algae species were included to provide inference of the ancestral condition (Lewis and McCourt, 2004). Representatives of each bryophyte clade (Marchantiophyta, Bryophyta and Anthocerotophyta) were chosen according to Shaw and Renzaglia (2004). The sampling of lycophytes and monilophytes follows recent studies on the phylogeny of extant ferns (Pryer et al., 2001; Schneider et al., 2004; Smith et al., 2006). The selection of (basal) seed plants (Gnetales, gymnosperms, angiosperms) followed Qiu and colleagues (2006). The analysis of angiosperms was restricted to ten species (mainly model monocots and eudicots) already available at NCBI Genbank.

2.2. Amplification, cloning, and sequencing

DNA was isolated from 0.5 to 3 g fresh plant material using a modified CTAB method after Doyle and Doyle (1990). PCR was carried out using several PCR systems [Eppendorf *Taq*-polymerase, Eppendorf HotStart *Taq* Polymerase, Eppendorf *Triple Master Taq* Polymerase, *Qiagen* Long-range PCR System, peQlab SAWADY *Taq* Polymerase]. PCR was set up following the manufacturers' protocols and modified specifically according to each species requirements. Various sets of primers anchored within the LSU, 5S and SSU coding regions were designed to amplify the intergenic spacer region(s) of both 35/45S repeats and the non-transcribed spacer of the 5S rDNA cluster (summarized in Supplementary material). A detailed description of PCR conditions and accession numbers for each species/lineages is provided in Supplementary material.

PCR products were gel purified employing the Nucleo Spin Extract II kit (Macherey–Nagel). Subsequent cloning was carried out using the pGEM®-T Easy Vector Systems (Promega) or TOPO® XL PCR Cloning Kit (Invitrogen) following the manufacturers' instructions. Plasmid isolation was performed via alkaline lysis (Sambrook and Russell, 2001). Plasmids were sequenced bi-directionally and via primer-walking on Beckman-Coulter CEQ 8000 or ABI 3700 Prisma, respectively. Alternatively, plasmids were sequenced at Macrogen Ltd., South Korea or at the Centro de Sequenciación, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid, Spain.

2.3. Southern hybridization

Sequence information was double checked using Southern Blot hybridization. RNase treated DNA was digested with *BamHI* (Promega). The digestion was separated in 0.7% agarose gels and DNA transferred to a positively charged nylon membrane; membranes were baked at 80 °C for 30 min and subsequently UV-crosslinked for 2 min (Sambrook and Russell, 2001). Probing for the 45S repeat was carried out using a LSU-probe amplified from *Lycopodium phlegmaria* (partial LSU-fragment covering 1500 bp towards the 3′-end of the LSU gene). 5S rDNA was hybridized with a probe representing the entire coding sequence of the 5S-gene that was amplified from the fern *Phlebodium aureum*. Both PCR-fragments were cloned into the pGEM-TEasy vector and sequenced prior to use. Labeling was performed using Biotin-11-dUTP (Fermentas) in a PCR-reaction using T7promoter and SP6 as amplification primers, 1 mM dCTP, dGTP, dATP and 0.4 mM dTTP/0.6 mM

biotin-dUTP, and 1.5 U *Taq*-polymerase. Labeled probes were purified employing the Nucleo Spin Extract II Kit (Macherey–Nagel) according to the manufacturer's instructions. Filters were hybridized following the protocol and buffer compositions described by Sambrook and Russell (2001). Both pre-hybridization and hybridization were carried out at 60 °C overnight. Membranes were washed twice for 10 min at 60° using pre-warmed 2× SSC + 0.1% SDS-buffer. Bound biotin-labeled probes were detected employing the Biotin Chromogenic Detection Kit (Fermentas) according to the manufacturer's instructions.

2.4. Sequence data analysis

DNA sequences were edited manually using the Phylogenetic Data Editor (PhyDE®, www.phyde.de). Sequences were cross-checked via BLASTn (local Blast-search and via http://www.ncbi.nlm.nih.gov/BLAST, www.COSMOSS.org and DOE-Joint Genome Institute). Sequences were manually aligned (as far as possible) and compared to closely related species via group specific substitution patterns in the genes for ribosomal LSU, 5S and SSU. Sequence data and datasets have been deposited at the EMBL sequence database; accession numbers are supplied in Supplementary material. Furthermore, single nrDNA- and 5S-rDNA contigs or scaffolds of the draft genomes from whole-genome sequencing projects of *Physcomitrella patens* (Rensing et al., 2008), *Selaginella moellendorffii* (JGI draft sequence 1.0; Banks et al., 2011), *Chlamydomonas rheinhardii* (Merchant et al., 2007), *Chlorella vulgaris* C169 (JGI draft sequence 1.0), *Oryza sativa* (http://www.plantgdb.org/OsGDB), *A. thaliana* were incorporated into the dataset (The Arabidopsis Genome Initiative, 2000).

The secondary structure of the 5S rRNA and some spacer parts of the 45S repeat as well as structure plots were analyzed employing MFold at the Rensselaer bioinformatics web server (Zuker, 2003). Calculations were performed for an example of each lineage using constraint information such as experimentally proven helix structures and bulges according to models of Luehrsen and Fox (1981) and Barciszewska et al. (1994), the latter of which includes information on putative tertiary interactions as well.

3. Results

3.1. Sequence analyses of land plant nrDNAs

3.1.1. LSU–5S spacer (IGS1)

PCR-products spanning a putative LSU–5S spacer (IGS1) could be easily generated under standard conditions for the selected algal representatives *Pseudendoclonium basiliense*, *C. vulgaris* (both Chlorophytes), *Klebsormidium flaccidum* and *Chara vulgaris* (Charophytes) and most early diverging land plants (Table 1; Supplementary material) employing the primer combination IGS1F1 (situated ~80 bp from the LSU 3′-terminus) and IGS1R1 (situated 72 bp into the 5S 5′-coding region; Supplementary material). Thus, using this primer combination, verification of the physical linkage of 5S rRNA genes to the 45S rDNA repeat units could be adduced for algae, liverworts, mosses, most "eusporangiate" and most leptosporangiate ferns (Table 1, Fig. 3). However, within liverworts (*Bucegia romanica*, *Pellia epiphylla*, *Jamesoniella autumnalis*, and *Plagiochila adiantoides*) this amplification strategy often resulted in multiple PCR products. Apart from the desired liverwort IGS1 sequences, cloning and sequencing of the respective amplicons often resulted in various unspecific or fungal sequences as identified via BLASTn. This result is in line with the frequent fungal associations (mycothallus) observed in liverworts. Amplification attempts of a putative LSU–5S spacer in hornworts yielded several products containing the hornwort 5S coding sequence (CDS), but no trace of the LSU could be detected in the surrounding sequence parts. However, Southern Blot hybridization indicated the L-type organization. Amplification of the IGS1 for lycophytes yielded numerous truncated sequences lacking either 5S rDNA or the ribosomal LSU gene. Co-localization, however, was deduced analyzing the available genome data of *Selaginella moellendorfii* (version 1.0). Most fern (monilophyte) lineages, especially the early diverging lines ("eusporangiate" ferns) and early diverging leptosporangiates possess very large genomes that are potentially coinciding with numerous (intact, silenced or pseudogenized) rDNA loci. Combination of different amplification primers eventually resulted in distinct IGS1 sequences for: whisk, ophioglossid and marattiod ferns, horsetails and nearly all leptosporangiate ferns, showing a co-localization of rDNA elements for these taxa. IGS1 amplifications for gymnosperm (*Cycas*, *Gingko*, *Gnetum*) and basal angiosperms (e.g. *Amborella*, *Nymphaea*, *Victoria*) did not yield any PCR products, irrespective of the employed primer combination. Amplification of the LSU–SSU intergenic spacer however, were successful for *Gingko*, *Gnetum* and *Victoria*. All sequences derived from these cloned PCR products lacked a 5S RNA coding gene region. Structure and composition of the IGS region of those taxa is similar to that of previously described angiosperm rDNA intergenic spacer regions (e.g. Hemleben et al., 1988).

3.1.2. 5S cluster

Amplification of the 5S rDNA cluster using the different 5S primer combinations yielded unspecific or no amplicons for algae, bryophytes, lycophytes, and ferns, excepting the heterosporous (water) ferns (*Marsilea quadrifolia*, *Salvinia oblongifolia*). In several cases, PCR products obtained from bryophytes and early diverging tracheophytes were either identified as 5′- and 3′-truncated gene fragments consisting solely of the conserved 5S promoter boxes or, in the case of the derived leptosporangiate ferns (tree ferns and polypod ferns), as 5S rDNA derived *Cassandra* retroelements. In water ferns, however, a true 5S rDNA cluster could be amplified and verified by sequencing. The obtained 5S sequences showed the typical base composition pattern characteristic for the majority of ferns with L type arrays. Sequences of the 5S rDNA cluster were obtained easily for gymnosperms and basal angiosperms. In addition, tandemly arrayed 5S rRNA genes could also be amplified for hornworts. However, no more than two genes in a row were detected in *Anthoceros*. And in all cases, one of the two 5S rRNA genes showed a deletion of seven base pairs within the CDS (directly upstream of promoter box A), although overall sequence divergence was low (in terms of nucleotide substitutions and/or insertion/deletions). One clone (out of 26) contained a duplication of the 5S gene promoter box C. Similarly, screening for putative 5S clusters in *Lycopodium*, *Selaginella* and *Isoetes* only yielded 5S rRNA gene fragments of unknown genomic localization. In all cases, short (~60–70 bp) gene fragments corresponded to 5S rDNA domains β and γ; domain α, however, could not be detected.

3.1.3. Secondary structures of 5S rDNA and Cassandra retroelements

In contrast to most other land plant lineages, 5S sequences from the Gnetales (*Gnetum gnemon* and *Gnetum urens*) and lycophytes, in particular those from *Selaginella*, displayed a high number of nucleotide substitutions within the 5S CDS. However, stable secondary structures according to previous structure models after Luehrsen and Fox (1981), and Barciszewska et al. (1994; Fig. 4; Supplementary material) could be inferred due to compensating base pair changes (CBC) suggesting that functionality is not impaired. Structure models for bryophyte, lycophyte and monilophyte 5S rDNA including observed nucleotide substitutions within the coding region are provided in Supplementary material. The open ("exposed") sides, that were thought to interact closely with each other (loop D and C) as well as the transcription factor

Table 1
Summary of the nrDNA organization types in Streptophytes and Chlorophytes as revealed by this study. '×' indicates the presence of a certain organization type (L-type, S-type); absence is indicated by '–' (abbr.: WGS - whole genome draft sequence, SB – Southern Blotting).

Lineage	L-type	S-type	Remarks
Chlorophyte algae	×	×	– L-type: *Chlorella vulgaris*, *Pseudendoclonium basiliense* – S-type: *Tetraselmis chuii*, *Chlamydomonas rheinhardtii* (Marco and Rochaix, 1980) – No 5S rDNA cluster found by Blast search against the *Chlorella* WGS; – no evidence of additional 5S rDNA genes being inserted in the IGS of the 45S repeat in the *Chlamydomonas* as identified by Blast search against WGS
Streptophyte algae	×	–	– No additional 5S cluster detected by PCR – Linkage confirmed by sequencing (*Klebsormidium*) and SB of *Cosmarium*, *Chara*
Liverworts	×	–	– No additional 5S cluster detected by PCR – Linkage confirmed by sequencing and SB
Mosses	×	–	– No additional 5S cluster detected by PCR – Physical linkage confirmed by sequencing and SB – No 5S cluster detected by Blast search against the *Physcomitrella* WGS
Hornworts	×	–	– 5S linkage confirmed by Southern Blot, yet no sequence evidence of physical linkage, may indicated that 5S rDNA is encoded on the opposite strand
Lycophytes	×	–	– Physical linkage confirmed by Southern Blotting and Blast-Searches against *Selaginella moellendorffii* WGS – No 5S cluster or other tandemly arrayed 5S rDNA genes detected by PCR screens – Numerous truncated 5S like elements scattered across the genome
Monilophytes	×	×	– L-type: 'eusporangiate' + most leptosporangiate ferns; S-type: heterosporous (water) ferns – The heterosporous water fern genus *Salvinia* might show a transitional state as suggested by SB-results – *Cassandra* retrotransposons only found in leptosporangiate ferns
Gymnosperms	–	×	– S-type identified by PCR screens and SB and available FISH-studies (e.g. Liu et al., 2003)
Angiosperms	–	×	– S-type ancestral as identified by PCR screens and SB – Re-insertion of 5S rDNA into the IGS of the 45S repeat in the asterid genus *Artemisia* (Garcia et al., 2007, 2009)

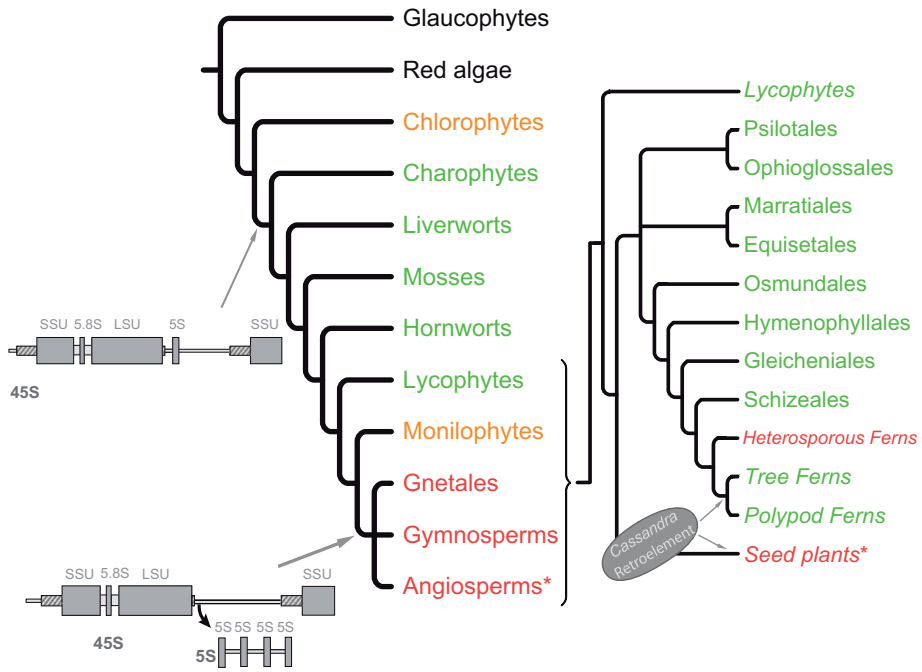
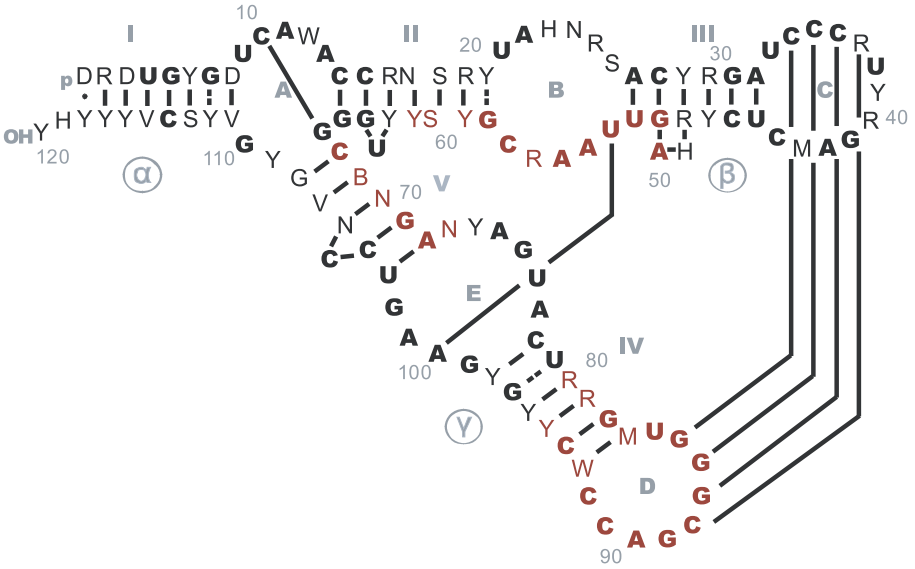


Fig. 3. Summary of nrDNA arrangements in land plants. A physical linkage of all rRNA genes is present in early land plants such as liverworts, mosses, hornworts, lycophytes and most of the fern lineages. Separate arrays for SSU-5.8S-LSU-repeat and 5S genes can be detected in seed plants and heterosporous (water) ferns. A change from L- to S-type organization was observed in monilophytes (water ferns) and chlorophyte algae (*Chlamydomonas*). Secondary re-insertion of 5S is highlighted by an asterisk. Arrows indicate the occurrence of the 5S derived transposon element *Cassandra* in some vascular plants (polypod ferns, seed plants) [land plant topology follows Qiu et al. (2006)].

IIIA (TFIIIA) binding loop E showed a very low substitution rate (Fig. 4). Nucleotide substitutions as well as CBCs were domain specific and reflected previous findings on functional 5S rRNA domains. Domain α accounted for most base pair changes within land plants, most of which are compensatory. Substitution rates in the domains β and γ were significantly lower and a minority of non-complementary base pair changes (72% CBC in domain β, 76% CBC in domain γ) was observed. Positions that were known to interact with reactive ribosomal centers were highly conserved within all examined species. Mutations within functional domains, however, occurred within the duplicated 5S gene (Ψ5S) found in *Equisetum hyemale* and in an aberrant copy amplified from *Gleichenia dicarpa* (clone F975SR1CC6); those 5S copies were therefore considered as pseudogenes.

A retroelement of the TRIM-class named *Cassandra* that harbors short sequence elements derived from 5S rDNA was detected



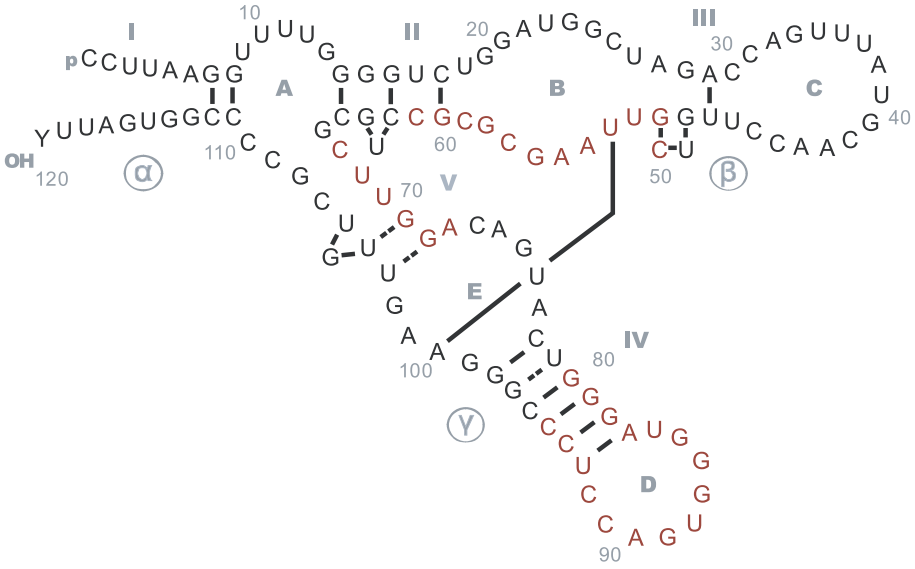


Fig. 4. Secondary structure models for land plant 5S rRNA and the *Cassandra* 5S-like element. Although nucleotide substitutions are seen frequently in some domains of the 5S rRNA, helical and loop structures remain highly conserved within land plants which is mainly due to compensatory base pair changes (CBCs). CBCs were not observed in *Cassandra* 5S-like elements, impairing the formation of stable helices in domains α and β and thus suggesting non-functionality of such elements. Dark gray/red letters highlight 5S promoter boxes in both the 5S rRNA and Cassandra-5S-like element, and nucleotide positions where no substitution was found within land plants are indicated by bold letters. Consensus models for single lineages are supplied in Supplementary material after Luehrsen and Fox (1981) and Barciszewska et al. (1994).

within the species-rich core leptosporangiate ferns (polypod ferns). *Cassandra* could be amplified and identified using primers nested within the C-Box of the 5S intragenic promoter (Figs. 3 and 4, Supplementary material). The autonomous transposable element *Cassandra* is well known from seed plants, and has already been described for two core-leptosporangiate ferns before (Kalendar et al., 2008). However, *Cassandra* elements could not be detected in other groups, such as early diverging fern lineages incl. "eusporangiate", early diverging leptosporangiate ferns and water ferns (Fig. 3), either by PCR-screening or southern hybridization.

3.1.4. IGS organization and variation and transcription regulating elements

Sequence analyses of the IGS showed a conserved structure for all land plants, while exhibiting a high diversity with regard to length and GC-content that does not correlate with organizational level. The average length was around 4–5 kb with an average GC content of ~60%. A general trend of accumulating species-specific GC-rich repetitive sequence motifs was observed. The number and lengths of these repeat elements were variable and largely accounted for the extreme IGS-length heterogeneity (1812 nucleotides (nt) in *Gingko biloba* versus 9207 nt in *M. polymorpha*, including the 5S CDS. Two small hairpin structures could be identified directly downstream of the LSU that were present throughout all analyzed taxa. The detected hairpins were surrounded by short sequence motifs that displayed a high similarity to previously reported (putative) termination and processing signals. The LSU–5S intergenic spacer (IGS1) of co-localized lineages often contained short sequence stretches of 5S intragenic regulatory regions derived from the polymerase III (*pol*III) promoter. In rare cases, box

A and C or parts of it were duplicated or repeated with only few nucleotide substitutions, resulting in multiple misprimed PCR-products. In contrast to *K. flaccidum* and *C. vulgaris*, *pol*III promoting signals described so far (such as the TATA-box at approx. −30), could not be detected in early land plant lineages and most monilophytes, regardless of the 5S localization type. The occurrence and prediction of TATA-like elements for 5S rDNA in early lineages is highly speculative as relevant sequences appear to be highly divergent and do not show any similarity to previously reported 5S transcription-regulating motifs. In few cases, slightly different motifs could be identified within a correct distance (e.g. GATA in *Todea*, TTTA in *Angiopteris* and *Gleichenia*, CATA at −34 or TACA at −32 in *Lygodium*, CATA at −32 or TATG at −30 in *Vandenboschia*), however, their functionality remains unclear. The transcription termination signal (i.e. long poly-T stretch shortly downstream of the 5S rRNA gene), however, was identified in all lineages, although with occasional modifications, yet maintaining a polypyrimidine stretch. In contrast to available data on angiosperms, neither a TATA-element nor other transcription signaling motifs could be identified in the 5S–5S intergenic spacer in an appropriate distance to the 5S transcription initiation site for water ferns, gymnosperms and basal angiosperms. Termination signals were, however, very prominent in all lineages. IGS/NTS sequences of the horsetail *E. hyemale* revealed a duplication of the 5S rRNA gene in the spacer region and the insertion of three different tRNAs (*trnC*, *trnG*, *trnY*) as well as a putative tRNA pseudogene (Ψ*trnC*). The first 5S rRNA gene is located 491 bp downstream of the LSU and appears as a functional version, whereas the second gene copy displays several nucleotide substitutions suggesting a pseudogenization (Ψ5S). A duplication of 5S rRNA genes within the intergenic spacer of the 45S repeat has been reported previously, yet a co-localization of several tRNAs and 5S rDNA inside the 45S repeat has not been found so far. The 5′-flanking region of the second 5S rDNA copy lacks a "TATA-box" and other regulatory elements for 5S transcription. However a poly-A stretch (5′-AAAAAAGGAAA-3′) is present only a few nucleotides upstream of the 5S genic regions as well as several short repetitive motifs, both typical indicators of transposon activity. Organization of the IGS2 was nearly identical to the IGS of S-Type lineages. AT-rich stretches were often detected upstream of putative promoter elements or promoter repeats. In early land plant lineages and gymnosperms, however, the specific ribosomal initiator (rInr), that has been inferred for a few angiosperms (consensus sequence for angiosperms: TAT KWR GGG) could not be identified. Instead, a similar motif could be predicted (e.g. CACACGGGGTG for *G. gnemon*) as a putative rInr. These motifs were either embedded into large repeat elements (putative promoter elements) or enhancer-like short repeat stretches similar to those of *A. thaliana*. Putative promoter motifs have been reported previously for *F. hygrometrica* and *M. polymorpha* (TATGTGGGGG, GATAGGGGG; Capesius, 1997; Sone et al., 1999) which are similar to angiosperm motifs. However, experimental data to prove the 45S and 5S transcription initiation for early land plant lineages is missing.

3.2. Southern Blot hybridizations

A complicated structure of the 45S IGS or a localization of the 5S rRNA genes on the opposite strand might hamper adducing sequence evidence. In order to identify the state of rDNA organization, Southern hybridization was performed. For *C. vulgaris*, *Cosmarium botrytis*, *P. basiliense*, as well as most early diverging Streptophytes (including hornworts, lycophytes and whisk ferns), both LSU and 5S probes annealed to the same fragments, indicating L-type organization and thus confirming the results from PCR screens. However, additional binding of 5S rDNA to two short fragments could be detected for *Anthoceros* and *Equisetum* verifying the 5S rRNA gene duplication found by sequencing. No ladder-like patterns typical for repetitive DNA were detected, implying the absence of a 5S rDNA repeat cluster. In angiosperms, however, *Bam*HI ladder-like restriction patterns are often caused by 5S rDNA methylation. Absence of such a pattern due to the lack of differential methylations in 5S loci, thus, might not necessarily indicate absence of a separate 5S rDNA cluster. LSU hybridization yielded numerous fragments in *Mesostigma*, lycophytes and monilophytes. This might point towards the presence of several rDNA loci or heterogeneous repeat units in the genome. In *Mesostigma*, the LSU probe detected many short fragments (<0.5 kb) which seemed to be of equal size to the two faint 5S rDNA bands. *Tetraselmis chuii*, seed plants and heterosporous (water) ferns did not show a hybridization of both probes to the same fragments. The LSU probe was found to bind to rather big fragments (~3–5 kb in size). The 5S probe, however, annealed to much shorter fragments resulting in ladder-like patterns indicative of repetitive units. This confirmed the separation of rRNA genes into two clusters. In the water fern *Salvinia*, the 5S rDNA probe also annealed to the identified LSU fragment, suggesting either the presence of both organization types or the presence of a 5S rDNA-like gene in close proximity to the large ribosomal subunit. No such pattern could be found in any seed plant tested.

3.3. BLAST analyses of whole-genome data

Due to the current lack of whole-genome data for Chlorophytes and early diverging Streptophytes, BLASTn searches on non-seed plant whole genome data were limited to *P. patens*, *C. vulgaris*, *C. rheinhardii* and *S. moellendorfii*. BLASTn searches of the obtained IGS1 sequences from *C. vulgaris* against the available whole-genome data yielded hits in several scaffolds. Further analyses of these scaffolds revealed multiple 45S repeats in the *Chlorella* genome with the 5S rRNA genes embedded into the IGS in all cases. No such linkage was found in *Chlamydomonas* scaffolds. BLAST-searches against the shotgun reads and assembled contigs/scaffolds of *P. patens* confirmed the absence of distinct 5S rDNA arrays. Analysis of whole genome sequences from the lycophytes *S. moellendorfii* (version 1.0) confirmed the physical linkage of all rRNA genes. A total of eight scaffolds showed maximum identity and physical linkage of 5S rDNA to LSU and SSU. Furthermore, several scaffolds were detected carrying partial 5S rRNA genes. Most frequently, only the 3′-genic region (surrounding the internal promoter sequences of 5S rDNA) was present.

4. Discussion

4.1. Organization and mobility of nrDNA in land plants

In contrast to previous ideas early diverging Streptophytes, including the algal lineages ancestral to land plants and the paraphyletic bryophytes (Qiu et al., 2006), share a co-localization of all rDNA components within one 45S repeat unit (Table 1, Fig. 3). This is also true for the basal tracheophytes such as lycophytes and monilophytes, exclusive of heterosporous (water) ferns. In seed plants and, surprisingly, water ferns, the gene encoding the 5S rRNA forms a separate cluster of tandemly arranged 5S rRNA genes. These transposition events, however, occurred apparently independently. Clearly, the formation of the 5S rDNA cluster as discovered in gymnosperms and angiosperms must have taken place in their common ancestor as can be concluded from the presence of the S-type organization in early diverging seed plants such as cycads and *Gingko* as well as *Victoria and Amborella*. In contrast, the S-type organization present in heterosporous (water) ferns is rather unexpected. There is no evidence of close relationships of

both groups either from paleobotanical data or from molecular phylogenetics and/or morphology. Thus, the most likely explanation is an independent formation of the separate 5S array after the divergence of water ferns from the remaining monilophyte lineages. The presumably synapomorphic S-type organization in seed plants, however, has been demonstrated unambiguously in previous works on Solanaceae (Borisjuk et al., 1997; Volkov et al., 2003; Komarova et al., 2004), Cucurbitaceae (Zentgraf et al., 1990; King et al., 1993), Fabaceae (Schiebel et al., 1989; Kato et al., 1990; Abirached-Darmency et al., 2005) and is further corroborated by whole-genome sequencing projects on *Oryza*, *Carica*, *Arabidopsis*, etc. In addition, the "mobility" of 5S rDNA repeats is independently inferred by the presence of both organization types among Chlorophytes. S-type organization has been reported for *C. rheinhardii* (Chlorophyceae; Marco and Rochaix, 1980) and *T. chuii* (Prasinophyceae), while L-type organization was observed for *C. vulgaris* (Trebouxiophyceae) and *P. basiliense* (Ulvophyceae) in the present study; this distribution does not allow us to infer the ancestral nrDNA organization state. The co-localization of all four rRNA coding regions is, however, shared by all streptophyte algae studied here (*Chara*, *Klebsormidium*, *Cosmarium*) apart from *Mesostigma* where the current data is not unambiguous. Thus, it seems plausible that the insertion of the 5S rDNA into the 45S rDNA array occurred already in the ancestral Streptophyte lineage before the emergence of land plants. Due to lacking evidence from Chlorokybophyceae, Coleochaetophyceae, and uncertainties concerning the nrDNA organization in *Mesostigma*, however, it is not certain, that the physical linkage of all rRNA genes is ancestral to *all* Streptophyte lineages. Our results suggest that similarly to land plants, the locus of the 5S rRNA gene changes several times independently in chlorophyte green algae (e.g. separate clusters in *Tetraselmis*, *Chlamydomonas*, co-localization in *Chlorella*, *Pseudendoclonium*). Therefore, the herein investigated taxa are not sufficient to uncover patterns of green algal rDNA-organization. A much deeper sampling of, in particular, chlorophyte algae, will be necessary.

In contrast to seed plants, the additional presence of co-localized 5S rDNA could not be ruled out for water ferns, as Southern hybridization indicated the presence of both organizational types – perhaps, one of which may be coding whereas the other may be pseudogenic. Based on the current phylogeny of land plants (Qiu et al., 2006), and if the L-type organization represents the ancestral state, the presence of S-type organization in water ferns and seed plants clearly suggests two independent 5S rDNA translocation events. Interestingly, the reversal, i.e. re-insertion of 5S rDNA into the 45S array has recently been reported for the angiosperm genus *Artemisia* (Asteraceae) by Garcia et al. (2007, 2009). The insertion of two tandemly arranged and putatively active 5S rRNA genes into the IGS implies has been shown to coincide with the absence of 5S rDNA clusters. Apart from the ambivalent results obtained for water ferns, we found no indication of isolated or retired arrays of 5S rDNA in any other L-type species, which is concordant with previous findings (Sone et al., 1999; Garcia et al., 2009). It can be assumed that the L-type organization in that case eliminates a separate nuclear 5S rDNA locus. This observation is further corroborated by analysis of whole-genome sequence data from *P. patens*, *S. moellendorfii* and *Chlorella* sp. NC64A, where 5S rDNA only occurs in the intergenic spacer of 45S-repeat units but could not be found elsewhere in the genome. The independent switches from S- to L-type in distinct groups together with its reversal in *Artemisia* reasonably implies that more 5S rDNA relocations might have occurred during the evolution of green plants, than is reported here. It is also conceivable that transition forms still exist, i.e. organisms where both organization types co-exist within a species. In fact, randomly dispersed 5S rDNA genes occurring as single copies (i.e. not arranged tandemly) have been reported from various eukaryote genomes, e.g. those of fungi (Morzycka-Wroblewska et al., 1985; Bergeron and Drouin, 2008; reviewed in Dujon, 2010), and vertebrates (rat and mouse: Reddy et al., 1986; Drouin, 2000). However, the presence of both a functional 5S rDNA cluster and a 45S rDNA cluster containing all four rRNA genes within one transcription unit has not yet been detected there.

What is the advantage of (physically) linking 5S rDNA to the large ribosomal subunits or formation of a separate 5S cluster, respectively? There are several reports from various organisms where variant 5S gene copies are found, and that these copies obviously evolve differently (Piper et al., 1984; Brow, 1987; Leah et al., 1990; Cloix et al., 2000). Some 5S rDNA repeats have been suggested to lack maintenance by concerted evolution or are subjected to biased or incomplete homogenization (e.g. Ganley and Scott, 2002). The general structure of the IGS is composed of several types of repeat units of different length (that mainly contribute to the extreme IGS length variation) in both S- and L-type organisms. One class of repeats generally occurs shortly downstream (or as part of) the 3′-ETS region and has therefore been denominated as a termination repeat (Schiebel et al., 1989; Moss and Stefanovsky, 1995; Paule and White, 2000). Another class of repeat units can be seen upstream of the 5′-ETS, surrounding the promoter element(s). Some of these elements have been shown to act as enhancers or contribute to read-through enhancement (Mitchelson and Moss, 1987; Moss et al., 1992; Moss and Stefanovsky, 1995; Borisjuk et al., 1997; Paule and White, 2000; Volkov et al., 2003). In L-type species, this second repeat class is localized downstream of the 5S rRNA gene. Repeat units, in general, are usually more developed in the IGS compared to the 5S rDNA spacers that harbor only short or no repeat motifs. These former IGS repeats, however, are considered to facilitate a more effective and more stable homogenization of 45S rDNA via e.g. homologous recombination leading to concerted evolution. Altogether, this might explain the preferred 5S-linkage to larger gene families. It is known that the overall repeat-size (which includes the total size of each higher order repeat unit, e.g. a 45S transcription unit) affects the homogenization efficiency and rate (Amstutz et al., 1985; Watt et al., 1985; Jinks-Robertson et al., 1993; Cabral-de-Mello et al., 2010). Furthermore, the number of 5S gene copies would be regulated and more or less equivalent to the number of the remainder rRNA species. In addition, the localization of all rRNA coding regions within the nucleolus-organizing region (NOR) during active transcription stages might better foster the assembly of ribosomes.

On the other hand, co-localization might, however, sterically interfere with possible read-through enhancements and polymerase recycling (of both *pol*I and *pol*III) during ribosomal transcription, which would – in return – favor the separation of rDNA clusters. Thus, mechanisms promoting concerted evolution might have been factors in the re-location of 5S rDNA repeat unit (Fig. 5). The initial re-location might have been an accidental deletion from, or insertion into, one (out of many) repeat units of the 45S array. Considering the effect of homogenization of a gene family, this mutation (loss, insertion) will most likely be erased by CE within a few generations – or, alternatively, it is spread. In any case, the more or less concerted re-location of 5S rDNA might have occurred in a common ancestor that did possess only a *few* rDNA repeat units which would favor quick spreading. Interestingly, the occurrence of a 5S rDNA-derived transposon is not unique to land plants. It has been observed in animals as well, where recently a non-autonomous short interspersed nuclear element (SINE3) has been reported for vertebrates (Kapitonov and Jurka, 2003; Gogolevsky et al., 2009). The emergence of 5S derived transposons in several eukaryote lineages that show switches in rDNA organization patterns might be a side effect of the relocating mechanisms. Previous studies have shown that transposable elements might have mediated the translocation of single 5S rDNA gene

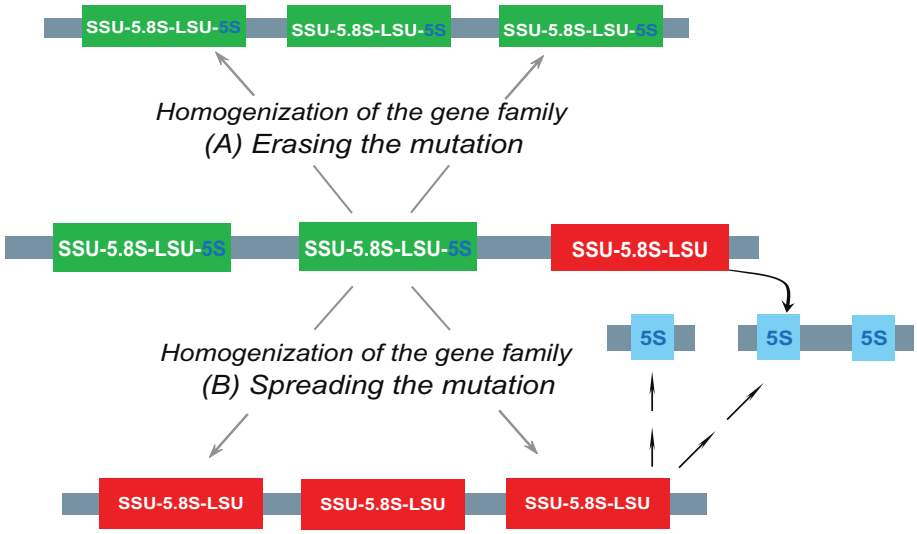


Fig. 5. Relocation of 5S rDNA under concerted evolution. Possible scenario for the deletion of one 5S rDNA gene from the intergenic spacer of a 45S repeat can either be erased (a) or spread (b) by mechanisms that produce concerted evolution. See Sections 4.1 and 4.3 for a detailed explanation.

copies or are responsible for 5S gene duplications (Drouin, 2000; Garcia et al., 2009). However, concerted translocation of several copies by transposable elements for any gene is not known so far and appears unlikely as sole trigger of 5S rDNA cluster formation although an involvement of (retro)transposon activity acting in combination with other mechanisms cannot be excluded. Involvement of extrachromosomal covalently closed circular DNA (ecccDNA) as proposed in some works (Drouin and de Sa, 1995b, Garcia et al., 2009) or even a helitron-like mechanism are other possible players, especially considering the fact that rDNA undergoes unusual replication ways in some organisms (e.g. rolling-circle replication; Hourcade et al., 1973; Rochaix and Bird, 1975; Bakken, 1975; Backert et al., 1996; Cohen et al., 2005).

4.2. Regulatory elements and 5S functionality

Analysis of eukaryotic 5S rRNA structure models (as proposed by Luehrsen and Fox, 1981; Gewirth et al., 1987; Barciszewska et al., 1994, 1996; Dinman, 2005) that are based upon both experimental evidence and computer modeling suggests full functionality of L-type 5S rRNA genes (Fig. 4) and S-type organized 5S rDNA copies of all taxa studied here. Refined analysis regarding tertiary structure models (after McDougall and Nazar, 1983; Barciszewska et al., 1994) can even corroborate previous models by incorporating structural information and allow the inference of a putative consensus structure for land plant 5S rRNAs (Fig. 4). Secondary structure analyses of the 5S rRNA revealed a region of higher mutation rate in the stem region of the 5S rRNA for early diverging land plants, with a high degree of compensating base pair changes (especially in domain α), as illustrated in the consensus structure for land plants (Fig. 4; further details on nucleotide substitution patterns for each land plant lineage is provided in Supplementary material). The TATA-signal upstream of the 5S genic region has been shown to play an important role as a regulatory element in 5S transcription. Our data, however, suggests that this may only true for eudicots, as neither a conserved TATA-element between −30 and −3, nor other essential elements known for *pol*III transcription could be identified in most land plants including basal angiosperms and monocots. For example, TATA-signals for the transcription of 5S rDNAs are absent in the basal angiosperm *Victoria regia*, a water lily, as well as in the different 5S rDNA clusters of *Triticum aestivum* (Appels et al., 1980) and *Oryza* as well as for *Avena* species (Zhu et al., 2008). These latter 5S rDNA clusters are reported to contain a TATA-stretch within the 5S–5S NTS shortly downstream of the transcription termination signal, but not within a previously reported distance to the transcription initiation site (Röser et al., 2001). The role of those TATA-motifs for the transcription of oat grass (*Avena*) 5S rDNA is uncertain, and they might just represent coincidental motifs. It may be possible that regulating motifs upstream of the 5S genic region are lineage specific, given the current data. A perfect TATA-element mediating the binding of factors for the *pol*III transcription complex is only seen in *E. hyemale* and *Osmunda javanica*, while all other studied fern lineages seem to possess an altered signal, which differs in at least one nucleotide. The series AGGG occurs between −30 and +1 in all studied L-type ferns, and thus may play a role in transcription factor binding. Nevertheless, further recognition elements described previously seem to be absent or are heavily modified. This is an especially interesting case, since flowering plant signals are highly similar to those of vertebrates (*Xenopus*: Murphy et al., 1989) and invertebrates (*Bombyx:* Morton and Sprague, 1984). However, the apparent absence or divergence of such upstream regulatory elements (TATA-box, etc.) does not imply non-functionality of 5S genes as this is a recurrent feature of early land plant lineages. On the contrary it suggests a slightly different transcription regulation of 5S rDNA that can only be revealed by future experiments on the transcription and regulation of 5S rDNA in early land plants.

The problematic prediction of transcription initiation is not unique to 5S rDNA regulating elements in non-seed plants, but also accounts for the large ribosomal subunits, i.e. the 45S repeat. Although some motifs show a high similarity to previously identified critical promoter elements; further experimental evidence is required for early land plant lineages. Contrary to 5S upstream regulating elements, it is not surprising that *Pol*I regulating elements are difficult to predict as these have been shown early to be highly lineage specific (Moss and Stefanovsky, 1995; Paule and White, 2000).

4.3. 5S rDNA array formation

The aforementioned points directly lead to the question what causes the formation of a 5S rDNA cluster? The observation of the frequent change of 5S rDNA localization occurring within all eukaryotic kingdoms (as well as in prokaryotes) might indicate

that the translocation is likely to be linked to a very ancient molecular mechanism. Several hypotheses concerning the putative transposition of the 5S rRNA gene have been discussed earlier (Drouin and de Sa, 1995; Garcia et al., 2009). These range from recombination events to retrotransposition and ecccDNA involvement. The discovery of a 5S derived retroelement and its wide distribution among tracheophytes has offered a possible theory for the translocation of 5S rDNA in plants (Kalendar et al., 2008; Garcia et al., 2009). Several additional lines of evidences indicate that the duplication of the 5S RNA gene within the IGS of e.g. *E. hyemale* could indeed be mediated by retrotransposon activity as has also been discussed for 5S gene duplication in the IGS of *Artemisia* (Garcia et al., 2009). Since the IGS is, however, often packed with direct and/or inverted repeats as well as homopolymeric stretches whose origins and functions are widely unknown, the interpretation of poly-A stretches as traces of transposon activity is often only speculative and should be considered with some care.

4.4. Cassandra 5S-like elements

Interestingly, the PCR screens did not detect *Cassandra* elements within early land plant and basal monilophyte lineages, including water ferns. Similarly, extensive blast-searches in the draft genomes of *P. patens*, *S. moellendorfii* and *C. vulgaris* confirmed the absence of *Cassandra* elements with no indication of other 5S derived transposable elements. BLAST-searches against early genome releases of *S. moellendorfii* yielded numerous truncated 5S-like sequences that contained perfectly conserved internal promoter elements, but none was associated with known coding elements. Thus, *Cassandra* elements might not account for nrDNA re-organization events in chlorophyte algae and water ferns. In contrast, PCR screens within derived leptosporangiate ferns and seed plants using the primer combination IGS1R1 and IGS2F2 often yielded *Cassandra* retroelements. Alignment and analysis of the 5S-like sequence within the terminal repeat of *Cassandra*-5S-like elements with true 5S rDNA clearly shows extreme sequence divergence (>68%). *Cassandra* 5S like elements cannot form stable 5S rRNA structures indicated by the secondary structure estimations (Fig. 4, Supplementary material). In contrast to 5S rRNA sequences, nearly all base pair changes observed in *Cassandra* are not compensatory, indicating a neutral mutation that may lead to the incapability of forming tertiary interactions. Furthermore, the high sequence divergence observed in *Cassandra* 5S like elements, makes it virtually impossible to form a stable 5S rRNA molecules (Fig. 4, bottom). In most cases high sequence similarity of 5S rRNA and *Cassandra*-5S is only obvious within the internal promoter regions. In particular, 5S promoter boxes A and C are highly conserved, while the intermediate element (IE or Box B) is variable and highly divergent even within one species and likely to be deprived of its functionality. The retained 5S promoter might, however, serve or assist the transcription of the retroelement. It is known from several studies that retroelements are subject to either polymerase II or III transcription (e.g. Boeke and Corces, 1989; Marschalek et al., 1992; Kumar and Bennetzen, 1999; Wessler, 2006; Myakishev et al., 2008; Gogolevsky et al., 2009). Thus, it is not surprising that promoter elements from all *pol*III processed gene types have been found in mobile elements including promoter elements of tRNAs, 7SL RNA and 5S rRNA. Recently, a 5S rRNA like element (P5SM) found exonized within the TFIIIA coding sequence was also shown to regulate its own transcription by alternative splicing of TFIIIA which is a transcription factor known to be crucial for *pol*III transcription (Hammond et al., 2009). In addition, the 5S promoter boxes A and C were intact, while the IE was modified. The remainder genic region was strongly diverged showing an insertion of several nucleotides in length. Taking together these findings, it is imaginable that 5S promoters may play more supporting roles in other transcription or regulating pathways of eukaryote organisms than has currently been postulated.

5. Summary and conclusions

This study revealed that, in green algae and land plants at least one core element of ribosomal DNA is subject to dynamic rearrangements. In addition, the study indicates that generalizations derived from a single model plant, such as *Arabidopsis* or *Physcomitrella*, need to be considered with some care, especially if no phylogenetic backup is provided. However, in the light of the current phylogenetic concept of land plants, the observed physical linkage of all rRNA genes in representatives of streptophyte algae, liverworts, mosses, hornworts, lycophytes and monilophytes, except water ferns, indicates that the L-type organization may be regarded as the ancestral state of rDNA organization within land plants and possibly also for streptophyte algae (although more data from green algal taxa is necessary to verify a "streptophyte-ancestral" hypothesis for L-type organization). Several independent 5S transposition events may have occurred during the evolution of land plants, as illustrated by the convergent 5S rDNA cluster formation in water ferns and seed plants, as well as by the reported reversal from S- to L-type repeats in the eudicot genus *Artemisia*. Both organization types provide advantages and disadvantages ranging from more efficient homogenization of gene copies, equivalent copy numbers of all rRNA genes and close proximity during transcription for L-type repeats, while for S-type organization might facilitate polymerase recycling.

Irrespective of the organization type, the structure of the 5S rRNA molecule is highly conserved and stable among all plant lineages implying full functionality, despite elevated substitution rates in various lineages such as the Gnetales or lycophytes. Almost half of the observed nucleotide substitutions were compensatory indicating that selection is acting on the coding sequences to enable the formation of stable secondary structures. The transposable *Cassandra*-5S-like elements are apparently not capable of adopting stable secondary and tertiary structures. Missing or highly diverged regulatory elements of 5S rDNA transcription in early land plants suggests that they may possess different transcriptional regulatory mechanisms compared to most eudicot model plants.

Although independent transposition events could be inferred in green algae and land plants, the trigger for 5S-relocation and the relocating mechanisms remain unclear. However, *Cassandra* might be rejected as a primary relocating mechanism in water ferns and the chlorophyte *Chlamydomonas*, as this retroelement appears to be restricted to polypod ferns and angiosperms.

Acknowledgments

This research was initially funded by the European Commission's BIODIBERIA HUMAN POTENTIAL PROGRAMME (D.Q.), and later received support by SYNTHESYS (S.W., D.Q.) that was financed by the European Community Research Infrastructure Action under the FP6 "Structuring the European Research Area" Programme (http://www.synthesys.info) to perform research at the Royal Botanical Garden, Madrid, Spain. We thank and highly appreciate technical support by Monika Ballmann (University of Bonn). The authors would like to thank Barbara Ditsch (Botanical Garden of Dresden), Wolfram Lobin (Botanical Garden of Bonn), Nikolai Friesen (Botanical Garden Osnabrück), Frank Müller (TU Dresden), Thomas Friedl (SAG Göttingen) and Michael Dilger for providing fresh plant material as well as Ingrid Essigmann-Capesius for continuous usage of her DNA-collection. We appreciate helpful comments and suggestions by Liz Zimmer and an anonymous reviewer on an earlier version of the manuscript. We are especially

grateful for the generous support of the people at the Royal Botanical Garden Madrid, Spain. Sincere thanks are due to Christoph Neinhuis (TU Dresden, Germany) for providing lab space during early days of this study.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.06.023.

References

Abirached-Darmency, M. et al., 2005. Variation in rDNA locus number and position among legume species and detection of 2 linked rDNA loci in the model *Medicago truncatula* by FISH. Genome 48, 556–561.

Amstutz, H. et al., 1985. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. Cell 40, 879–886.

Appels, R. et al., 1980. Molecular and chromosomal organization of DNA sequences coding for the ribosomal RNAs in cereals. Chromosoma 78, 293–311.

Backert, S. et al., 1996. Rolling-circle replication of mitochondrial DNA in the higher plant *Chenopodium album* (L.). Mol. Cell. Biol 16, 6285–6294.

Bakken, A., 1975. Replication of amplifying ribosomal deoxyribonucleic acid in rolling circles in *Xenopus laevis* oocytes. J. Histochem. Cytochem. 23, 463–474.

Banks, J.A. et al., 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science. doi:10.1126/science.1203810.

Barciszewska, M.Z. et al., 1994. A new model for the tertiary structure of 5S ribonucleic acid in plants. Plant Mol. Biol. Rep. 12, 116–131.

Barciszewska, M.Z. et al., 1996. Ribosomal 5S rRNA: tertiary structure and interactions with proteins. Biol. Rev. 71, 1–25.

Bergeron, J., Drouin, G., 2008. The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungal species. Curr. Genet. 54, 123–131.

Boeke, J.D., Corces, V.G., 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. 43, 403–434.

Borisjuk, N.V. et al., 1997. Structural analysis of rDNA in the genus *Nicotiana*. Plant Mol. Biol. 35, 655–660.

Brow, D.A., 1987. In vitro transcripts of a yeast variant 5 S rRNA gene exhibit alterations in 3′-end processing and protein binding. J. Biol. Chem. 262, 13959–13965.

Brown, T.A., 2002. Genomes, second ed. Garland Science, Oxford.

Cabral-de-Mello, D.C. et al., 2010. Chromosomal mapping of repetitive DNAs in the beetle *Dichotomius geminatus* provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. Heredity 104, 393–400.

Capesius, I., 1997. Analysis of the ribosomal RNA gene repeat from the moss *Funaria hygrometrica*. Plant Mol. Biol. 33, 559–564.

Cloix, C. et al., 2000. Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. Genome Res. 10, 679–690.

Cohen, S. et al., 2005. Evidence for rolling circle replication of tandem genes in *Drosophila*. Nucl. Acids Res. 33, 4519–4526.

Cruces, J. et al., 1989. The 5S rRNA-histone repeat in the crustacean *Artemia*: structure, polymorphism and variation of the 5S rRNA segment in different populations. Nucl. Acids Res. 17, 6283–6297.

Dinman, J.D., 2005. 5S rRNA: structure and function from head to toe. Int. J. Biomed. Sci. 1, 2–7.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. Focus 12, 13–15.

Drouin, G., 2000. Expressed retrotransposed 5S rRNA genes in the mouse and rat genomes. Genome 43, 213–215.

Drouin, G., de Sa, M., 1995. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. Mol. Biol. Evol. 12, 481–493.

Drouin, G. et al., 1992. Variable arrangement of 5S ribosomal genes within the ribosomal DNA repeats of arthropods. Mol. Biol. Evol. 9, 826–835.

Dujon, B., 2010. Yeast evolutionary genomics. Nat. Rev. Genet. 11, 512–524.

Ganley, A.R., Scott, B., 2002. Concerted evolution in the ribosomal RNA genes of an *Epichloë* endophyte hybrid: comparison between tandemly arranged rDNA and dispersed 5S rrn genes. Fungal Genet. Biol. 35, 39–51.

Garbary, D.J. et al., 1993. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. Plant Syst. Evol. 188, 237–269.

Garcia, S. et al., 2007. Extensive ribosomal DNA (18S-5.8S-26S and 5S) colocalization in the North American endemic sagebrushes (subgenus Tridentatae, Artemisia, Asteraceae) revealed by FISH. Plant Syst. Evol. 267, 79–92.

Garcia, S. et al., 2009. Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. Chromosoma 118, 85–97.

Gerbi, S.A., 1986. The evolution of eukaryotic ribosomal DNA. Biosystems 19, 247–258.

Gewirth, D.T. et al., 1987. Secondary structure of 5S RNA: NMR experiments on RNA molecules partially labelled with nitrogen-15. Biochemistry 26, 5213–5220.

Gilson, P.R. et al., 1995. Organisation and sequence analysis of nuclear-encoded 5S ribosomal RNA genes in cryptomonad algae. Curr. Genet. 27, 239–242.

Gogolevsky, K.P. et al., 2009. 5S rRNA-derived and tRNA-derived SINEs in fruit bats. Genomics 93, 494–500.

Goldsbrough, P.B., Cullis, C.A., 1981. Characterisation of the genes for ribosomal RNA in flax. Nucl. Acids Res. 9, 1301–1310.

Goldsbrough, P.B. et al., 1982. Sequence variation and methylation of the flax 55 RNA genes. Nucl. Acids Res. 10, 4501–4514.

Goodpasture, C., Bloom, S.E., 1975. Visualization of nucleolar organizer regions in mammalian chromosomes using silver staining. Chromosoma 53, 37–50.

Gruendler, P. et al., 1991. RDNA intergenic region from *Arabidopsis thaliana* structural analysis, intraspecific variation and functional implications. J. Mol. Biol. 221, 1209–1222.

Hammond, M.C. et al., 2009. A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. Nat. Struct. Mol. Biol. 16, 541–549.

Hemleben, V. et al., 1988. Organization and length heterogeneity of plant ribosomal RNA genes. In: Kahl, G. (Ed.), Architecture of Eukaryotic Genes. VCH Verlagsgesellschaft mbH, Weinheim, pp. 371–383.

Hourcade, D. et al., 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. Proc. Natl. Acad. Sci. USA 70, 2926–2930.

Jinks-Robertson, S. et al., 1993. Substrate length requirements for efficient mitotic recombination in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 13, 3937–3950.

Kalendar, R. et al., 2008. Cassandra retrotransposons carry independently transcribed 5S RNA. Proc. Natl. Acad. Sci. USA 105, 5833–5838.

Kapitonov, V.V., Jurka, J., 2003. A novel class of SINE elements derived from 5S rRNA. Mol. Biol. Evol. 20, 694–702.

Kato, A. et al., 1990. The structure of the large spacer region of the rDNA in *Vicia faba* and *Pisum sativum*. Plant Mol. Biol. 14, 983–993.

Kawai, H. et al., 1997. Linkage of 5S ribosomal DNA to other rDNAs in the chromophytic algae and related taxa. J. Phycol. 33, 505–511.

King, K. et al., 1993. Molecular evolution of the intergenic spacer in the nuclear ribosomal RNA genes of Cucurbitaceae. J. Mol. Evol. 36, 144–152.

Komarova, N. et al., 2004. Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. Plant Mol. Biol. 56, 439–463.

Kumar, A., Bennetzen, J.L., 1999. Plant retrotransposons. Annu. Rev. Genet. 33, 479–532.

Leah, R. et al., 1990. Nucleotide sequence of a mouse 5S rRNA variant gene. Nucl. Acids Res. 18, 7441.

Lewis, L.A., McCourt, R.M., 2004. Green algae and the origin of land plants. Am. J. Bot. 91, 1535–1556.

Liu, Z. et al., 2003. Intragenomic and interspecific 5S rDNA sequence variation in five Asian pines. Am. J. Bot. 90, 17–24.

Long, E.O., Dawid, I.B., 1980. Repeated genes in eukaryotes. Annu. Rev. Biochem. 49, 727–764.

Luehrsen, K.R., Fox, G.E., 1981. Secondary structure of eukaryotic cytoplasmic 5S ribosomal RNA. Proc. Natl. Acad. Sci. USA 78, 2150–2154.

Marco, Y., Rochaix, J., 1980. Organization of the nuclear ribosomal DNA of *Chlamydomonas reinhardii*. Mol. Gen. Genet. 177, 715–723.

Marschalek, R. et al., 1992. Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. Mol. Cell. Biol. 12, 229–239.

McDougall, J., Nazar, R.N., 1983. Tertiary structure of the eukaryotic ribosomal 5 S RNA. Accessibility of phosphodiester bonds to ethylnitrosourea modification. J. Biol. Chem. 258, 5256.

Merchant, S.S. et al., 2007. The *Chlamydomonas* Genome reveals the evolution of key animal and plant functions. Science 318, 245–250.

Mishler, B.D., Churchill, S.P., 1984. A cladistic approach to the phylogeny of the "Bryophytes". Brittonia 36, 406–424.

Mitchelson, K., Moss, T., 1987. The enhancement of ribosomal transcription by the recycling of RNA polymerase I. Nucl. Acids Res. 15, 9577–9596.

Morton, D.G., Sprague, K.U., 1984. In vitro transcription of a silkworm 5S RNA gene requires an upstream signal. Proc. Natl. Acad. Sci. USA 81, 5519–5522.

Morzycka-Wroblewska, E. et al., 1985. Concerted evolution of dispersed *Neurospora crassa* 5S RNA genes: pattern of sequence conservation between allelic and nonallelic genes. Mol. Cell Biol. 5, 46–51.

Moss, T., Stefanovsky, V.Y., 1995. Promotion and regulation of ribosomal transcription in eukaryotes by RNA polymerase I. Prog. Nucl. Acids. Res. Mol. Biol. 50, 25–66.

Moss, T. et al., 1992. Readthrough enhancement and promoter occlusion on the ribosomal genes of *Xenopus laevis*. Biochem. Cell Biol. 70, 324–331.

Murphy, S. et al., 1989. Common mechanisms of promoter recognition by RNA polymerases II and III. Trends Genet. 5, 122–126.

Myakishev, M. et al., 2008. PCR-based detection of Pol III-transcribed transposons and its application to the rodent model of ultraviolet response. Cell Stress Chaperon. 13, 111–116.

Palmer, J.D. et al., 2004. The plant tree of life: an overview and some points of view. Am. J. Bot. 91, 1437–1445.

Pardue, M.L. et al., 1973. Location of the genes for 5S ribosomal RNA in *Xenopus laevis*. Chromosoma 42, 191–203.

Paule, M.R., White, R.J., 2000. Transcription by RNA polymerases I and III. Nucl. Acids Res. 28, 1283–1298.

Piper, P.W. et al., 1984. A minor class of 5S rRNA genes in *Saccharomyces cerevisiae* X2180-1B, one member of which lies adjacent to a Ty transposable element. Nucl. Acids Res. 12, 4083–4096.

Procunier, J.D., Tartof, K.D., 1975. Genetic analysis of the 5S RNA genes in *Drosophila melanogaster*. Genetics 81, 515–523.

Pryer, K.M. et al., 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409, 618–622.

Qiu, Y.-L. et al., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA 103, 15511–15516.

Quandt, D., Stech, M., 2002. Molecular systematics of bryophytes in context of land plant phylogeny. In: Sharma, A.K., Sharma, A. (Eds.), Plant Genome. Oxford & IBH Publishing, New Delhi, pp. 268–295.
Reddy, R. et al., 1986. Some gene variants for 5 S RNA are dispersed in the rat genome. J. Biol. Chem. 261, 10618–10623.
Rensing, S.A. et al., 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69.
Rocha, P.S.C.F., Bertrand, H., 1995. Structure and comparative analysis of the rDNA intergenic spacer of *Brassica rapa*. Eur. J. Biochem. 229, 550–557.
Rochaix, J., Bird, A.P., 1975. Circular ribosomal DNA and ribosomal DNA: replication in somatic amphibian cells. Chromosoma 52, 317–327.
Rogers, S.O., Bendich, A.J., 1987. Ribosomal RNA genes in plants: variability in copy number and in intergenic spacer. Plant Mol. Biol. 9, 509–520.
Röser, M. et al., 2001. Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). Mol. Phyl. Evol. 21, 198–217.
Rubin, G.M., Sulston, J.E., 1973. Physical linkage of the 5 S cistrons to the 18 S and 28 S ribosomal RNA cistrons in *Saccharomyces cerevisiae*. J. Mol. Biol. 79, 521–530.
Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
Sastri, D.C. et al., 1992. An overview of evolution in plant 5S DNA. Plant Syst. Evol. 183, 169–181.
Schiebel, K. et al., 1989. Termination of transcription of ribosomal RNA genes of mung bean occurs within a 175 bp repetitive element of the spacer region. Mol. Gen. Genet. 218, 302–307.
Schneeberger, R.G. et al., 1989. Chromosomal and molecular analysis of 5S RNA gene organization in the flax, *Linum usitatissimum*. Gene 83, 75–84.
Schneider, H. et al., 2004. Ferns diversified in the shadow of angiosperms. Nature 428, 553–557.
Schubert, I., Künzel, G., 1990. Position-dependent NOR activity in barley. Chromosoma 99, 352–359.
Shaw, J., Renzaglia, K., 2004. Phylogeny and diversification of bryophytes. Am. J. Bot. 91, 1557–1581.
Smith, A.R. et al., 2006. A classification for extant ferns. Taxon 5, 705–731.
Sone, T. et al., 1999. Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. Plant Mol. Biol. 41, 679.
The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.
Vakhitov, V. et al., 1986. Determination of the number of copies of genes coding for 5S rRNA and tRNA in the genomes of 43 species of wheat and *Aegilops*. Genetika 22, 676–683.
Volkov, R.A. et al., 2003. Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. Petota (genus *Solanum*). Mol. Phyl. Evol. 29, 187–202.
Watt, V.M. et al., 1985. Homology requirements for recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 82, 4768–4772.
Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. USA 103, 17600–17601.
Zentgraf, U. et al., 1990. Length heterogeneity of the rRNA precursor in cucumber (*Cucumis sativus*). Plant Mol. Biol. 15, 465–474.
Zentgraf, U. et al., 1998. Different transcriptional activities in the nucleus. In: Behnke et al. (Eds.), Progress in Botany. Springer, Berlin, Heidelberg, New York, pp. 131–168.
Zhu, X. et al., 2008. Molecular and cytological characterization of 5S rDNA in *Oryza* species: genomic organization and phylogenetic implications. Genome 51, 332–340.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31, 3406–3415.