npg
Heredity (2012) 108, 640–646
© 2012 Macmillan Publishers Limited All rights reserved 0018-067X/12
www.nature.com/hdy

Heredity

ORIGINAL ARTICLE

Early evolutionary colocalization of the nuclear ribosomal 5S and 45S gene families in seed plants: evidence from the living fossil gymnosperm *Ginkgo biloba*

JA Galián[1], M Rosato[1] and JA Rosselló[1,2]

In seed plants, the colocalization of the 5S loci within the intergenic spacer (IGS) of the nuclear 45S tandem units is restricted to the phylogenetically derived Asteraceae family. However, fluorescent *in situ* hybridization (FISH) colocalization of both multigene families has also been observed in other unrelated seed plant lineages. Previous work has identified colocalization of 45S and 5S loci in *Ginkgo biloba* using FISH, but these observations have not been confirmed recently by sequencing a 1.8 kb IGS. In this work, we report the presence of the 45S–5S linkage in *G. biloba*, suggesting that in seed plants the molecular events leading to the restructuring of the ribosomal loci are much older than estimated previously. We obtained a 6.0 kb IGS fragment showing structural features of functional sequences, and a single copy of the 5S gene was inserted in the same direction of transcription as the ribosomal RNA genes. We also obtained a 1.8 kb IGS that was a truncate variant of the 6.0 kb IGS lacking the 5S gene. Several lines of evidence strongly suggest that the 1.8 kb variants are pseudogenes that are present exclusively on the satellite chromosomes bearing the 45S–5S genes. The presence of ribosomal IGS pseudogenes best reconciles contradictory results concerning the presence or absence of the 45S–5S linkage in *Ginkgo*. Our finding that both ribosomal gene families have been unified to a single 45S–5S unit in *Ginkgo* indicates that an accurate reassessment of the organization of rDNA genes in basal seed plants is necessary.
Heredity (2012) **108,** 640–646; doi:10.1038/hdy.2012.2; published online 22 February 2012

Keywords: rDNA; intergenic spacer; 45S–5S unit; ribosomal pseudogenes gymnosperm; phylogenetic relict

INTRODUCTION

Ribosomes are cellular structures that are ubiquitous in the cytoplasm, mitochondria and chloroplasts of living organisms. They are universally formed by two subunits that are composed of ribosomal RNA (rRNA) molecules and about 80 ribonucleoproteins that play important roles in correctly folding and packaging the rRNAs (Lafontaine and Tollervey, 2001).

Given the crucial role of ribosomes in the protein synthesis pathway, the ribosomal genes (rDNA) are some of the most abundant tandemly repeated DNA sequences in eukaryotes, with highly conserved structure and organization throughout the tree of life (Long and Dawid, 1980). Nuclear ribosomal DNA belongs to two universal gene families in eukaryotes, the 5S rDNA and 45S rDNA (alternatively called 35S rDNA, for example, Komarova *et al.*, 2008). The 5S rDNA repeat unit consists of a $\sim$120 bp gene and a relatively short intergenic spacer (IGS), usually between 100 and 900 bp (Sastri *et al.*, 1992). The major cluster (45S rDNA) correspond to the nucleolar organizer regions, and consists of three coding regions (18S, 5.8S and 25S/26S/28S, the two former gene names indistinctly used for plants and the latter used for animal organisms), internal transcribed spacers between genes (ITS1 and ITS2) and a long IGS separating adjacent unit repeats (Rogers and Bendich, 1987; Jorgensen and Cluster, 1988).

The 5S and 45S rDNA families typically occur as independent loci in most eukaryotic phylogenetic lineages, as revealed by molecular cytogenetic techniques (fluorescent *in situ* hybridization, FISH), polymerase chain reaction (PCR) and sequencing. However, integration of the 5S gene within the 45S rDNA family appears to have occurred independently in unrelated evolutionary groups (Maizels, 1976; Batts-Young and Lodish, 1978; Vahidi *et al.*, 1991; Capesius, 1997; Sone *et al.*, 1999; Vitturi *et al.*, 2002; Bergeron and Drouin, 2008). In these cases, the 5S genes are included within the IGS region separating the 25S/26S/28S and 18S coding regions, although other rare organizations have also been described (Drouin and Moniz de Sá, 1995).

For a long time, vascular plant species were characterized as having unlinked rDNA loci, in contrast to the linked structure detected in early land plants (bryophytes; Sone *et al.*, 1999). Recently, however, it has been unequivocally shown for the first time, using molecular cytogenetics, Southern blot hybridization, PCR-based techniques and DNA sequencing, that 5S genes can be embedded within the 45S multigene family in angiosperms (Asteraceae; García *et al.*, 2009; García *et al.*, 2010). The detailed survey within Asteraceae has revealed that in about 25% of the species in this family, the two ribosomal gene families have been combined into a single, linked 45S–5S unit (García *et al.*, 2010). These authors suggested that the 5S gene integration within the 45S multigene family probably occurred once in Asteraceae, but might have repeatedly occurred elsewhere during plant evolution.

Recently, Wicke *et al.* (2011) assessed the rDNA organization in land plants, using PCR-based techniques, Southern blot hybridization and DNA sequencing. Their results implied that unlinked 45S–5S rDNA arrays were identified for all seed plants checked, including the

[1]Jardín Botánico de Valencia, Universidad de Valencia, Valencia, Spain and [2]Marimurtra Botanic Garden, Carl Faust Fdn, Blanes, Catalonia, Spain
Correspondence: Dr JA Rosselló, Jardín Botánico de Valencia, Universidad de Valencia, c/Quart 80, 46008 Valencia, Spain.
E-mail: rossello@uv.es
Received 3 August 2011; revised 7 December 2011; accepted 3 January 2012; published online 22 February 2012

10 gymnosperms and the single basal angiosperm lineage experimentally analyzed (Nympheales), and the scrutiny of 10 ribosomal 45S IGS sequences available in public databases (Wicke *et al.*, 2011). These results suggest that in seed plants linked 45S–5S units were exclusive of the Asteraceae family, a derived angiosperm lineage, and that the linkage of both ribosomal gene families likely occurred 7–10 million years ago (Mya) (García *et al.*, 2010).

Some of the results obtained by Wicke *et al.* (2011) in seed plants are intriguing. The juxtaposition of 45S and 5S fluorescent-based signals in dividing cells or interphase nuclei, as observed by FISH, has traditionally been used as putative evidence for the linkage of 45S and 5S rRNA genes in plants. In seed plants, FISH colocalization of both multigene families has been observed not only in the Asteraceae (García *et al.*, 2010), but also in other unrelated angiosperm lineages such as *Tulipa* (Mizuochi *et al.*, 2007), *Rhoeo* (Golczyk *et al.*, 2005), *Linum* (Muravenko *et al.*, 2004), *Silene* (Široký *et al.*, 2001), *Brassica* (Snowdon *et al.*, 2000; Hasterok *et al.*, 2001), as well as in a couple of gymnosperm genera: *Ginkgo* (Nakao *et al.*, 2005) and *Podocarpus* (Murray *et al.*, 2002). However, given the shortcomings of the axial-resolution detection limits of classical FISH (between 2000 and 10 000 kb; Figueroa and Bass, 2010), such molecular cytogenetic methods could generate crude estimates or visual artifacts concerning the organization of the 45S and 5S rDNA families in plants.

Former work has identified colocalization of 45S and 5S in *Ginkgo* using FISH (5S genes are apparently linked to the two active 45S loci; Nakao *et al.*, 2005), but this merit a closer investigation as recently Wicke *et al.* (2011) failed to detect such linkage in the IGSs they sequenced.

The maidenhair tree (*Ginkgo biloba* L.) is the oldest extant genus among seed plants and was already distinguished by Charles Darwin as a living fossil on the basis of its phylogenetic isolation, relictual distribution and its conspicuous morphological similarities to Mesozoic and Cenozoic congenérics (Major, 1967; Royer *et al.*, 2003). *G. biloba* is the sole survivor of the Ginkgoalean clade, a once dominant and diverse gymnosperm clade (consisting of up to six families and 19 genera; Tralau, 1968) that existed and flourished early in the Permian and achieved maximum diversity during the Jurassic and Early Cretaceous (Royer *et al.*, 2003). Some authors, on the basis of the identical morphology shared between modern *G. biloba* and the fossil *G. adiantoides* (Unger) Heer have suggested their conspecificity, implying that *G. biloba* was already present in the earth from latest Cretaceous to middle Miocene (Royer *et al.*, 2003).

Given the contradictory results previously reported by two independent teams, in this paper we aimed to discern whether 45S and 5S rDNA genes were combined into a single linkage unit in the phylogenetic gymnosperm relict *G. biloba*, using molecular cytogenetics, and DNA sequencing of the nuclear ribosomal 45S IGS.

MATERIALS AND METHODS

Plant sampling

A female accession from an old *Ginkgo* plant was obtained from the living collections of the Botanical Garden of Valencia University, and this was used as a source for DNA extraction. The tree was acquired about 1850 from an unknown source, but according to Del Tredici (1991), European *Ginkgo* trees were imported for the first time from Japan in the early eighteenth century. In addition, several young plants growing in the garden were transplanted to pots for cytogenetic analyses.

DNA extraction

Total genomic DNA was isolated from fresh leaves using the DNAeasy Plant Minikit (Qiagen Inc, Valencia, CA, USA) following the manufacturer's instructions.

IGS amplification, cloning and sequencing

The nuclear ribosomal IGS fragment was amplified according to protocols designed for long PCR products using the universal plant primers 25 F and 18S R. Primers 25 F (5′-AAATATGGGACGAGGTATTGTAAG-3′) and 18S R (5′-AAGCATATGACTACTGGCAGGAT-3′) were designed to anneal to the 3′ end of the conserved region of the 26S rDNA gene in the forward direction and the 5′ end of 18S rDNA gene in the reverse orientation, respectively.

PCR was conducted with total reaction volumes of 50 μl, with each reaction containing approximately 5 ng of genomic DNA, 0.5 μM of each primer, 2.5 mM $MgCl_2$, 2 mM Tris–HCl (pH 8.0), 5 mM KCl, 0.0001% bovine serum albumin, 250 μM dNTPs and 2.5 U of DNA polymerase (TaKaRa LA Taq; Takara Biotechnology Inc., Takara, Japan). The amplification conditions were as follows: the PCR mix (without the DNA polymerase) was subjected to an initial denaturation step conducted at 94 °C for 1 min, after which the temperature was lowered to 80 °C, the DNA polymerase was added and the reaction was incubated for 5 min at 80 °C. After these preincubation steps, 40 amplification cycles were performed on a Primus (MWG-Biotech, Milton Keynes, UK) thermal cycler, with each cycle consisting of a denaturation step at 93 °C for 1 min, annealing at 55 °C for 1 min and primer elongation at 72 °C for 6 min and 15 s. The PCR reaction ended with a final elongation of 10 min at 72 °C, and the amplified fragments were separated on 1% (w/v) agarose gels.

To prevent damage to the amplified DNA during the cloning process, PCR fragments were excised from the agarose gel using the UV-Free Gel Purification Kit (Invitrogen, Paisley, UK). The purified PCR fragments were cloned into the pCR-XL-TOPO (TA Cloning Kit; Invitrogen) according to the manufacturer's instructions. Plasmids were transformed into chemically competent *Escherichia coli* TOP 10 cells (Invitrogen), and plasmid extractions of the recombinant clones were performed using a REAL Miniprep Turbo Kit (REAL Durviz SL, Valencia, Spain). Purified plasmids were sequenced with an ABI 3100 Genetic Analyzer using the ABI BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA). DNA sequencing was performed using a primer walking strategy, whereby vector-specific (M13F and M13R) and internal primers were used to generate partial sequences. This was repeated until sequences were obtained that overlapped in the forward and reverse directions. The sequences of the internal primers designed for sequencing can be obtained upon request.

Sequence analyses

The sequences determined by cloning and sequencing were compared to GenBank database sequences using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1997). The sequences were aligned using CLUSTALX ver. 1.83 (Thompson *et al.*, 1997). The EMBL-EBI Available Analysis Tools (http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+applSelect) were used for sequence analysis, and the BioEdit Sequence Alignment Editor was used for dot matrix analysis. A search for sequence repeats and motifs was performed using the Repeats Finder Tool from EMBL-EBI and Oligonucleotides Repeats Finder (http://wwwmgs.bionet.nsc.ru/mgs/programs/oligorep/InpForm.htm). Results derived from bioinformatics methods were manually refined during direct inspection of the results.

Cytogenetic analyses

Chromosome preparations. Seedlings and young plants were cultivated in pots in the greenhouse to obtain actively growing root apices. The root tips were removed and pretreated with 0.002 M 8-hydroxyquinoline for 6 h at 0 °C, fixed in 3:1 ethanol:acetic acid and stored at –20 °C until used. The root tips were washed in 10 mM citrate buffer at pH 4.6, and then macerated in a mixture of 2% (v/v) cellulase (Calbiochem, Merck KGaA, Darmstadt, German) in citrate buffer (pH 4.6) and 20% pectinase (from *Aspergillus niger*) in 40% glycerol in 10 mM citrate buffer (pH 4.6), and this reaction was incubated for 90 min at 37 °C. The spreading procedure, used to prepare nuclei and chromosomes for *in situ* hybridization, was conducted as described by Zhong *et al.* (1996).

Probes for in situ hybridization. The 1.84 kb IGS product cloned from *G. biloba* was converted to a FISH probe via labeling with biotin-16-dUTP according to a standard PCR-based procedure. The pTa794 fragment, a cloned 410 bp *Bam*HI fragment of the 5S rDNA (including the 120 bp coding sequence for the 5S rRNA and the IGS), was isolated from *Triticum aestivum* L.

(Gerlach and Dyer, 1980) and converted to a FISH probe by labeling it with digoxigenin-11-dUTP via a nick translation procedure.

In situ hybridization and detection. We followed the *in situ* hybridization protocol used by Rosato *et al.* (2008), and probe detection was conducted using the method of Hanson *et al.* (1995) with minor modifications. After detection, the chromosomes were mounted in Vectashield antifade (Vector Laboratories Inc., Burlingame, CA, USA) with 5 μg/ml of 4′,6-diamidino-2-phenylindole used as a counterstain. Hybridization signals were analyzed using an epifluorescence Olympus microscope, with an appropriate filter set, and equipped with an Olympus Camedia C-2000-Z digital camera. The images were optimized for best contrast and brightness using image processing software (Adobe Photoshop v. 7.0).

RESULTS

Amplification and sequencing of IGS variants

Following specific protocols and using DNA polymerases capable of amplifying long PCR products (up to 17.5 kb using human DNA according to the Takara polymerase manufacturer's guide) resulted in the amplification of six putative IGS products (0.5, 0.7, 0.8, 0.9, 1.8 and 6.0 kb). Sequencing of these cloned products revealed two main groups. The first group included the 1.8 and 6.0 kb clones and showed structural features of plant IGS sequences, with the longest amplicon containing the 5S nuclear gene. The 1.8 kb clone was a truncated version of the 6.0 kb product with a 4.2 kb deletion between nucleotides 278 and 4573 near the 5′ end. The second group included the amplified products <1 kb in length, and these did not show any significant homology with the plant ribosomal sequences (other than the amplification primers) available in public nucleotide databases or with the structure and nucleotides of the 1.8 and 6.0 kb products. The nucleotide sequences reported in this paper have been submitted to GenBank under accession numbers JQ 279501–JQ 279502.

Structure and organization of the 6.0 kb IGS fragment

The 6.0 kb fragment was 6021 bp in length, had a 51.22% G+C content and was derived by trimming the 26S and 18S coding regions. Dot-blot analysis showed that most of the spacer sequence was composed of tandem repetitive elements (Figure 1) clustered at three subrepeat regions (A, B and C) and separated by short stretches of non-repeat sequences (Figure 1). The A, B and C subrepeat regions were 1275, 1865 and 1038 bp, respectively, and each contained conserved and imperfect copies of single motifs of 50, 10 and 219 bp, respectively. A single copy of the 5S nuclear gene was located between the A and B subrepeat regions exactly 2086 nucleotides from the 5′ end of IGS. The 5S gene was 121 bp in length and was inserted into the coding strand in the same direction of transcription as the 45S rRNA genes. The IGS sequence of *Ginkgo* showed several conserved motifs when compared with other angiosperm spacers, but also contained some unique features. Part of the *Ginkgo* motif was found imperfectly duplicated in 65 nucleotides downstream of the first occurrence. A putative transcription initiation site (TIS), TTTCAGGCGTG, was identified between the B and C boxes 4381 nucleotides from the 5′ end of the IGS. This motif was partially repeated twice from the adenine at the +1 position (AGGCGTG) to downstream locations at positions 4954 and 5230 within box C. The predicted secondary structure of the TIS motif revealed that it occurs at the beginning of a stem-loop structure (Figure 2) and is A+1 to the second nucleotide of the stem. IGS subrepeats are included between or near the promoter region and might act as additional promoter or enhancer elements. Thus, the putative TIS in *Ginkgo* is preceded by the B repetitive block (at position −173), which is composed of several AT-rich sequences (58.71% A+T content for the entire block). In the 5′-ETS, a GCGGCATGTG motif was present and shared significant similarity with GCGNATGAGTGT, a conserved sequence observed in angiosperms that is located 100 nucleotides downstream of the ETS processing site.

Figure 1 Structure of the intergenic spacers of the 45S rDNA cistron found in *G. biloba* (6.0 and 1.8 kb variants), showing the regions of the subrepeats (A, B, C), the location of the 5S gene, the putative TIS and the transcription termination sites (TTS).

Structure and organization of the 1.8 kb IGS variant

The smallest IGS variant amplified was 1814 bp in length (excluding coding regions) and had a G+C content of 60.14%. This 1.8 kb variant was a modified version of the putative, functional 6 kb IGS, with the main difference being an inferred deletion of the central portion of the 6 kb IGS. Thus, the 5′ ends of both IGS versions could be aligned, and a sequence alignment showed more than 86% similarity between nucleotides 1–366 of the 1.8 kb spacer and 1–214 of the 6.0 kb IGS

Figure 2 Predicted RNA secondary structure of the first 50 nucleotides beginning with the putative TIS sequence in *Ginkgo* (TTTCAGGCG). This stem-loop structure is not affected by folding of the entire 5′-ETS sequence.

variant. Similarly, nucleotides from 367 to 1814 of the smaller 1.8 kb variant and nucleotides 4573 to 6021 of the longer 6 kb IGS were more than 97% similar. The huge deletion distinguishing the two IGS sequences resulted in a complete loss of repetitive boxes A and B, as well as a small fragment of the 5′ end of the C box in the 1.8 kb variant. Some structural motifs involved in IGS processing, such as the pyrimidine-rich motif CCCCCACCCC (present at the 5′ end) and the putative termination motif CCCTCCC (located 68 bp downstream from the former), were present in the 1.8 kb variant. However, the putative transcription termination site and the GACTTGGC motif thought to be involved in transcription termination were not identified. Moreover, the 5S gene and all surrounding structural motifs linked to gene transcription in the spacer region were completely absent from the 1.8 kb fragment. These results suggest that the 1.8 kb variant is linked to a non-functional cistron.

Structural features of the 5S gene and upstream and downstream sequences

The 5S rDNA domain is composed of a 41 bp sequence in the 5′ end region that occurs after the A subrepeat region, a 5S coding sequence of 121 bp and a 33 bp 3′ spacer that occurs before the B subrepeat region. Comparison of the 5S coding sequence with a previously determined 5S coding sequence from RNA transcripts (Hori *et al*., 1985; GenBank accession number M10433) revealed a highly similar, but not identical, 5S sequence. The sequence was 120 bp in length and was 98% identical to the 5S sequence of *Ginkgo*, with two transitional mutations (at positions 35 and 48 of the alignment, encoding T-to-C and A-to-G mutations, respectively) and one cytosine insertion at position 104. The sequenced 5S gene contained most of the universally conserved motifs found in plant 5S sequences, including the 5′ (GGG) and 3′ ends (CCTC), the GC motif located four nucleotides upstream from the CCTC end, a slightly modified angiosperm box A sequence (AGTTAAGCACGC, instead of AGTTAAGCGTGC), and the strictly conserved box C sequence (GGATGGGTGACCTCCTG). Both the RNA and DNA 5S sequences lacked similarity to the intermediate element, GCGAGA, located in flowering plants between the A and C boxes. Moreover, both the upstream and downstream 5S spacers contained several functional elements involved in gene transcription, including a putative TATA box (ATAATAA) located 33 nucleotides upstream of the 5′ end of the 5S gene and a prymidine-rich stretch (TCTTCCTTTTT) involved in efficient transcription termination located just downstream of the 3′ end of the gene. The latter motif was duplicated after 9 bp of the first copy of the sequence, and it preceded a 10 bp duplicated sequence at the 3′ end of the 5S gene (TTGCACCCTC). The single deviant motif in the upstream spacer region lacked the universal cytosine at −1 nucleotide upstream of the 5′ end of the 5S gene, as it was replaced with a thymine. Patterns of retrotransposon sequences or other autonomous mobile elements were not detected in the flanking 5S gene spacers or elsewhere in the whole IGS fragment.

A single 45S–5S organization occurs within the *Ginkgo* genome

PCR experiments were performed to assess whether different 5S rRNA gene organizations and copy numbers were present in the ribosomal 45S–5S units in the *G. biloba* genome. Attempts to amplify tandem 5S units, by PCR with universal primers anchored in the gene, repeatedly failed to produce positive results. This outcome may have suggested that any 5S array composed of at least two consecutive 5S gene units, either within the 45S–5S unit or as pure 5S arrays, were not present in the genome. Furthermore, PCR experiments designed to assess whether 45S–5S units contained the 5S gene in an inverted orientation, with respect to that found in the 6.0 kb IGS variant, did not produce positive results. These results suggest that a single structure of the 45S and 5S rRNA genes is present in the *Ginkgo* genome.

Chromosomal colocalization of the 45S and 5S rRNA genes

Consistent with previous results (Nakao *et al*., 2005), four 5S and 45S rDNA signals colocalized to the same positions on two chromosome pairs in the previously identified nucleolar organizer regions (chromosome pairs 1 and 7, according to Nakao *et al*. (2005)). The 1.8 kb IGS clone lacking the 5S gene was used as a probe to assess whether this variant was present in the 45S–5S ribosomal unit or elsewhere in the *Ginkgo* genome. FISH experiments, using a fluorescent antibody enhancer designed to improve the sensitivity of the detection signal, showed that the 1.8 kb IGS variant was present exclusively on the satellite chromosomes bearing the 45S–5S genes, without detectable minor signals in other genomic positions (Figure 3).

DISCUSSION

Comparison of the structure and organization in IGS in *Ginkgo* and flowering plants

While considerable, mainly phylogenetic, effort has been devoted to the sequencing of a small region of the IGS (the 5′-ETS), full

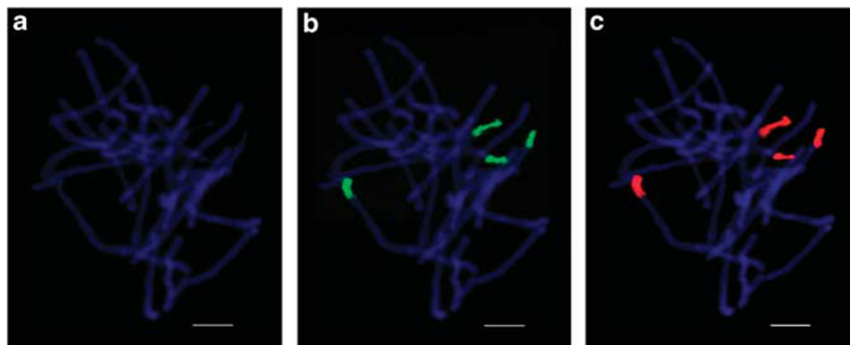

Figure 3 FISH localization of rDNA on mitotic metaphase chromosomes derived from root tips of *G. biloba*. (**a**) 4′,6-Diamidino-2-phenylindole (DAPI) staining was used to visualize the 24 chromosomes via counterstaining the chromosomal DNA. (**b**) *In situ* hybridization of the 1.8 kb variant is shown as green signals at four terminal sites. (**c**) FISH was conducted with a 5S rDNA probe for hybridization, and fluorescence was detected as red signals at four terminal sites. The 1.8 kb IGS rDNA and the 5S rDNA are colocalized exclusively at the terminal nucleolar organizer regions (NOR). Scale bar: 10 μm.

sequences of the ribosomal 45S IGS are available for less than a hundred out of the approximately 250 000 extant species of seed plants.

Because the IGS of *Ginkgo* described in this work is the first characterized putative functional structure from gymnosperms, caution should be exercised when considering it as a standard model for the whole group. In fact, the length of the putative functional IGS isolated from *Ginkgo* (6.0 kb) is far less than the estimated 20 kb unit of *Pinus sylvestris* (Karvonen *et al.*, 1993). By contrast, it fits well within the length ranges of known angiosperms, which range from 1.8 kb in *Glycine soja* (Nickrent and Patrick, 1998) to approximately 12.2 kb in *Trillium* species (Rogers and Bendich, 1987). In addition, the putative 5′-ETS region of *Ginkgo* (1641 bp) is similar in length to the ETS regions found in some angiosperms and bryophytes (Doelling *et al.*, 1993; Sone *et al.*, 1999; Fernández *et al.*, 2000).

Moreover, the presence of motifs and domains involved in the critical steps of start and finalization of the rRNA transcription (including rDNA promoters, promoter/enhancer elements, transcription termination site) and the TATA box involved in the transcription by RNA polymerase I have been recognized to be present in the 6.0 kb IGS.

Similar to the IGS characterized for most vascular plant species, the promoter region appears to be located between the repetitive areas (between blocks B and C, see Figure 1). The hypothesized TIS sequence in *Ginkgo* departs from the putative consensus promoter sequence reported for plants, that is, TATA(R)TA(N)GGG (Perry and Palukaitis, 1990; Doelling *et al.*, 1993). However, it shows resemblance to other putative, deviant angiosperm (*Quercus petrae/robur*, TCTTTAGGGGGG; *Hordeum bulbosum*, CATATAGCGCT) and bryophyte promoters (*Marchantia polymorpha*, TATATGAGGGG; *Funaria hygrometrica*, TACTATGTGGGG/GATAGGGGG). Increasing evidence suggests that restrictions on the TIS primary sequence could be much more relaxed than previously thought (Fan *et al.*, 1995; Capesius, 1997; Sone *et al.*, 1999; Maggini *et al.*, 2008; Bauer *et al.*, 2009).

The putative TIS in *Ginkgo* is preceded by a region composed of several AT-rich sequences (within the B repetitive block, at position −173), and that could form the domain of the promoter sequence known as the upstream binding elements, which is known to confer species specificity to polymerase I (Gerstner *et al.*, 1988; Zentgraf and Hemleben, 1992; Fan *et al.*, 1995). It has been noted that the promoter sequence in long IGS regions are usually duplicated or contain multiple repeats that enhance the polymerase activity via an increase in transcription initiation (Doelling *et al.*, 1993; Capesius, 1997; Bauer *et al.*, 2009). The lack of promoter duplication in *Ginkgo*, and its associated enhancement of transcription, might be counteracted by the presence of large repeat units (comprising several AT-rich motifs that are duplicated throughout the entire B box), which occur proximal to the promoter region. They have been hypothesized to serve as additional specificity enhancers for protein binding, playing an important role in enhancing transcription activity (Maggini *et al.*, 1992; Bauer *et al.*, 2009). A short sequence, CACAAACACAAA, found 98 nucleotides upstream of the TIS, was similar to the angiosperm motif CCAAAAAAGA, which has been implicated in protein binding (Bauer *et al.*, 2009).

The pyrimidine-rich motif sequence, CCCCCACCCC, found at the 5′ end of the IGS (3′-ETS) was highly similar to, but shorter than, a portion of the conserved motif present at the beginning of the IGS of some plant species (Kelly and Siegel, 1989; Perry and Palukaitis, 1990; Gruendler *et al.*, 1991; Borisjuk and Hemleben, 1992; Borisjuk *et al.*, 1997). This motif has been suggested to play a role in the finalization of the rRNA transcription as a transcription termination site.

The presence of ribosomal IGS pseudogenes in *Ginkgo* best reconciles contradictory results concerning the 45S–5S linkage

The results obtained in this work clearly suggest that the two ribosomal gene families have been unified to a single 45S–5S linkage unit in the genome. Our work corroborates previous findings based on molecular cytogenetic approaches, showing a juxtaposition of 45S and 5S fluorescent-based signals (Nakao *et al.*, 2005), suggesting, but do not probing, a colocalization of both multigene families in gymnosperms. On the other hand, we do not support the results of Wicke *et al.* (2011), who failed to detect such linkage in *Ginkgo*. Their results were based on the amplification and sequencing of three identical 2589 bp fragments (GenBank accession numbers FR 695705–FR 695707), lacking any 5S gene in the IGS. After trimming the 26S and 18S coding regions from their accessions, the resulting IGS is 1812 bp long, showing virtually the same length as our 1814 bp IGS variant. Furthermore, the comparison of both 1.8 kb variants shows the same structure and a 99.45% of sequence similarity, suggesting that they belong to the same IGS type unit.

Several lines of evidence, including their smaller length, lack of most of the modular structures present in the putative functional 6.0 kb IGS long (A and B boxes) and the absence of several critical motifs and domains hypothesized to be involved in structural and regulatory functions in plants and other eukaryotic species, strongly suggest that the 1.8 kb IGS variants isolated from *Ginkgo*, by Wicke *et al.* (2011) and ourselves, belong to a pseudogenized and non-functional 45S ribosomal cistron. The fact that the 1.8 kb IGS variant has not been amplified in the *Ginkgo* genome outside the 45S ribosomal loci (Figure 3) contrast with the results reported in other species, where parts of the IGS have been dispersed through the genome through drastic amplification processes (for example, Stupar *et al.*, 2002; Macas *et al.*, 2003). Colocalization of the 6.0 and 1.8 kb IGS variants within the same 45S locus suggest that the mechanisms of concerted evolution (Dover, 1994) are not fully operating in the ribosomal multigene family in this species.

Spurious results leading to incorrect conclusions have often been obtained when nuclear ribosomal pseudogenes are present in plant genomes and are not adequately identified during the experimental and analytical research stages (Mayol and Rosselló, 2001; Nieto Feliner and Rosselló, 2007). The results obtained by Wicke *et al.* (2011) for *Ginkgo* suggest that this is another case where the incorporation of unnoticed ribosomal pseudogenes could severely affect the reliability of the evolutionary signal recovered from nuclear ribosomal data (Mayol and Rosselló, 2001).

Early evolutionary colocalization of the 45S and 5S gene families in seed plants

The 45S–5S unit arrangements were not considered to be present in seed plants until recently, when García *et al.* (2009, 2010) detected this type of linkage in a wide sample of taxa belonging to the phylogenetically advanced Asteraceae family. Our finding that the 45S–5S loci are linked in *Ginkgo* suggests a colocalization of the two nuclear ribosomal gene families in the early evolutionary history of seed plants, indicating that a reassessment of the organization and molecular evolution of rDNA genes of basal plants is still necessary, as the results of Wicke *et al.* (2011) should be viewed with great caution.

The oldest *Ginkgo* fossils date back to Jurassic times (for example, *G. yimaensis* and *G. huttonii*, ca. 170 Mya; Zhou and Zhang, 2003),

and analysis of the reproductive organs of these plants has been used to infer that the evolutionary lineages of *Ginkgo* have experienced long periods of morphological stasis.

Molecular clock estimations suggest that *Ginkgo* diverged with the other gymnosperms excludes Cycadales more than 300 Mya (Won and Renner, 2006), about 20 Mya after the appearance of the crown group of seed plants (between 325 and 385 Mya; Won and Renner, 2006). Although these dates only estimate the maximum time since the of appearance of the 45S–5S linkage in *Ginkgo*, we speculate that the molecular events leading to the restructuring of the *Ginkgo* ribosomal loci are much older than the estimated age of the reunion of 45S and 5S gene families in angiosperms that likely occurred 7–10 Mya (García *et al*., 2010).

Molecular cytogenetic techniques, such as FISH, have not detected colocalization of the 45S–5S genes in any of the 31 analyzed species of Cycadales belonging to the gymnosperm families Stangeriaceae, Zamiaceae and Cycadaceae, but this could be due to lack of resolution of the technique, as the most finer fiber-FISH technique was not used. Although Cycadales have been described as basal group of extant seed plants and have an even deeper phylogenetic position than *Ginkgo* (Mathews, 2009), molecular data have convinced some authors that the Cycadales form a sister relationship with *Ginkgo* (Wu *et al*., 2007). Unfortunately, the presence of the 45–5S linkage in Ginkgoales, and its absence in Cycadales, further confounds which rRNA gene structure was present in the common ancestor of seed plants.

DATA ARCHIVING

Sequence data have been submitted to GenBank accession numbers JQ 279501–JQ 279502.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by funds from Spanish Ministry of Education and Science (Project No. CGL2010-22347-C02-01), the Catalan Government (Consolidated Research Group 2009SGR608) and by a PhD grant from Spanish Ministry of Education and Science to JA Galián. We thank G Nieto Feliner for insightful comments that improved the paper.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W *et al*. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res* **25**: 3389–3402.

Batts-Young B, Lodish HF (1978). Triphosphate residues at the 50 ends of rRNA precursor and 5S RNA from *Dictyostelium discoideum*. *Proc Natl Acad Sci USA* **75**: 740–744.

Bauer N, Horvat T, Birus I, Vicic V, Zoldos V (2009). Nucleotide sequence, structural organization and length heterogeneity of ribosomal DNA intergenic spacer in *Quercus petraea* (Matt.) Liebl. and *Q. robur* L. *Mol Genet Genom* **281**: 207–221.

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.

Borisjuk L, Hemleben V (1992). Nucleotide sequence of potato rDNA intergenic spacer. *Plant Mol Biol* **21**: 381–384.

Borisjuk NV, Davidjuk YM, Kostishin SS, Miroshnichenco GP, Velasco R, Hemleben V *et al*. (1997). Structural analysis of rDNA in the genus *Nicotiana*. *Plant Mol Biol* **35**: 655–660.

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

Del Tredici P (1991). Ginkgos and people: a thousand years of interaction. *Arnoldia* **51**: 2–15.

Doelling JH, Gaudino RJ, Pikaard CS (1993). Functional analysis of *Arabidopsis thaliana* rDNA gene and spacer promoters *in vivo* and by transient expression. *Proc Natl Acad Sci USA* **90**: 7528–7532.

Dover G (1994). Concerted evolution, molecular drive and natural selection. *Curr Biol* **4**: 1165–1166.

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

Fan H, Yakura K, Miyanishi M, Sugita M, Sugiura M (1995). *In vitro* transcription of plant RNA polymerase I-dependent rRNA genes is species-specific. *Plant J* **8**: 295–298.

Fernández M, Polanco C, Ruiz ML, Pérez de la Vega M (2000). A comparative study of the structure of the rDNA intergenic spacer of *Lens culinaris* Medik and other legume species. *Genome* **43**: 597–603.

Figueroa DM, Bass HW (2010). A historical and modern perspective on plant cytogenetics. *Brief Funct Genom* **9**: 95–102.

García S, Lim KY, Chester M, Garnatje T, Pellicer J, Vallès J *et al*. (2009). Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Chromosoma* **118**: 85–97.

García 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.

Gerlach WL, Dyer TA (1980). Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. *Nucleic Acid Res* **8**: 4851–4855.

Gerstner J, Schiebel K, von Waldburg G, Hembelen V (1988). Complex organization of the length heterogeneous 5′external spacer of mung bean (*Vigna radiata*) ribosomal DNA. *Genome* **30**: 723–733.

Golczyk H, Hasterok R, Joachimiak AJ (2005). FISH-aimed karyotyping and characterization of Renner complexes in permanent heterozygote *Rhoeo spathacea*. *Genome* **48**: 145–153.

Gruendler P, Unfried I, Pascher K, Schweizer D (1991). rDNA intergenic region from *Arabidopsis thaliana*: structural analysis, intraspecific variation and functional implication. *J Mol Biol* **221**: 1209–1222.

Hanson RE, Zwick MS, Choi S, Islam-Faridi MN, McKnight TD, Wing RA *et al*. (1995). Fluorescent *in situ* hybridization of a bacterial artificial chromosome. *Genome* **38**: 646–651.

Hasterok R, Jenkins G, Langdon T, Jones RN, Maluszynska J (2001). Ribosomal DNA is an effective marker of *Brassica* chromosomes. *Theor Appl Genet* **103**: 486–490.

Hori H, Lim B-L, Osawa S (1985). Evolution of green plants as deduced from 5S rRNA sequences. *Proc Natl Acad Sci USA* **82**: 820–823.

Jorgensen R, Cluster P (1988). Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies useful at several taxonomic levels. *Ann Missouri Bot Gard* **75**: 1238–1247.

Karvonen P, Karjalainen M, Savolainen O (1993). Ribosomal RNA genes in Scots pine (*Pinus sylvestris* L.): chromosomal organization and structure. *Genetica* **88**: 59–68.

Kelly RJ, Siegel A (1989). The *Cucurbita maxima* ribosomal DNA intergenic spacer has a complex structure. *Gene* **80**: 239–248.

Komarova NY, Grimm GW, Hemleben V, Volkov RA (2008). Molecular evolution of 35S rDNA and taxonomic status of *Lycopersicon* within *Solanum* sect. *Petota*. *Plant Syst Evol* **276**: 59–71.

Lafontaine DLJ, Tollervey D (2001). The function and synthesis of ribosomes. *Nat Rev Mol Cell Biol* **2**: 514–520.

Long EO, Dawid IG (1980). Repeated genes in eukaryotes. *Ann Rev Biochem* **49**: 727–764.

Macas J, Navrátilová A, Mészáros T (2003). Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on *Vicia sativa* chromosomes. *Chromosoma* **112**: 152–158.

Maggini F, Gelati MT, Spolverini M, Frediani M (2008). The intergenic spacer region of the rDNA in *Olea europaea* L. *Tree Genet Genomes* **4**: 293–298.

Maggini F, Tucci G, Demartis A, Gelati MT, Avanzi S (1992). Ribosomal RNA genes of *Phaseolus coccineus I*. *Plant Mol Biol* **18**: 1073–1082.

Maizels N (1976). *Dictyostelium* 17S, 25S, and 5S rDNAs lie within a 38 000 base pair repeated unit. *Cell* **9**: 431–438.

Major RT (1967). The *Ginkgo*, the most ancient living tree. *Science* **157**: 1270–1273.

Mathews S (2009). Phylogenetic relationships among seed plants: persistent questions and the limit of molecular data. *Am J Bot* **96**: 228–236.

Mayol M, Rosselló JA (2001). Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol Phyl Evol* **19**: 167–176.

Mizuochi H, Marasek A, Okazaki K (2007). Molecular cloning of *Tulipa fosteriana* rDNA and subsequent FISH analysis yields cytogenetic organization of 5S rDNA and 45S rDNA in *T. gesneriana* and *T. fosteriana*. *Euphytica* **155**: 235–248.

Muravenko OV, Amosova AV, Samatadze TE, Semenova OY, Nosova IV, Popov KV *et al*. (2004). Chromosome localization of 5S and 45S ribosomal DNA in the genomes of *Linum* L. species of the section *Linum* (Syn. *Protolinum* and *Adenolinum*). *Russ J Genet* **40**: 193–196.

Murray BG, Friesen N, Heslop-Harrison JS (2002). Molecular cytogenetic analysis of *Podocarpus* and comparison with other gymnosperm species. *Ann Bot* **89**: 483–489.

Nakao Y, Taira T, Horiuchi S, Kawase K, Mukai Y (2005). Chromosomal difference between male and female trees of *Ginkgo biloba* examined by karyotype analysis and mapping of rDNA on the chromosomes by fluorescence *in situ* hybridization. *J Jpn Soc Hort Sci* **74**: 275–280.

Nickrent DL, Patrick JA (1998). The nuclear ribosomal DNA intergenic spacers of wild and cultivated soybean have low variation and cryptic subrepeats. *Genome* **41**: 183–192.

Nieto Feliner G, Rosselló JA (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phyl Evol* **44**: 911–919.

Perry KL, Palukaitis P (1990). Transcription of tomato ribosomal DNA and the organization of the intergenic spacer. *Mol Gen Genet* **221**: 102–112.

Rogers SC, Bendich AJ (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol Biol* **9**: 509–520.

Rosato M, Castro M, Rosselló JA (2008). Relationships of the woody *Medicago* species (section *Dendrotelis*) assessed by molecular cytogenetic analyses. *Ann Bot* **102**: 15–22.

Royer DL, Hickey LJ, Wing SL (2003). Ecological conservatism in the 'living fossil' *Ginkgo*. *Paleobiology* **29**: 84–104.
Sastri DC, Hilu K, Appels R, Lagudah ES, Playford J, Baum BR (1992). An overview of evolution in plant 5S DNA. *Plant Syst Evol* **183**: 169–181.
Široký J, Lysák MA, Doležel J, Kejnovský E, Vyskot B (2001). Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Res* **9**: 387–393.
Snowdon RJ, Friedt W, Koèhler A, Koèhler W (2000). Molecular cytogenetic localization and characterization of 5S and 25S rDNA loci for chromosome identification in oilseed rape (*Brassica napus* L.). *Ann Bot* **86**: 201–204.
Sone T, Fujisawa M, Takenaka M, Nakagawa S, Yamaoka S, Sakaida M *et al.* (1999). Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. *Plant Mol Biol* **41**: 679–685.
Stupar RM, Song J, Tek AL, Cheng Z, Dong F, Jiang J (2002). Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. *Genetics* **162**: 1435–1444.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25**: 4876–4882.
Tralau H (1968). Evolutionary trends in the genus *Ginkgo*. *Lethaia* **1**: 63–101.
Vahidi H, Purac A, Leblanc JM, Honda BM (1991). Characterization of potentially functional 5S rRNA-encoding genes within ribosomal DNA repeats of the nematode *Meloidogyne arenaria*. *Gene* **108**: 281–284.
Vitturi R, Colomba MS, Pirrone AM, Mandrioll M (2002). rDNA (18S–28S and 5S) colocalization and linkage between ribosomal genes and (TTAGGG)(*n*) telomeric sequence in the earthworm *Octodrilus complanatus* (Annelida, Oligochaeta, Lumbricidae), revealed by single- and double-color FISH. *J Hered* **93**: 279–282.
Wicke S, Costa A, Muñoz J, Quandt D (2011). Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol Phyl Evol* **61**: 321–332.
Won H, Renner SS (2006). Dating dispersal and radiation in the gymnosperm *Gnetum* (Gnetales)—clock calibration when outgroup relationships are uncertain. *Syst Biol* **55**: 610–622.
Wu C-S, Wang Y-N, Liu S-M, Chaw S-M (2007). Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. *Mol Biol Evol* **24**: 1366–1379.
Zentgraf U, Hemleben V (1992). Complex formation of nuclear proteins with the RNA polymerase I promoter and repeated elements in the external transcribed spacer of *Cucumis sativus* ribosomal DNA. *Nucl Acids Res* **20**: 3685–3691.
Zhong X, Fransz PF, Wennekes-Van Eden J, Zabel P, Van Kammen A (1996). High-resolution mapping on pachytene chromosomes extended DNA fibres by fluorescence *in situ* hybridisation. *Plant Mol Biol Rep* **14**: 232–242.
Zhou Z, Zhang S (2003). The missing link in *Ginkgo* evolution. *Nature* **423**: 821–822.