Research Signpost
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Recent Advances in Pharmaceutical Sciences VI, 2016: 49-75 ISBN: 978-81-308-0566-5
Editors: Diego Muñoz-Torrero, Àngela Domínguez and Àngels Manresa

4. Organisation, expression and evolution of rRNA genes in plant genomes

Sònia Garcia[1,2], Paula Cortés[1], Xènia Fernàndez[1], Teresa Garnatje[2] and Aleš Kovařík[3]

[1]Laboratori de Botànica – Unitat associada CSIC, Facultat de Farmàcia, Universitat de Barcelona Avinguda Joan XXIII s/n, 08028 Barcelona, Catalonia, Spain; [2]Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain
[3]Department of Molecular Epigenetics, Institute of Biophysics, ASCR, Kralovopolská 135 CZ-61265 Brno, Czech Republic

Abstract. Here we present an overview on the results of eight years of research line devoted to the organisation, expression and evolution of rRNA genes in plant genomes. We describe how we discovered a new rDNA arrangement in genus *Artemisia*, which includes all rRNA genes in a single unit (L-type). This was the first time that such organisation was found in seed plants. We further explored family *Asteraceae* in depth, to which Artemisia belongs, to find that the L-type arrangement may be present in about 25% of its species. Later on we move to gymnosperms to describe the landscape of rRNA arrangements in a representative sample of its diversity. We assess the expression rate of the L-type rRNA in several L-type species, which is comparable to that of species with separated arrangement of rRNA genes (S-type). Finally, we present the resource www.plantrdnadatabase.com which includes information on type of rDNA arrangement, number and position of rDNA loci in plants.

Correspondence/Reprint request: Dr. Sònia Garcia, Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, 08038 Barcelona, Spain. E-mail: soniagarcia@ibb.csic.es

Introduction

It is broadly accepted that a previous RNA world predated the present one based on DNA, RNA and proteins, probably thanks to the ability of RNA to both store genetic information and to act as an enzyme [1]. However, the mechanism by which RNA turned into the DNA system is still unknown. Nevertheless, several facts, such as the central role of RNA in the translation process and the ability to self-replicate, support this hypothesis. Notably, many RNA molecules, known as ribozymes or RNA enzymes, can catalyse biochemical reactions. Indeed, there are plenty of ribozymes in present-day cells, which could be considered as living fossils. Although ribosomes are composed by RNA and proteins, the catalytic site is formed by RNA only; besides, the deciphering of the 3D structure of ribosome revealed that peptide bond formation (the chemical reaction binding amino acids to form proteins) is catalysed by an adenine residue in the ribosomal RNA [2].

Ribosomal RNAs comprise 80% of the RNA found in a typical cell. The different types of ribosomal RNA (rRNA) account for 60% of the mass of the ribosomes. Ribosomal RNAs are codified by rRNA genes (rDNA), and there can be from 50 to 50,000 copies of rRNA genes per cell. In eukaryotes there are four ribosomal RNA genes. The 18S, 5.8S, and 26S rRNA genes are encoded in a single operon (called 35S rDNA in plants or 45S rDNA in animals and the 5S rRNA gene is encoded outside this operon, by 5S rDNA. Ribosomal DNAs are usually tandemly organised forming long arrays on one or several chromosomal loci. While in prokaryotes all rRNA genes form the same operon and the same RNA polymerase transcribes all of them, in some species of early diverging eukaryote groups e.g. yeast, *Saccharomyces cerevisiae*, although all rRNA genes are organised in a single unit and clustered in tandem, the 18S, 5.8S and 26S rRNA genes are transcribed by RNA polymerase I and the gene for 5S rRNA is transcribed by RNA polymerase III (Pol III). In eukaryotes, the 35S/45S (encoding 18S-5.8S-26S rRNA) and 5S genes are usually arranged separately and they are also transcribed by different polymerases, RNA polymerase I and III, respectively. Typically, in angiosperms, the tandemly arranged 35S rDNA unit is 9–20 kb long occurring in several thousands of copies at one or multiple chromosomal loci [3]. The unit comprises a ~5.0-kb genic region encoding for 18S–5.8S–26S genes and a more variable intergenic spacer (IGS). The 5S rDNA unit is much shorter and is composed of a 120-bp genic region and a 100–1,000-bp spacer [4]. Both the 35S and 5S rRNA genes are

tandemly arranged forming clusters of thousands of regularly spaced units. Figure 1 summarises the possible organisation of rRNA genes in different organisms.

Because of their abundance and sequence conservation, the nuclear genes encoding rRNA have been the subject of much research amongst molecular biologists and cytogeneticists, particularly for inferring evolutionary relationships between species. There are high levels of sequence homogeneity both with coding and non-coding parts of rDNA units within 5S or 35S rDNA loci, a phenomenon known as concerted evolution. Concerted evolution is the process by which the members of a multigene family are homogenised within a species but are different between species.

Independent control of transcription probably enabled physical separation of both loci in chromosomes, an arrangement that is typical for most eukaryotic organisms. Nevertheless, it seems that there are several exceptions to this rule. For example, 5S genes linkage to other repetitive sequences including 35S, histone genes or the trans-spliced leader has been

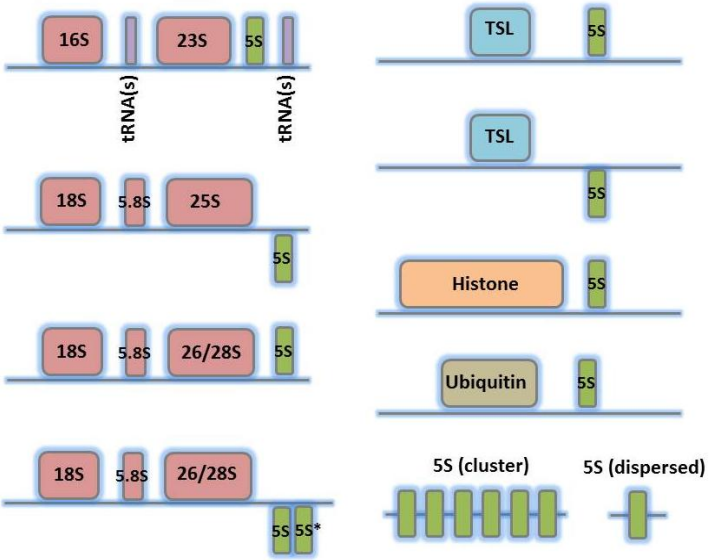

Figure 1. Arrangement of rRNA genes within (or outsides) the tandem repeat units of various multigene families found in different taxonomic groups. TSL=trans-spliced leader : (*) Pseudogene.

demonstrated [5]. These linked arrangements are found among diverse biological taxa including nematodes [6], fungi [7], crustaceans [8], slime moulds [9] or mosses [10, 11], and they are believed to represent transition states between linked (prokaryotic) and unlinked (mostly eukaryotic) arrangements.

1. Organisation of ribosomal RNA genes in *Artemisia*

As pointed previously, the 5S and 35S ribosomal DNA (rDNA), encoding the four major ribosomal RNA species, occur at separate loci in most plants. In some algae, bryophytes and ferns, they are at the same locus (linked arranged) but such molecular organisation had not been described in any angiosperm until our work in genus *Artemisia*. Our research group has been interested for more than 20 years in the evolutionary aspects of this genus [12], one of the most interesting within the family Asteraceae. Genus *Artemisia* is found in the most evolutionary advanced part of the family, belonging to subfamily Asteroideae and tribe Anthemideae. It consists in about 500 species distributed mainly in the Northern hemisphere, with few representatives in the Southern [13]. *Artemisia* species are mostly perennial and landscape dominating in grasslands and semidesert areas of Asia and North America [14]. Certain species (Figure 2) have medical interest, such as *Artemisia annua*, the source of artemisinin, a powerful treatment against malaria. Others are used as a condiment, such as *A. dracunculus*, (tarragon) or as the source of alcoholic beverages such as absinth (*A. absinthium*). Many species are used as forage and some can behave as weeds (*A. vulgaris*). The genus is quite well known from the evolutionary point of view (for a review see [12]) and plenty of karyological and cytogenetic studies have been devoted to the genus. Polyploidy and dysploidy are common in the genus, with ploidy levels up to 16x.

The first molecular cytogenetic study performed in *Artemisia* showed the pattern of ribosomal RNA genes localisation by fluorescent *in situ* hybridisation (FISH), in the *Artemisia campestris* complex [15]. In that study it became evident that both 5S and 35S rDNAs were colocalised, that is, the rDNA probes hybridised in the same chromosomal location, at the end of several submetacentric chromosomes. A further study by [16] deepened in the apparently anomalous organisation of rRNA genes in *Artemisia* and using Southern blot hybridisation, polymerase chain reactions (PCR), fluorescent *in situ* hybridisation, cloning and sequencing the genomic organisation of 5S and 35S rDNA in *Artemisia* was revealed (Figures 3 and 4).

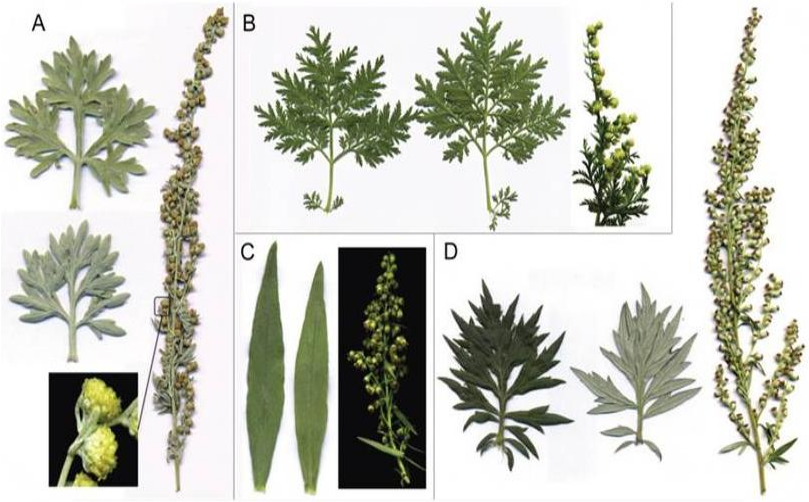

Figure 2. Leaves and synflorescences of four *Artemisia* species. (A) *Artemisia absinthium*. (B) *Artemisia annua*. (C) *Artemisia dracunculus*. (D) *Artemisia vulgaris*. From [12].

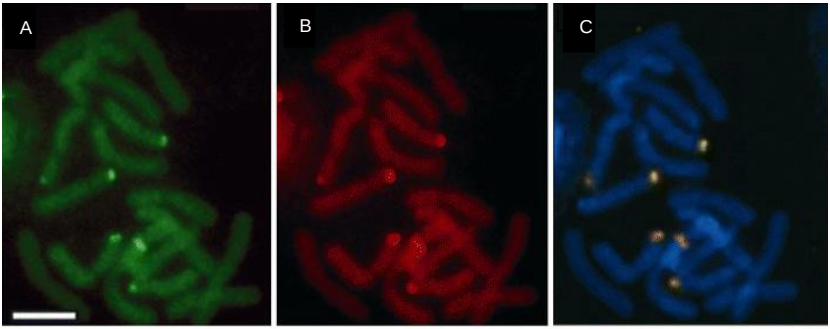

Figure 3. FISH on *Artemisia tridentata* showing the 35S probe (A, green) and the 5S probe (B, red) hybridising at the same loci. C, composite image including DAPI (blue) staining. Scale bar 10 µm. Pictures extracted from [16].

In *Artemisia*, the rRNA genes are all arranged in a single operon containing the 35S genes and the 5S gene in inverted orientation within the large intergenic spacer (IGS) of the 35S rDNA. The 5S gene insertion is typically located next to (about 150 bp) the end of the 26S gene. Also there

is a second less conserved insertion further downstream. The first 5S gene insertion presents all structural features of a functional gene, while the second is most likely a pseudogene since it had a deletion in the internal promoter region (Figure 4). Heterogeneity in unit structure may reflect ongoing homogenisation of variant unit types without fixation for any particular variant.

The study investigated several species within the genus and in all cases the same organisation of rRNA genes was inferred. Besides, closely and not so closely related genera, such as *Ajania* or *Nipponanthemum*, did also present the same arrangement (Figure 5). We hypothesized therefore that the linked arrangement probably evolved before the divergence of *Artemisia* from the rest of Asteraceae (>10 Myrs). Most likely, the activity of transposable elements is involved in this particular arrangement of rRNA genes and, probably, processes of concerted evolution took care of the homogenisation in a given genome (since no unlinked units were detected in *Artemisia*).

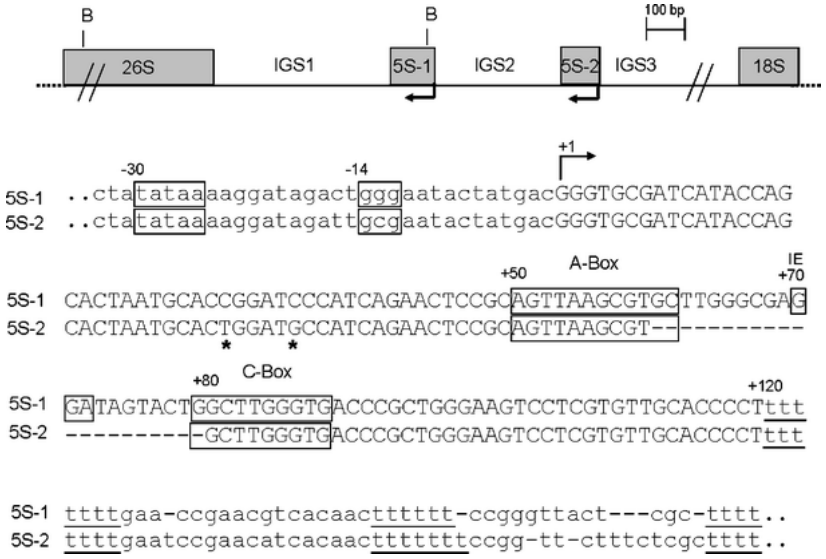

Figure 4. Representation of the 35S intergenic spacer of *Artemisia* containing two 5S copies. While the first one (5S-1) is a complete 5S rRNA gene, the second (5S-2) lacks part of the sequence, including essential regulatory regions, which indicate that it may be a pseudogenised copy. Extracted from [16].

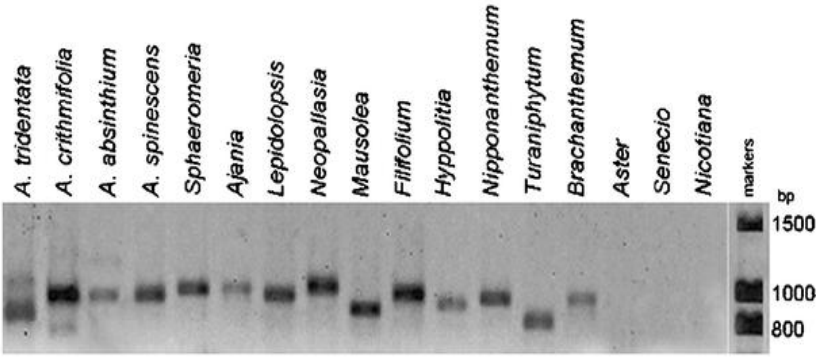

Figure 5. Gel electrophoresis of PCR products obtained after amplification of genomic DNAs with the 5SL-F and 26S-F primers. The sizes of products ranged between 800–1,000 bp. Extracted from [16].

2. Organisation of rRNA genes in family Asteraceae

From the observations established that both rRNA genes have been unified to a single 35S-5S unit in the genus *Artemisia*, we further aimed to reveal the extent, distribution and mechanisms leading to the linked organisation of rDNA in the Asteraceae, the family to which genus *Artemisia* belongs. The Asteraceae or Compositae are the largest angiosperm family with 1,620 genera and 23,600 species [17], and Asteraceae species embrace about 8-10% of all flowering plants. A majority of species of Asteraceae are herbaceous although some trees exist in the family (particularly in the American continent). Asteraceae are easily recognizable by certain morphological characters including fused anthers (syngenesia), a fruit with a single ovule (cypsela), and, in particular, by a specialized inflorescence termed capitulum (Figure 6). The family is widespread and cosmopolitan, growing everywhere but in the Antarctica [17]. Many species are edible and have huge economic interest (sunflowers, lettuces, artichokes, etc.), others are medicinal (calendula, chamomile, etc.) but several of them are also noxious weeds or invasive (wild safflower, cardoon, etc.). Several phylogenetic analyses have been conducted on the family, considered to hold 12 major subfamilies, four of which (Asteroideae, Carduoideae, Cichorioideae and Mutisioideae) comprise 99% of its species diversity [17].

Figure 6. A poster of twelve kinds of capitula of the Asteraceae (not representing their whole diversity), belonging to the two most representative subfamilies: Asteroideae and Cichorioideae: (1) yellow chamomile, *Anthemis tinctoria* (Asteroideae); (2) crown daisy, *Glebionis coronarium* (Asteroideae); (3) *Coleostephus myconis* (Asteroideae); (4) chrysanthemum - *Glebionis* sp. (Asteroideae); (5) sow thistle - *Sonchus oleraceus* (Cichorioideae); (6) chicory - *Cichorium intybus* (Cichoireideae); (7) treasure flower - *Gazania rigens* (Cichorioideae); (8) Mexican sunflower - *Tithonia rotundifolia* (Asteroideae); (9) field marigold - *Calendula arvensis* (Asteroideae); (10) Ox-eye daisy - *Leucanthemum vulgare* (Asteroideae); (11) common hawkweed - *Hieracium lachenalii* (Cichorioideae); (12) Cape daisy - *Osteospermum ecklonis* (Asteroideae). Picture by Dori Alvesgaspar and Tony Wills, licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

Although the family is so vast, little had been published about rDNA position, organisation and structure in the whole group. Specific research works centered in certain genera, such as *Tragopogon*, *Centaurea*, *Helianthus* and *Hipochaeris*, among others [18, 19, 20, 21, 22], showed a separate organisation of rRNA genes, in contrast with the situation found in *Artemisia*. Considering the large size of the family, however, we still did not know how frequently an arrangement as the one in *Artemisia* could be found, and this was the purpose of our study. Using molecular (PCR, Southern blot, sequencing, quantitative PCR) and cytogenetic (FISH) methods, we examined ribosomal DNA structure and organisation in selected species (about 200) representing Asteraceae diversity [23].

The linked rDNA arrangement (Figure 7) was found within three large groups in subfamily Asteroideae: tribe Anthemideae (93% of the studied cases), tribe Gnaphalieae (100%) and in the "Heliantheae alliance" (23%). The remaining five tribes of the Asteroideae displayed separated arrangement of rDNA (Figure 8), as did the other groups in the Asteraceae.

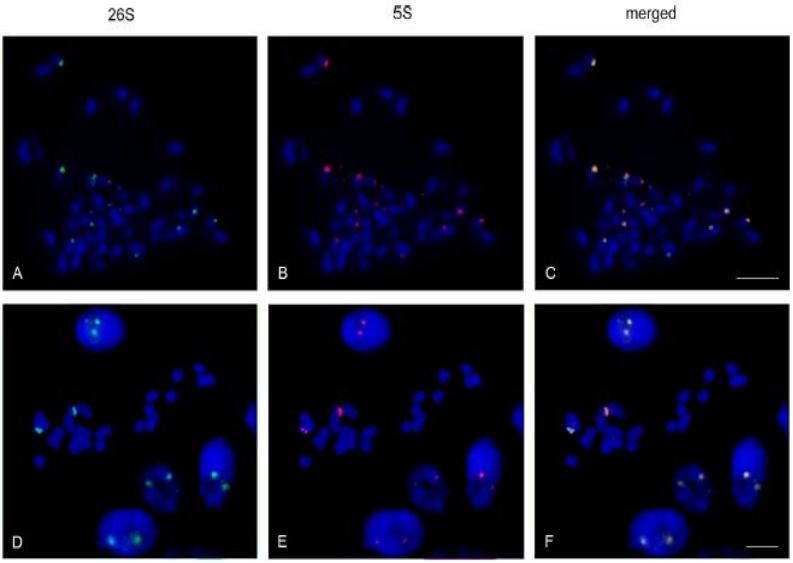

Figure 7. FISH of metaphase chromosomes and interphase nuclei of species that evolved linked arrangement of rRNA genes: (A-C) *Tagetes patula* (2n = 48) and (D-F) *Helichrysum bracteatum* (2n = 22). The 35S (26S probe) and 5S loci are labelled in green and red, respectively. Pictures extracted from [23].

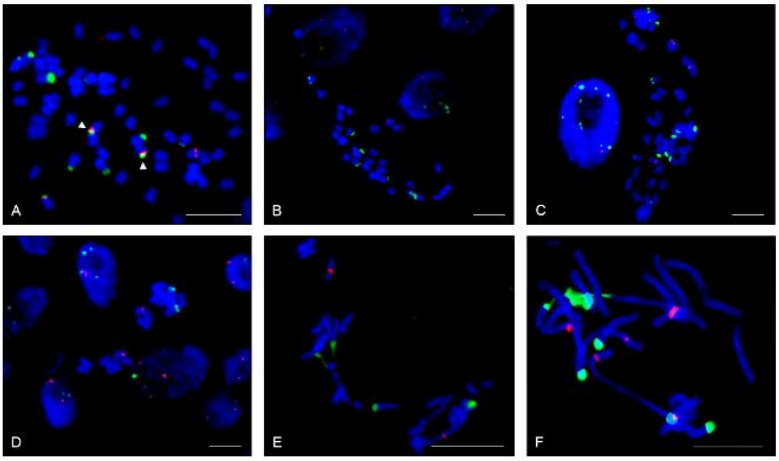

Figure 8. FISH of metaphase chromosomes and interphase nuclei of species that evolved unlinked arrangement of rRNA genes. The individual slides show merged 26S and 5S signals: (A) *Dahlia pinnata* (2n = 64), (B) *Helianthus annuus* (2n =34); (C) *Chrysanthemum zawadskii* (2n = 54), (D) *Aster alpinus* (2n = 18); (E) *Calendula officinalis* (2n = 28); (F) *Tragopogon mirus* (2n = 24). Arrowheads in (A) indicate juxtaposition of 35S and 5S arrays. The 35S (26S probe) and 5S loci are labelled in green and red, respectively. Pictures extracted from [23].

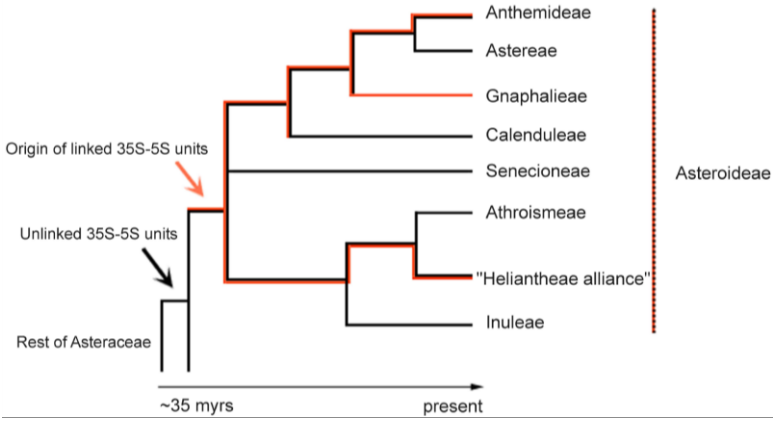

Figure 9. Phylogenetic distribution of rDNA arrangement in the subfamily Asteroideae. Red and black lines indicate evolutionary trajectories of linked and unlinked units, respectively. Extracted from [23].

Therefore, two contrasting modes of rDNA organisation are present in the Asteraceae. According with the distribution of these L-type genera in the phylogeny it seems that the linked arrangement arose independently several times in the largest subfamily, Asteroideae (Figure 9).

Yet, there are differences regarding the tribe; for example, 100% of studied species from Gnaphalieae present the L-type arrangement but conversely, a variable number of species with unlinked rRNA genes were found in Anthemideae (7%) and in the "Heliantheae alliance" (77%). This variability is reminiscent to that of primitive eukaryotes (yeast, other fungi) in which 5S genes are linked either with other rDNAs or other repetitive genes. Indeed, differences in the rDNA organisation are found even between closely related genera such as *Elachanthemum* (S-type) and *Artemisia* (L-type).

Therefore, could rDNA arrangement have a potential use as a phylogenetic tool? Since two closely related genera such as the previously mentioned present contrasting rDNA organisations, the most intuitive answer would be no. Nevertheless, on the basis of the L-type arrangement shared by most Anthemideae, Gnaphalieae and members of the "Heliantheae alliance", perhaps phylogenetic relationships between these three groups should be reassessed, as they could be more closely related than currently considered in Asteraceae phylogenetic treatments [24]. Moreover, the sequence between 26S and 5S genes in L-type species, termed IGS1 (see Figure 4), is not as conserved as the genic regions, displaying significant length and sequence polymorphism. When this sequence is used for a phylogenetic reconstruction, the results are congruent with one of the most used region for inferring phylogenies, the ITS1. For example, between *Artemisia absinthium* and *Artemisia tridentata*, the divergence of ITS1 was around 4%, with 10 substitutions in 250 bp while for IGS1 the divergence was around 7%, with 35 substitutions per 488 bp. Therefore the latter could be used as a suitable marker to infer phylogenetic relationships within the L-type species (the equivalent region in S-type species remains to be tested).

Since the separate organisation of rRNA genes is the most common across eukaryotes, it is likely that this is the most primitive one from which the linked arrangement evolved, although at present this still remains to be proved. According to our results, about 25% of the Asteraceae would present exclusively the linked arrangement; this means that at some point(s) the linked arrays would have overwritten the unlinked ones through concerted evolution. Some hypotheses explaining the mechanisms

of rDNA rearrangements would imply DNA recombination and/or RNA mediated transposition [25].

According to the results in Asteraceae, it is possible that the linked rDNA arrangement is more common than previous assumptions, which considered this organisation as an exceptional finding. Further in-depth studies at the family level are needed, not only to detect similar rDNA arrangements within plant families but also to find the evolutionary advantage, if any, that a linked arrangement of rRNA genes would imply for the organisms. The next step in our quest on rDNA organisation was an in-depth look to gymnosperms' rDNAs, since previous data pointed to certain prevalence of the L-type arrangement in this group.

3. Organisation of rRNA genes in gymnosperms

Gymnosperms were a rather unexplored field for rDNA research at the moment in which our paper dealing with rDNA organisation in these plants was released [26]. Gymnosperms, despite their large diversity in morphology and ecology, are only 14 families and about 1,000 species. This is much reduced in comparison with angiosperms (estimated to be about 352,000 species) or even with certain of its families, like the previously mentioned Asteraceae or the Orchidaceae, both with more than 20,000 species each [26].

There are four orders classically considered in gymnosperms: Pinales, Cycadales, Gnetales and Ginkgoales. The former is the largest and most abundant in terms of species and coverage; it includes about 650 species mostly belonging to the division Pinophyta. Cycadales are the second largest and they currently include three families and 130 species; Gnetales hold three families and 70 species and the latter, Ginkgoales, has a single extant species, *Ginkgo biloba*.

Most molecular cytogenetics research work on gymnosperms has been indeed carried out in family Pinaceae, the most widespread of the whole group. The arrangement of rRNA genes found in most gymnosperms up to date was the separate organisation, with 5S and 35S rRNA genes located in different positions, and usually in different chromosomes. However, the non-Pinaceae gymnosperms are more variable in rDNA organisation. In *Podocarpus* (family Podocarpaceae and still within Pinales, as Pinaceae) FISH revealed colocalisation of 5S and 35S rRNA genes in all loci [27]; similar results were found in certain Cupressaceae, including *Cryptomeria japonica* [28] and in *Ginkgo biloba* [29]. These studies, however, did not

show whether the 35S and 5S were physically linked (as in *Artemisia* and other Asteraceae, see parts 1 and 2 of this article), because rDNA *in situ* hybridisation in metaphase chromosomes does not allow enough resolution to interpret the molecular organisation.

Additionally, PCR screens and sequence analysis revealed a separate organisation of rRNA genes in *Cycas*, *Ginkgo* and *Gnetum* [30]. On the basis of the latter and other studies, it had been proposed that in gymnosperms the 5S and 35S repeats hold the separated configuration typical for vascular plants. Nevertheless, [29] found (using a PCR-based strategy) that in *G. biloba* 5S genes were integrated in the 26S-8S spacer. However, there were also rDNA clones without 5S genes, pointing to some heterogeneity of rDNA arrays. Although *in situ* hybridisation results favoured a dominant linked arrangement, it was unclear which was the proportion of separate units (if any). With this in mind, as well as with the purpose of drawing the global picture of rDNA arrangements in gymnosperms, we conducted an overall study on rDNAs of representatives of the main groups of gymnosperms, with an approach that included PCR, cloning, sequencing and Southern blot hybridisation [26]. Our study provides coverage of basic rDNA organisation for at least one representative of all gymnosperm orders, as well as a 15% increase in coverage of genera and an almost 30% increase in coverage of families of gymnosperms.

3.1. Uniform S-type rDNA arrangement in Pinaceae (Pinales) and Cycadales

Most previous molecular cytogenetic studies had been previously focused on family Pinaceae as it holds economically and ecologically interesting genera such as *Abies*, *Pinus* or *Picea*. Partially colocalising 5S and 35S rDNAs had been observed on chromosomes of *Abies alba* [31], *Picea* [32] and *Pinus* [33] species. In our survey, however, Southern blot did not reveal cohybridisation of 35S (using a 18S probe) and 5S probes in any of these, therefore it is likely that 5S and 35S rDNA signals are one next to the other, but not physically linked. Fluorescent *in situ* hybridisation performed to the species *Cupressus dupreziana* also showed separate arrangement of rRNA genes (Figure 10). Besides, Southern hybridisations showed quite homogeneous signal pattern (Figure 11), possibly the consequence of interloci homogenisation, i.e. efficient concerted evolution acting on these genes.

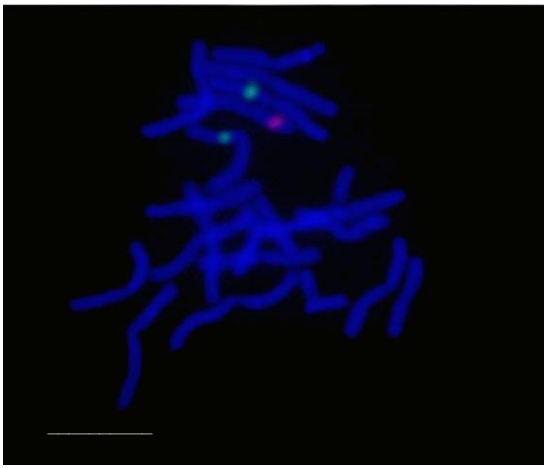

Figure 10. FISH to *Cupressus dupreziana* chromosomes (2n=22). Two 35S sites depicted in green and a single 5S site depicted in red. Scale bar 10 µm.

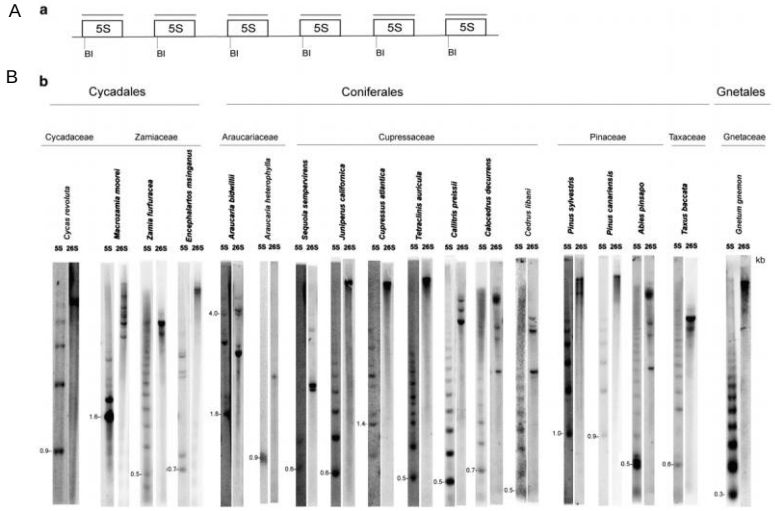

Figure 11. Southern blot hybridisation analysis of species with S-type arrangement of 5S and 35S units. The 5S tandem arrays with conserved *Bam*HI sites and probe hybridisation regions (lines above boxes) are schematically drawn in (A). Genomic DNAs were digested with *Bam*HI and hybridised on blots with the 5S and 26S rDNA probes (B). The sizes of 5S monomeric units are indicated. Extracted from [26].

The representative of Cycadales in this sample also shows a pattern consistent with a separate arrangement of rRNA genes. In this regard, the fact that recent molecular phylogenetic approaches point towards cycadophytes being more ancient than *Ginkgo* [34, 35], with a linked rDNA arrangement, complicates the interpretation on the ancestral condition of rDNA organisation in seed plants.

3.2. Both S- and L-type rDNAs in non-Pinaceae Pinales and in Gnetales.

In this group, we observed certain variations in rDNA organisation. Most members of Araucariaceae, Cupressaceae and Taxaceae showed separation of both 5S and 35S loci (Figure 10), however genera *Podocarpus* and the closely related *Afrocarpus* (Podocarpaceae) show clear linkage of 26S and 5S units (Figure 12).

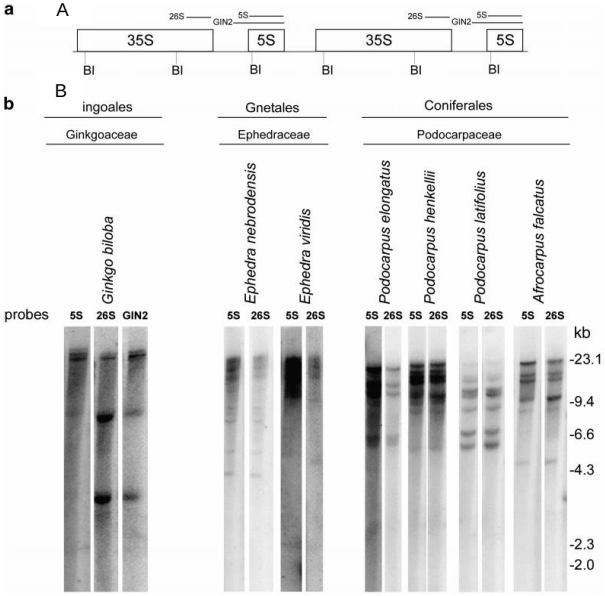

Figure 12. Southern blot hybridisation analysis of species with L-type arrangement of 5S and 35S units. Restriction map showing conserved *Bam*HI sites and regions of probe hybridisation (thin lines above boxes) are schematically drawn in (A). Genomic DNAs were digested with *Bam*HI and hybridised on blots with the 5S, 26S and GIN2 (containing IGS1 from *Ginkgo*) DNA probes (B). Extracted from [26].

The 35S–5S arrays are homogeneous and there is no evidence of separate loci. We estimated, based on Southern blot mapping and cloning, that the 5S insertions are found 6–9 kb downstream from the 26S gene and 42-kb upstream of 18S gene. In *Podocarpus* and *Afrocarpus*, it seems that the direction of transcription would be opposite as the rRNAs are encoded from complementary DNA strands (Figure 13). Also, we found cohybridisation of 26S, 18S and 5S probes in the Ephedraceae (Gnetales) species of our sample (Figure 12), consistent with FISH performed to the chromosomes of *Ephedra distachya* showing overlapping 5S and 35S signals (Figure 14). We also found additional, pseudogene-like copies, of 5S rDNA, together with one apparently functional 5S unit. According to Southern blot mapping and sequencing of *Ephedra* species, their rDNA units are similar to

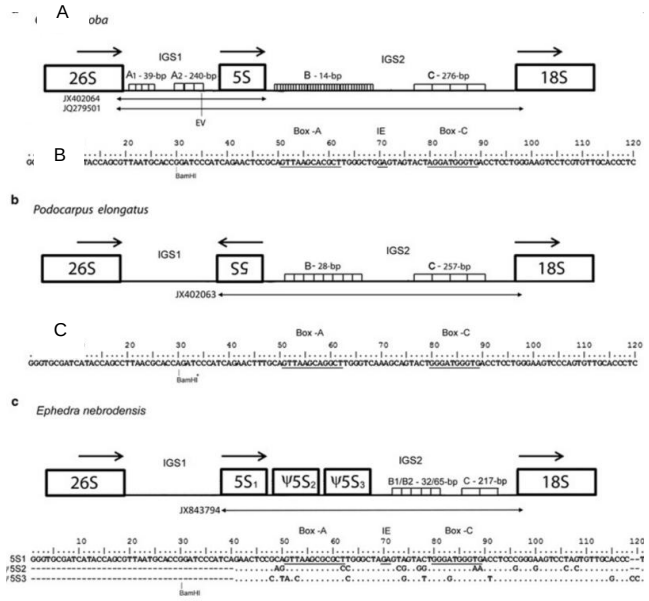

Figure 13. Schematic representation 35S–5S units in *Ginkgo biloba* (A), *Podocarpus elongatus* (B), and *Ephedra nebrodensis* (C). Arrows above boxes indicate direction of transcription. The lines with arrows represent clones carrying 5S insertions. Subrepeated portions of IGS are indicated and labelled as (A–C). The 5S coding region sequences are shown below each graph, with conserved regulatory elements underlined. The positions of *Bam*HI sites are indicated; in *Podocarpus* the site is mutated (asterisk). Extracted from [26].

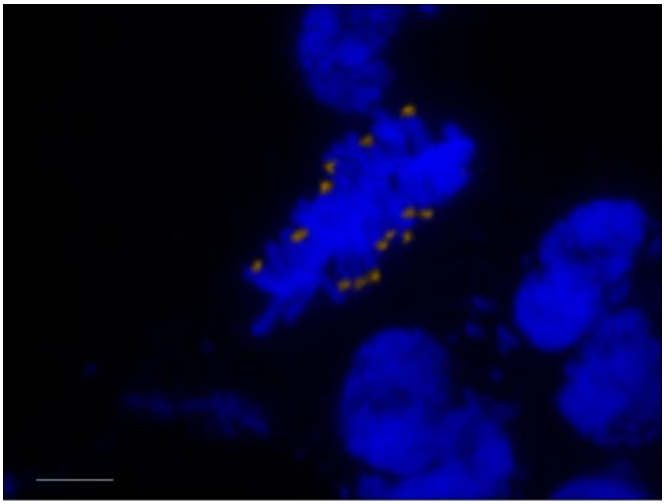

Figure 14. FISH to *Ephedra distachya* chromosomes (2n=28), showing 14 linked (5S-35S) rDNA signals. Scale bar 10 µm.

those of Pinaceae (Pinophyta) in many aspects, both being clearly different from the angiosperm pattern. Duplicated 5S copies had been also detected in other L-type species [26, 30]. However it seems that, again, not all Gnetales would display the L-type arrangement: the sequenced IGS clone from *Gnetum gnemon* appears to lack a 5S insertion (see Southern blot profile of the species in Figure 11, clearly showing separate 5S and 26S hybridization signals), but more research is needed to confirm this point.

3.3. Homogenised L-type rDNAs in Ginkgoales

Fluorescent *in situ* hybridisation showed linkage of 35S and 5S signals on *Ginkgo* chromosomes [28], but the PCR cloning analysis of [30] failed to demonstrate 5S linkage to the large rDNA cluster. In our assays, we also found a 5S insertion in direct orientation located downstream from the 26S gene (Figure 13), so the data support the hypothesis that the predominant organisation of *Ginkgo* rDNAs is L-type (Figure 12). Besides, the different orientation of 5S insertions in *Ginkgo* and *Podocarpus* suggests their independent origins.

The S-type arrangement dominates most angiosperm lineages [30] although in some families the L-type arrangement is quite frequent, as we

have shown. The phylogenetic relationships (Figure 15) indicate that the ribosomal RNA genes have changed their fundamental genomic organisation at least three times during the evolution of gymnosperms, similarly to our findings in Asteraceae.

Sequence analysis of 5S insertions shows interesting differences between angiosperms and gymnosperms. In angiosperms the 5S insertions occur in a non-repetitive part of the IGS quite close to the 26S gene, while in gymnosperms the 5S gene is located distally to the 26S and 18S genes, embedded in a highly repetitive DNA region as in bryophytes [11]. Up to our knowledge, the 5S genes of L-type angiosperms are encoded by the opposite DNA strand to the 35S genes while in gymnosperms both direct and inverse orientations occur. Moreover, retroelement signatures in close proximity of the 5S insertions in the IGS have been found in angiosperms [36], but not in gymnosperms.

Both cycads and *Ginkgo* are considered to be early diverging, living-fossil taxa [37] and it is remarkable that they show contrasting L- and S-type arrangements, respectively. Since both arrangements are also present in modern plants, it is likely that both are evolutionary neutral. However, the overwhelming prevalence of S-type arrangement in seed plants and vertebrates still needs to be explained.

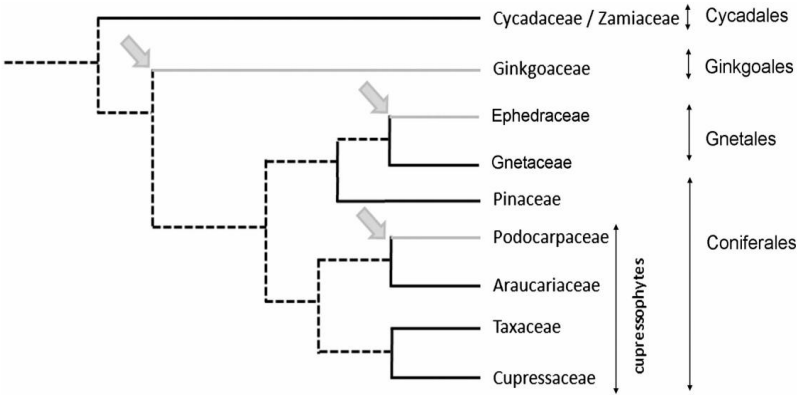

Figure 15. A proposed organisation of 5S–35S genes in gymnosperm phylogeny. Grey lines—L-type arrangement; black lines—S-type arrangement. Grey arrows show putative integration events of 5S genes into the 26S–18S IGS. Dashed lines—putative ancestors having either L- of S-type units. Extracted from [26].

4. Expression of rRNA genes in L-type arrangement

Despite numerous data confirming the existence of the L-type arrangement of rRNA genes in certain plants groups, the activity of such linked genes remained unknown. Therefore our next step was studying the homogeneity and expression of 5 S genes in several species from family Asteraceae, known to contain linked 35 S-5 S units, together with the determination of their methylation status by bisulfite sequencing [38]. Our purpose was to know which proportion (if any) of linked genes is expressed and functional.

For an adequate ribosome synthesis there has to be a stoichiometric ratio of rRNA molecules, but there are certain differences regarding the regulation of the transcription of the individual genes taking part of the ribosomal units. The 35S transcript is produced by RNA polymerase I and it is processed in the nucleolus to become 18S, 5.8S and 26S rRNA molecules. The 5S transcription is performed by RNA polymerase III and requires an internal promoter within the genic region (internal control region, ICR). This promoter consists on the A-box, an internal element (IE) typically two nucleotides, and a C-box (Figure 16). The internal promoter sequence is highly conserved among plants and animals [39]. The regulation of the synthesis of rRNAs is also epigenetically controlled: non-transcribed rRNA genes can be silenced by epigenetic processes involving chromatin modifications including DNA methylation [40, 41].

Using an approach that included RT-PCR, cloning, bisulfite sequencing and FISH to determine the expression, methylation status and chromatin condensation levels we showed that in several representative plant species the linked 5S genes are expressed and dominantly contribute to the cellular 5S rRNA pools [38].

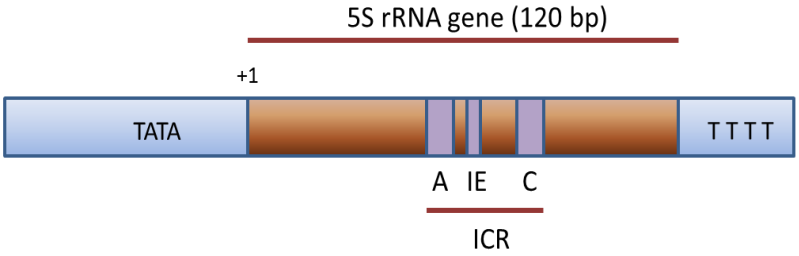

Figure 16. Structure of the 5S rRNA gene, showing the internal control region (ICR), which is subdivided into the A box, the intermediate element (IE), and the C box.

Another question that arose was the presence of minor separate rRNA genes (S-type arrangement) in species which are mostly L-type. Previously, Southern blot had shown almost totally homogenized linked 5S-35S units in several L-type genera of Asteraceae. However, several separate 5S tandems were also recovered in certain species (although in most L-type genomes there were no separate tandems). The interpretation was that these rare separate tandems likely came from loci that did not hybridise with 26S probe on Southern blots, although quantitative estimates by RT-PCR indicate that these are less than 10% in L-type genomes. Indeed, four types of rDNA arrangement were found: 1) completely homogenized L-type arrangement (with no 5S tandems outsides the linked units), as found in genera *Helichrysum*, *Matricaria* or *Tagetes*; 2) L_s-type, with most but not all rDNA units homogenized to the linked arrangement and some separate 5S tandems, as in genera *Artemisia* and *Gnaphalium*; 3) S_L-type, with a dominance of independent 5S tandems but few L-type rDNA units, like in genus *Elachanthemum*; 4) S-type arrangement with 5S independent tandems only, as found in most angiosperms.

4.1. Which 5S is transcribed?

In L_s genomes both L-type and S-type loci encode potentially transcribed genes, although the latter are much less abundant. It is known that only a fraction, sometimes quite reduced, of rRNA genes is transcribed in the cell while the rest is epigenetically inactivated, so it is interesting to set the origin of 5S rRNA transcripts. We suggest that the contribution to the total 5S rRNA pool from the scarce separate 5S tandem arrays is low, if any. Indeed, the linked 5S are being expressed, because:

1) 5S transcripts including IGS1 sequences (beyond the first termination signals) were identified, implying that transcribed 5S sequences actually stem from linked 5S genes,
2) no read-through transcripts were detected from tandemly arranged 5S-5S in L_s-type species,
3) the RNAs derived from linked genes adopted a secondary structure typical of a functional molecule according to the RNA folding simulations performed (Figure 17),
4) linked arrays contained undermethylated and decondensed chromatin fractions, likely corresponding to active genes.

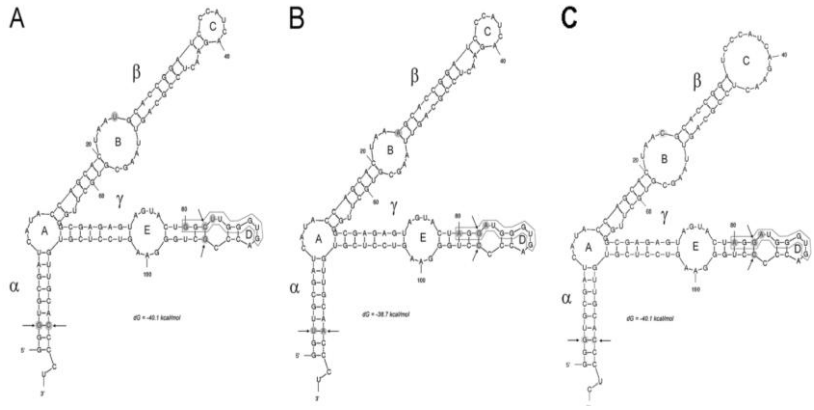

Figure 17. Secondary structure models for 5S rRNA from *Artemisia absinthium* (A), *Tagetes patula* (B) and *Helichrysum bracteatum* (C). Variable nucleotides highlighted by gray shading. Arrows indicate compensatory substitutions. The C-box element is boxed. Structural domains and loops are respectively in Greek and Latin letters. Extracted from [38].

4.2. Divergences in plant C-box consensus sequence

As pointed out previously, 5S rDNA contains regulatory elements (ICR) within the genic region, contrasting to most genes. The sequence and position of these elements (Figure 18) is highly conserved across eukaryotes [42]. However, it seems that certain Anthemideae species have evolved a variant of the C-box (C*) that is different from the angiosperm consensus. Interestingly, mutations in this region are related with reduced transcription rates in *Arabidopsis* [43]. Nevertheless, these changes in C-box do not affect the secondary structure of 5S rDNA which implies that the interactions between rRNA and transcription factor III are not damaged. The C*-box is found in some, but not all, L-type species (and the other way round) so it seems that there is no relationship between the occurrence of the promoter variants and the genomic arrangement of rRNA genes.

4.3. 5S rDNA methylation profiles of L-type and S-type species

The CG, CHG and CHH methylation profiles typically found in plant repetitive DNA [44] are also found in L-type and S-type species. Therefore it

Figure 18. Aligned C-box sequences from different plant and animal species. Shaded letters indicate conserved nucleotides in plants. Cassandra is a non-autonomous retroelement homologous to 5S gene (in this case, from *Rosa rugosa*). Variable nucleotides within the unique C*box are underlined and in bold. Extracted from [38].

seems that the tandem arrangement of 5S is not essential for its methylation, as 5S methylation patterns of L-type species are similar to those of S-type species. We found a significant rate of methylation in non-CG motifs in *Helianthus annuus* (with S-type arrangement) which shows a pericentromeric 5S locus, while much less methylated 5S are detected in *Tagetes patula* or *Artemisia absinthium*, presenting a terminal position of these genes. Previously researchers had found in *Arabidopsis* [45] that the 5S loci close to the centromeres were more methylated than other distally located, so it is likely that the specific location in chromosomes is related with its methylation rate.

4.4. Chromatin condensation and expression of rDNA

RNA polymerase I transcribes the 35S genes and occurs in the nucleolus and RNA polymerase III transcribes the 5S genes and is found in the nucleoplasm. Therefore transcription of both 5S and 35S cannot take place in the same unit at the same time, which may explain why we did not find products of bidirectional 26S-5S transcription. It is possible that the regulation of the transcription takes places at individual chromosome sites. While one chromosome homolog is acting as a NOR, transcribing 35S rRNA genes, the other could be transcribing the 5S rRNA genes. Figure 19 shows a late anaphase/telophase of *Helichrysum bracteatum* in which the nucleoli were assembled on one highly decondensed homolog, while the other was highly condensed and probably not involved in nucleolus assembly. The 35S-5S arrays closely associate with the nucleolus suggest that 5S transcription may occur in close proximity to the nucleolus, possibly at its periphery.

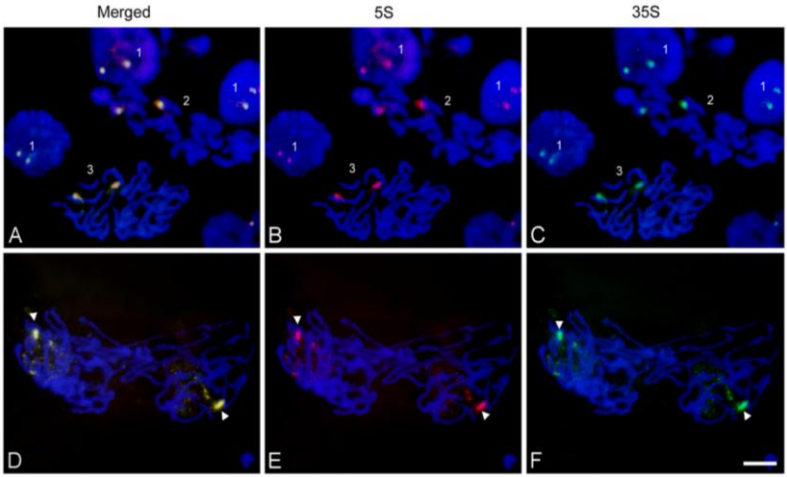

Figure 19. FISH of 5S (red) and 35S (green) rDNA probes to *Helichrysum bracteatum* nuclei. Pictures in the first row (A, B and C) show rDNA signals in interphase (field 1), metaphase chromosomes (field 2), and a prophasic cell (field 3). The second row (D, E and F) shows late anaphase/early telophase. One rDNA homolog is highly condensed (arrowheads) whereas the other is spread throughout the nucleus and decondensed. Pictures extracted from [38].

5. The Plant rDNA database

The idea of creating a database that stores information on number, position and structure of the 5S and 18S-5.8S-26S ribosomal DNA (rDNA) loci

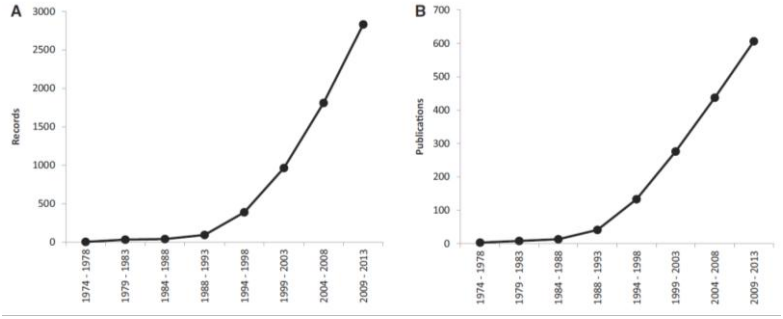

Figure 20. Mean number of records (A) and publications (B) reported per year over eight successive 5-year periods, between 1974 and 2013. Extracted from [49].

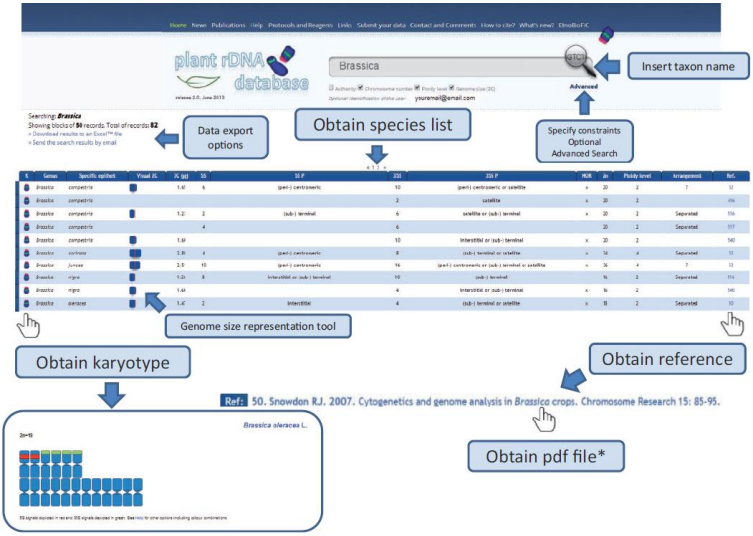

Figure 21. Flowchart showing the possible database outputs and options for a given search. The karyotype appears as a pop-up window, in which the 5S rDNAs are depicted in red and the 35S in green. Extracted from [49].

stemmed from our previous observations of particular rDNA linkages in certain plant groups (see parts 2 and 3 of this article). With the purpose of finding more linked rDNAs in plants, we started a review of any publication containing plant rDNA-FISH data. However, we realised that number, position and structure of rDNAs are useful cytogenetic traits that can help at species characterisation [46, 47], and we detected a growing interest within the scientific community to produce such data (Figure 20). Therefore, in order to store this information, and to make it accessible to a broader scientific community, an online resource, the Plant rDNA database (www.plantrdnadatabase.com) was developed in 2011 [48] and updated in 2013 [49]. The data comes from fluorescent *in situ* hybridisation experiments (FISH) and additional information is also provided, such as ploidy level, mutual arrangement of rRNA genes, genome size and life cycle. The webpage is intuitive and user-friendly, including different search options (Figure 21).

6. Further work

The next steps in our quest on ribosomal DNA will take a deeper look on the evolutionary role of the 5S rRNA gene across the tree of life. In particular, we will explore the evolutionary models of concerted evolution and birth-and-death and we will see how these adjust to globally explain the evolution of this ubiquitous and unquiet sequence. We are also interested in what is the participation of transposable elements in its mobility. Besides, we will continue working on the Plant rDNA database: a new update is planned to be released in 2016 and a global analysis of the data contained is also being prepared.

Acknowledgements

We acknowledge Professor Joan Vallès for his careful reading and corrections performed on the manuscript, as well as all members of our research team, www.etnobiofic.com which have participated at different stages of the project. This work was supported by Ministry of Economy and Competitiveness from the Government of Spain (CGL2010-22234-C02-01,02/BOS and CGL2013-49097-C2-2-P), the Generalitat de Catalunya, Government of Catalonia ("Ajuts a grups de recerca consolidats", 2009SGR0439 and 2014SGR514) and the IRBio (Institut de Recerca de la Biodiversitat, Universitat de Barcelona). SG benefited from a “Ramón y

Cajal" researcher contract of the Ministry of Economy and Competitiveness from the Government of Spain.

References

1. Gilbert, W. 1986, *Nature*, 319, 6018.
2. Lafontaine, D. L., Tollervey, D. 2001, *Nat. Rev. Mol. Cell Biol.*, 2, 514.
3. Hemleben, V., Zentgraf, U., Torres-Ruiz, R. A., Schmidt, T. 2000. In: *Progress in Botany.* Springer, Berlin. pp. 91-117.
4. Fulneček, J., Matyásek, R., Kovařík, A. 2006, *Genome*, 49, 840.
5. Drouin, G., de Sá, M. M. 1995, *Mol. Biol. Evol.*, 3, 481.
6. Vahidi, H., Purac, A., LeBlanc, J.M., Honda, B. M. 1991, *Gene*, 108, 281.
7. Bergeron, J., Drouin, G. 2008, *Curr. Genet.*, 54, 123.
8. Drouin, G., Hofman, J. D., Doolittle, W.F. 1987, *J. Mol. Biol.*, 196, 943.
9. Maizels, N. 1976, *Cell*, 9, 431.
10. Capesius, I. 1997, *Plant Mol. Biol.*, 33, 559.
11. Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, R., Yamato, K. T., Ohmido, N., Fukui, K., Fukuzawa, H., Ohyama, K. 1999, *Plant Mol. Biol.*, 41, 679.
12. Vallès, J., Garcia, S., Hidalgo, O., Martín, J., Pellicer, J., Sanz, M., Garnatje, T. 2011, *Adv. Bot. Res.*, 60, 349.
13. Tkach, N. V., Hoffmann, M. H., Röser, M., Korobkov, A. A.,Von Hagen, K. B. 2008, *Evolution*, 62, 184.
14. Ling, Y. R., Humphries, C. J., Shultz, L. 2006, *Flora of China*, 20, 845.
15. Torrell, M., Cerbah, M., Siljak-Yakovlev, S., Vallès, J. 2001, *Bocconea*, 13, 623.
16. Garcia, S., Lim, K. Y., Chester, M., Garnatje, T., Pellicer, J., Vallès, J., Leitch, A. R., Kovařík, A. 2009, *Chromosoma*, 118, 85.
17. Funk, V. A., Bayer, R. J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., Schilling, E., Panero, J. L., Baldwin, B. G., Garcia-Jacas, N., Susanna, A., Jansen, R. K. 2005, *Biol. Skr.*, 55, 343.
18. Pires, J. C., Lim, K. Y., Kovařík, A., Matyásek, R., Boyd, A., Leitch, A.R., Leitch, I. J., Bennett, M. D., Soltis, P.S., Soltis, D. E. 2004, *Am. J. Bot.*, 91, 1022.
19. Dydak, M., Kolano, B., Nowak, T., Siwinska, D., Maluszynska, J. 2009, *Hereditas*, 146, 152.
20. Cuellar, T., Belhassen, E., Fernandez Calvin, B., Orellana, J., Bella, J. L. 1996, *Heredity*, 76, 586.
21. Ruas, C. D., Vanzela, A. LL., Santos, M. O., Fregonezi, J. N., Ruas, P. M., Matzenbacher, N. I., de Aguiar-Perecin, M. L. R. 2005, *Genet. Mol. Biol.*, 28, 129.
22. Weiss-Schneeweiss, H., Stuessy, T. F., Siljak-Yakovlev, S., Baeza, C. M., Parker, J. 2003, *Pl. Syst. Evol.*, 241, 171.
23. Garcia, S., Panero, J. L., Siroky, J., Kovařík, A. 2010, *BMC Plant Biol.*, 10, 176.
24. Panero, J. L., Funk, V.A. 2008, *Mol. Phyl. Evol.*, 47, 757.
25. Drouin, G., de Sá, M. M. 1995, *Mol. Biol. Evol.*, 12, 481.

26. Garcia, S., Kovařík, A. 2013, *Heredity*, 111, 23.
27. Murray, B. G., Friesen, N., Heslop-Harrison, J. P. 2002, *Ann. Bot. London*, 89, 483.
28. Hizume, M., Shibata, F., Matsusaki, Y., Kondo, T. 1999, *Chromosome Science*, 3, 142.
29. Galián, J. A., Rosato, M., Rosselló, J. A. 2012, *Heredity*, 108, 640.
30. Wicke, S., Costa, A., Muñoz, J., Quandt, D. 2011, *Mol. Phyl. Evol.*, 61, 321.
31. Puizina, J., Sviben, T., Krajačić-Sokol, I., Zoldoš-Pećnik, V., Siljak-Yakovlev, S., Papeš, D., Besendorfer, V. 2008, *Pl. Biol.*, 10, 256.
32. Siljak-Yakovlev, S., Cerbah, M., Coulaud, J., Stoian, V., Brown, S. C., Zoldos, V., Papes, D. 2002, *Theor. Appl. Gen.*, 104, 505.
33. Cai, Q., Zhang, D., Liu, Z. L., Wang, X. R. 2006, *Ann. Bot.-London*, 97, 715.
34. Mathews, S. 2009. *Am. J. Bot.*, 96, 228.
35. Chaw, S. M., Aharkikh, A., Sung, H. M., Lau, T. C., Li, W. H. 1997, *Mol. Biol. Evol.*, 14, 56.
36. Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., Schulman, A. H. 2008, *Proc. Natl. Acad. Sci. USA.* 105, 5833.
37. Zhou, Z., Zheng, S. 2003, *Nature*, 423, 821.
38. Garcia, S., Khaitová, L. C., Kovařík, A. 2012, *BMC Plant Biol.*, 12, 95.
39. Orioli, A., Pascali, C., Pagano, A., Teichmann, M., Dieci, G. 2011, *Gene*, 493, 185.
40. Vaillant, I., Tutois, S., Cuvillier, C., Schubert, I., Tourmente, S. 2007, *Pl. Cell Physiol.*, 48, 745.
41. Blevins, T., Pontes, O., Pikaard, C. S., Meins, F. 2009, *PLoS One*, 4, e5932.
42. Specht, T., Szymanski, M., Barciszewska, M. Z., Barciszewski, J., Erdmann, V. A. 1997, *Nucl. Ac. Res.*, 25, 96.
43. Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., Tourmente, S. 2003, *Plant J.*, 35, 251.
44. Vanyushin, B. F., Ashapkin, V. V. 2011, *Biophys. Acta*, 1809, 360.
45. Vaillant, I., Tutois, S., Jasencakova, Z., Douet, J., Schubert, I., Tourmente, S. 2008, *Plant J.*, 54, 299.
46. Baeza, C. M., Schrader, O. 2005, *Gayana. Botánica*, 62, 47.
47. Xu, Y. H., Yang, F., Cheng, Y. L., Ma, L., Wang, J. B., Li, L. J. 2007, *Hereditas*, 29, 614.
48. Garcia, S., Garnatje, T., Kovařík, A. 2012, *Chromosoma*, 121, 389.
49. Garcia, S., Gálvez, F., Gras, A., Kovařík, A., Garnatje, T. 2014, *Database*, 2014, bau063.