Theor. Appl. Genet. 63, 361–365 (1982)

THEORETICAL AND APPLIED GENETICS

© Springer-Verlag 1982

Relative Rates of Divergence of Spacer and Gene Sequences Within the rDNA Region of Species in the *Triticeae*: Implications for the Maintenance of Homogeneity of a Repeated Gene Family

R. Appels and J. Dvořák
CSIRO, Division of Plant Industry, Canberra City, ACT (Australia)

Summary. The relative rates of divergence of 11 regions of the wheat rDNA cloned in pTA250 were estimated by measuring sequence change in 6 *Triticum* species. The Tm analysis of ^{32}P probes synthesized from the pTA250 regions and hybridized to DNA from the *Triticum* species provided an estimate of sequence change relative to *T. aestivum.* The results revealed a region of 1.2 kb preceding the 18S rRNA gene which was more conserved than the rest of the spacer. In addition the transcribed spacer between the 18S and 26S rRNA genes was shown to be poorly-conserved; the genes, as expected, were highly conserved. A model which proposes RNA as a co-factor in gene conversion is suggested to account for the observations.

Key words: *Triticeae* – Spacer – Divergence – rDNA – Homogeneity – Gene conversion

Introduction

The preceding two papers have provided data to establish a detailed map of the rDNA cloned in pTA250, and the variation of the "130" bp and "750" bp spacer sequences among species of *Triticeae* as well as populations of a given species. The comparison of the apparent rates of divergence of the "130" bp and "750" bp spacer sequences among the large numbers of species examined by Dvořák and Appels (1982) indicated that the "130" bp region accumulated changes more rapidly than the "750" bp region. In this paper we have sought to analyse this in more detail by using probes for 11 different regions of the rDNA unit. To estimate the relative level of conservation of the different sequences, six *Triticum* species covering a wide range of sequence variation on the basis of the data in Dvořák and Appels (1982) were compared. The Tm analysis of DNA hybrids formed between the radioactive sequence probes and genomic DNA from the various species was used to estimate the degree of sequence change in particular parts of the rDNA.

Materials and Methods

All procedures relating to the experiments reported are presented in the preceding two papers. The regions of the rDNA in pTA250 which were used as probes are summarized in Fig. 1. The regions are designated by the clones from which they originated i.e. from left to right, 250.16 Taq is the largest Taq fragment (340 bp) recovered from the 650 bp Hha fragment cloned in pTA250.16; 130.8 is one of the "130" bp spacer DNA fragments analysed in detail in Appels and Dvořák (1982); 250.15 Hae III.2 is a 250 bp Hae III fragment recovered from the 750 bp Hha fragment cloned in pTA250.15; 250.15 Hae III.1 is a 300 bp Hae III fragment recovered from the 750 bp Hha fragment cloned in pTA250.15; 250.17 Taq.2 is a 150 bp Taq fragment from the 650 bp Hinf fragment cloned in pTA250.16; 250.17 Taq.1 is a 450 bp Taq fragment from the 650 bp Hinf fragment cloned in pTA250.17; 250.10 is the 900 bp fragment from pTA250.10; 250.2 Dde.1 is the largest Dde fragment (1,000 bp) from the 3.6 kb of rDNA cloned in pTA250.2; 250.11 is the 500 bp fragment from pTA250.11; 250.5 is the 600 bp fragment from pTA250.5; 250.3 is the 900 bp fragment from pTA250.3. None of the sequences (except 250.2 Dde.1) cross-hybridized with each other to any major extent; some of the spacer sequences did show detectable cross-hybridization but this was minor compared to the homologous reaction. The sequence 250.2 Dde.1 overlaps 250.10 and 250.11.

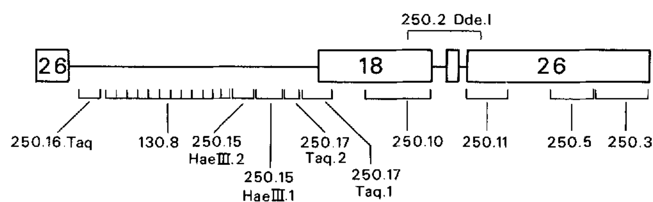

Fig. 1. Summary of location of segments from pTA250 used to assay divergence of different parts of the rDNA region

0040-5752/82/0063/0361/$ 01.00

Results

In Figures 2a, b the ΔTm's of respective hybrid molecules (relative to the Chinese Spring standard) are plotted on the Y axis vs an arbitrary "evolutionary distance" scale (X axis). The relative distribution of the *Triticum* species on the X axis was based on the Tm analyses of the "130" bp spacer sequences in Dvořák and Appels (1982). The species were arranged on this axis in such a manner so that the data could be described by linear regressions. Consequently the slope of the line described by the 130.8 bp sequence in Fig. 2a (for example) has no absolute meaning. The levels of divergence of the sequence are estimated relative to the 130.8 bp sequence since this appeared to show the maximum degree of divergence.

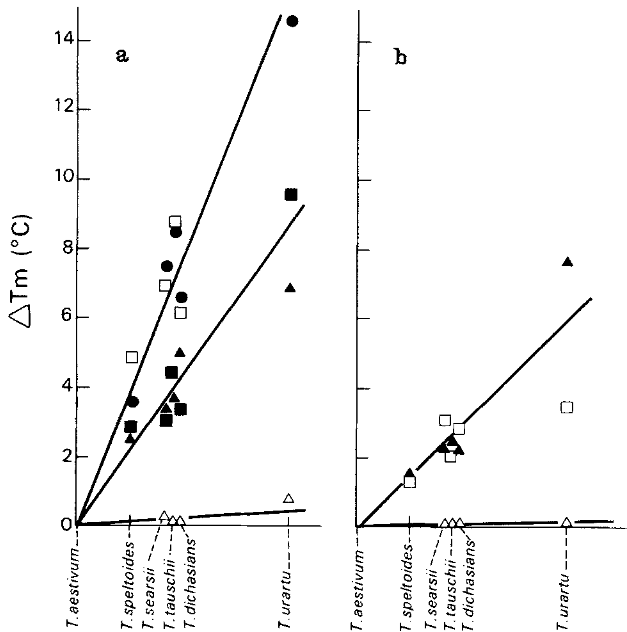

Fig. 2a and b. The relative levels of divergence of different parts of the rDNA region. A linear relationship between "evolutionary distance" and accumulation of sequence change (measured as a ΔTm relative to Chinese Spring) was assumed, to simplify the comparison between the different sequences. For each sequence the ΔTm of the 5 species shown were determined – the ΔTm of *T. aestivum* is 0 since this DNA (from the cultivar Chinese Spring) was used as a standard of comparison. **a** Probes are 130.8 (● – ●), 250.16 Taq (□ – □), giving a slope of 2.6 (SE 0.09) and 250.15 Hae III.1 (■ – ■), 250.17 Taq.2 (▲ – ▲) giving a slope of 1.4 (SE 0.08). The gene probe, 250.10 (△ – △) generates a slope not significantly different from 0. **b** Probes are 250.15 Hae III.2 (▲ – ▲), 250.17 Taq.1 (□ – □) giving a slope of 0.94 (SE 0.08). The gene probe 250.5 (△ – △) has a slope of 0. The sequence 250.2 Dde.1 is not plotted but is presented in Fig. 3. The sequences 250.3 and 250.11 are not shown in the graphs since the slopes of the regressions of these gene sequences were not significantly different from 0. The points for the 250.5 and 250.10 sequences hybridizing to *T. speltoides* are not shown for reasons detailed in the text (see also Fig. 5). The sequence 250.16 Taq did not hybridize with *T. urartu*

Two aspects of the data presented in Fig. 2a, b deserve comment. First the relative ranking of the species with each sequence, particularly the spacer sequences, is in general constant. The ranking of *T. dichasians* does change relative to *T. searsii* and *T. tauschii* (with sequences 250.15 Hae III.1, 250.17 Taq. 1 and 250.17 Taq. 2) and with one sequence (250.17 Taq. 1) *T. tauschii* changes its ranking relative to *T. searsii;* these discrepancies are not unexpected since the respective species are all related to Chinese Spring by similar degrees (using the DNA sequence probes). Secondly, the spacer sequences fall into two broad groups, one showing the minimum level of conservation and the other a significantly higher level of conservation. The gene sequences, as expected, showed high levels of conservation.

The sequence probe 250.2 Dde. 1 did not give clearly interpretable results in the Tm analyses because it spans a conserved gene region and a poorly conserved spacer region. Inspection of the curves in Fig. 3 clearly indicates the presence of a poorly conserved sequence(s) as well as a highly conserved sequence but the complexity of the curves prevents a simple estimation of the ΔTm relative to the Chinese Spring standard. It is interesting to note that the 250.2 Dde. 1 sequence spans the transcribed spacer region (Fig. 1); we can deduce that the poorly conserved sequence(s) derives from this region because the conserved gene sequences 250.10 and 250.11 overlap 250.2 Dde. 1. Although the conserved 5.8S rRNA gene is in this region, recent studies have shown that in mouse the region between the 5.8S rDNA and 28S rDNA is divergent at least on a broad evolutionary scale (Michot et al. 1982).

The data in Fig. 2 can be summarized in relation to the map of the rDNA unit in wheat (Fig. 4). The data has been normalized to assign the rDNA gene sequences a value of 0 with respect to their degree of divergence and the "130" bp sequences a value of 1.0. With regard to this normalization it should be noted that a clear exception to the general consistency in ranking of the species used to assess levels of conservation of a given sequence was found with the gene sequence probes (250.10, 250.11 and 250.3). The Tm of the hybrid formed between these sequences and *T. speltoides* DNA indicated a significant ΔTm relative to Chinese Spring DNA (Fig. 5) while the other species, *T. dichasians, T. searsii, T. tauschii* and *T. urartu* were virtually indistinguishable from Chinese Spring DNA. In assigning the gene sequences to have a level of divergence of 0 the behaviour of *T. speltoides* DNA was not included.

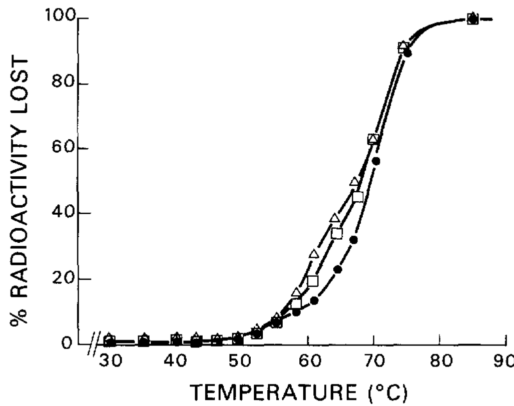

Fig. 3

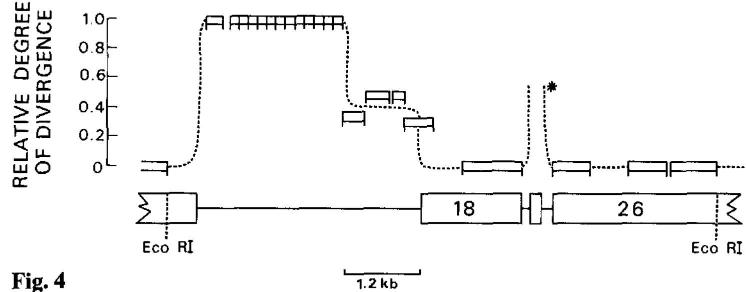

Fig. 4

Fig. 3. Tm curve of the 250.2 Dde.1 probe hybridized to Chinese Spring DNA (● – ●), *T. speltoides* (□ – □) and *T. searsii* (△ – △). The presence of a poorly conserved sequence is clearly indicated by the mismatching occuring in the lower parts of the curves

Fig. 4. Summary of the relative degree of divergence of the sequences assayed. The data from Fig. 2 was normalized with the 130.8 bp sequence being assigned a value of 1.0 and the rRNA gene sequences a value of 0. The 130.8 sequence showed the same degree of divergence as the population of "130" bp sequences (the "130" bp sequence) utilized by Dvořák and Appels (1982) and thus the entire "130" bp region is assigned a relative degree of divergence of 1.0. The * denotes that the exact position, and degree of divergence, of the sequence(s) in this region was not determined (see text and Fig. 3)

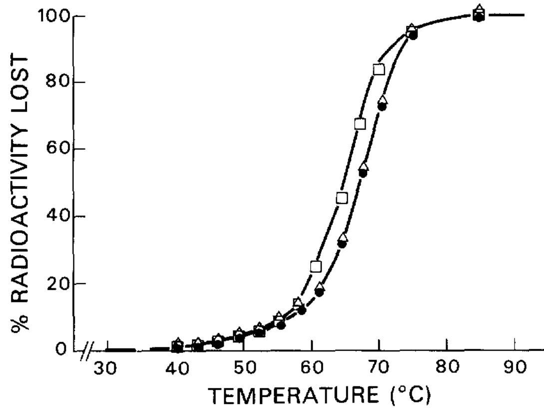

Fig. 5. Tm curve of the 250.10 probe hybridized to Chinese Spring DNA (● – ●), *T. speltoides* (□ – □) and *T. searsii* (△ – △). The species *T. tauschii T. dichasians* and *T. urartu* were essentially indistinguishable from *T. searsii* with this sequence

Discussion

The data presented illustrate the advantages and limitations of the Tm analysis (utilizing cloned probes) in measuring changes in nucleotide sequences within a group of phylogenetically related species. If the probe is too large and hybridizes to a DNA segment spanning regions differing in their degree of conservation, misleading estimates of levels of change in nucleotide sequence occur. Data in Fig. 2 indicate that in these situations the bias is towards a more conserved estimate of sequence difference. Small probes (100–) bp) however largely overcome this limitation an. allow the Tm analysis to provide rapid and reliable estimates of nucleotide sequence change in either repeated and possibly unique DNA sequences. The obvious advantage of this approach is that large numbers of individuals and species can be assayed to develop in detail the rules governing the evolution of DNA at the nucleotide sequence level. The exceptional behaviour of *T. speltoides* rRNA gene sequences, for example, in terms of their relative degree of conservation indicate that this species may be of particular interest for further study at the DNA sequence level. The changes detected in the *T. speltoides* 18S rRNA or 26S rRNA genes involve at most 8–10 base pairs but are nevertheless significant departures from the highly conserved nature of the rRNA genes in the other *Triticum* species assayed.

The asymmetrical nature of the distribution of the relatively more conserved sequences within the rDNA spacer region requires further comment. The spacer region with a relative degree of divergence of 1.0 begins within 100 bp of the end of the 26S rRNA gene and stops approximately 1.2 kb from the start of the 18S rRNA gene. The 1.2 kb region preceding the of the 18S rRNA gene was assigned a relative degree of divergence of 0.3–0.5. This suggests that this region must contain promoters etc., for the initiation of transcription or signals for the processing of rRNA transcripts, since these would be expected to be near the 18S rRNA gene (Nomura 1976; Perry 1976). Although this is undoubtedly the case to some degree, the 1.2 kb of

DNA is an unusually large region to be devoted to such a function(s) compared to other genes (for example see Breathnach and Chambon 1981). Furthermore it is difficult to envisage selection effectively operating on such a region to maintain a relatively high degree of conservation of the promoter-type (or processing signal) sequences. Loss or modification of a single promoter sequence, for example, would reduce the production of 18S and 26S rRNA from the genes associated with the promoter but would not produce defective rRNA products. Since the cereals, in general, have shown large quantitative differences in the amount of rDNA in the genome (Flavell and Smith 1974) a significant reduction in functional rDNA could probably be tolerated. It is thus difficult to consider selection alone operating to conserve promoter sequences (or other functional sequences), in a tandem array of 2000 or more rDNA units.

This problem has been considered by other authors and two explanations have been proposed for maintaining homogeneity within a tandem array of repeated gene sequences, in combination with some selection pressure (for recent discussions see Birky and Skavaril 1976; Coen et al. 1982). One of these involves unequal exchange among the sequence arrays. The other model, democratic gene conversion, postulates that a given unit of a tandem array of repeated sequence units has an equal probability of converting any other unit in an allelic or non-allelic array, by strand transfer and mismatch repair (for review see Radding 1978). Either model has to postulate selection to account for the differential levels of conservation of the spacer and genes in the rDNA region observed in this study as well as in other studies (Federoff 1979; Long and Dawid 1980). Although this is possible, an interesting alternative which may have the same consequences is the involvement of RNA as a co-factor in democratic gene conversion. The detailed mechanism is based on the in vitro observation that RNA can invade a DNA double helix to form an R-loop (Wellauer and Dawid 1977; White and Hogness 1977). This would allow the displaced DNA strand to invade the homologous DNA double helix and form a heteroduplex which is then subjected to excision repair processes leading to gene conversion (Radding 1978). The stochastic nature of this process results in the conservation of the sequence most often involved in the heteroduplex formation. The participation of RNA as a co-factor would provide a bias toward those DNA sequences which produce RNA sequences in highest abundance. In the rDNA system this would correspond to the rRNA gene sequences while transcribed spacer sequences, which are present as short-lived RNA molecules, would be in effectively lower concentrations. The model thus predicts that gene sequences would show the highest degree of conservation, transcribed spacer regions a lower degree and the non-transcribed spacer the lowest degree of conservation. Although speculative, the model of RNA participating as a co-factor in democratic gene conversion, perhaps in combination with other forces such as selection, would produce the type of pattern of sequence divergence observed in Figure 4.

We would predict that transcription of the rDNA precursor RNA initiates approximately 1.2 kb from the start of the 18S rRNA gene and that this 1.2 kb region is present in short-lived RNA molecules. Although processing of rRNA has not been studied in wheat the published data on rRNA processing in other plants suggest that the precursor is 5.98 kb–6.42 kb (recalculated using 1.55 kb for the 18S rRNA gene as a standard) in mung bean (Grierson and Loening 1972) and 7.53 kb in sycamore (Cox and Turnock 1973). These measurements would potentially leave 0.7 kb–1.13 kb and 2.24 kb of DNA preceding the 18S rRNA gene (respectively) which is transcribed into a short-lived RNA molecule. The 1.2 kb predicted above falls within this range. The model readily accounts for the poorly conserved region between the 18S and 26S rRNA genes since this transcribed spacer (except for the 5.8S rRNA) would also be in the short-lived RNA molecule class.

The concept of RNA as a co-factor in gene conversion may well have wide application and we expect that the further genetic and molecular analysis of the rDNA system in the cereals will contribute to establishing its validity.

Acknowledgements

The authors are grateful to Lyndall Moran for her expert assistance in carrying out the experiments reported. The authors thank Dr. J. Wood (Division of Mathematics and Statistics) for carrying out the linear regression analyses.

Literature

Appels, R.; Dvořák, J. (1982): The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor. Appl. Genet. **63**, 337–348

Birky, C.W.; Skavaril, R.V. (1976): Maintenance of genetic homogeneity in systems with multiple genomes. Genet. Res. **27**, 249–265

Breathnach, R.; Chambon, P. (1981): Organization and expression of eucaryotic split genes coding for protein. Ann. Rev. Biochem. **50**, 349–383

Coen, E.S.; Thoday, J.M.; Dover, G. (1982): Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. Nature **295**, 564–568

Cox, B.J.; Turnock, G. (1973): Synthesis and processing of ribosomal RNA in cultured plant cells. Eur. J. Biochem. **37**, 367–376

Dvořák, J.; Appels, R. (1982): Chromosome and nucleotide sequence differentiation in genomes of polyploid *Triticum* species. Theor. Appl. Genet., **63**, 349–360

Federoff, N.V. (1979): On spacers. Cell **16**, 697–710

Flavell, R.B.; Smith, D.B. (1974): Variation in nucleolar organizer rRNA gene multiplicity in wheat and rye. Chromosoma **47**, 327–334

Grierson, D.; Loening, V.E. (1972): Distinct transcription products of ribosomal genes in two different tissues. Nature, New Biol. **235**, 80–82

Long, E.O.; Dawid, I.B. (1980): Repeated genes in eukaryotes. Ann. Rev. Biochem. **49**, 727–764

Michot, B.; Bachellerie, J.-P.; Raynal, F.; Renalier, M.-H. (1982): Homology of the 5′-terminal sequence of 28S rRNA of mouse with yeast and *Xenopus*. FEBS Letters **140**, 193–197

Nomura, M. (1976): Organization of bacterial genes for ribosomal components: studies using novel approaches. Cell **9**, 633–644

Perry, R.P. (1976): Processing of RNA. Ann. Rev. Biochem. **45**, 605–629

Radding, C.M. (1978): Genetic recombination. Strand transfer and mismatch repair. Ann. Rev. Biochem. **47**, 847–880

Wellauer, P.K.; Dawid, I.B. (1977): The structural organization of ribosomal DNA in *Drosophila melanogaster*. Cell **10**, 193–212

White, R.L.; Hogness, D.S. (1977): R-loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* DNA. Cell **10**, 181–192

Received June 26, 1982
Communicated by J. S. F. Barker and H. F. Linskens

Dr. R. Appels
Division of Plant Industry CSIRO
Institute of Biological Resources
Black Mountain
P.O. Box. 1600
Canberra City, ACT 2601 (Australia)

Dr. J. Dvořák
Department of Agronomy and Range Science
University of California
Davis, CA 95616 (USA)

Announcements

Conference Report

Genetic Engineering in Eukaryotes

From July 26–August 6, a NATO Advanced Study Institute conference was held at Washington State University at Pullman, USA under the title "Genetic Engineering in Eukaryotes". The meeting, organized by P. Lurquin, brought together experts working in the field of plant, yeast and animal genetic engineering, which allowed a useful exchange of information obtained from various organisms.

In the plant field, many reports were concerned with the "natural genetic engineers" *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes,* which transfer DNA to plant cells, thereby causing the Crown Gall and Hairy Root system, respectively. Although the analysis of the Hairy Root phenomenon was started only recently, it is rapidly catching up with the analysis of Crown Gall, as evidenced by reports from Gordon and Huffman (Seattle) on Rh T-DNA structure and homology with plant DNA; Tepfer (INRA, Versailles), with a remarkable account on regeneration of different species from Rh roots; and by the Schell group (Cologne, Brussels, Ghent) on transcription of Rh T-DNA and homology with Ti plasmids. Crown Gall reports included complementation studies with avirulent bacteria mutated in regions outside the T-DNA (Iyer and Nester), and the description of a chromosomal transposon insertion mutant which is both unable to bind to plant cells and avirulent, thereby confirming that binding is an essential step in transformation (Nester). From the Schell group an octopine-producing but otherwise normal potato plant was reported, which was obtained from tumours induced by a transposon insertion mutant (Wöstemeyer). Sciaky (Brookhaven) reported the occurrence of a natural *Agrobacterium* mutant with a chromosomally derived, transposon-like sequence in the same region of the T-DNA as the artificial mutant used by Wöstemeyer. Apparently, *Agrobacterium* has not only invented genetic engineering of eukaryotes before man did, but it also seems to be capable of doing its own transposon insertion mutagenesis!

Reports on two other possible plant vector systems concerned Cauliflower Mosaic Virus (CaMV) and gemini viruses.

Shillito (FMI, Basel) reported results on transfer and phenotypical expression in plant cells of selectable marker genes incorporated in CaMV-derived vectors. These results, although promising, still need a detailed analysis of the supposedly transferred DNA.

The single stranded gemini viruses discussed by Coutts (Imperial College, London) might become important as vectors for monocots, either as a system in itself, or in combination with the Ti plasmid from *Agrobacterium.* Another possible alternative for introducing DNA in plant cells, and especially monocot cells (which are insensitive to *Agrobacterium*), is liposome-mediated DNA transfer. The Lurquin-Kleinhofs group (Pullman) presented recent progress in their approach to develop such a system, which will use plant-adapted bacterial genes packaged in liposomes to complement plant nitrate reductase mutants. The isolation of the relevant genes and their modification for expression in plants is well under way. Highly efficient transfer of viral RNA packed into