Retrotransposons as sources of *Quercus suber* genetic variabilities

Filomena Nóbrega[1], Rui Vidal[2], Raúl Sardinha[3] e Raúl Bruno de Sousa[4]

[1] Estação Florestal Nacional
[2] Faculdade de Farmácia da Universidade de Lisboa
[3] Instituto PIAGET
[4] Instituto Superior de Agronomia

Introduction

Retrotransposons (RTns) are transposable elements (TEs), also referred as mobile genetic elements (MGEs), of class I-type retroelements (RELs) that have been found, as integrated DNA segments, in all eukaryotic genomes investigated so far, including those of plants where they are present in multi-copy numbers ranging from just a few ones to thousands per genome. They have the ability to actively propagate themselves within genomes via reverse transcription of a RNA intermediate, thereby providing new genetic variations as well as tremendous effects on genome structures and gene functions. Since the RTns may contain or not long terminal repeats (LTRs), they are classified as LTR RTns and non-LTR RTns (also termed LINEs), respectively. On the other hand, LTR RTns may be either of the *Ty1-copia*-type or the *Ty3-gypsy*-type, this depending to the gene order of the enzymes RT and Int (Figure 1).

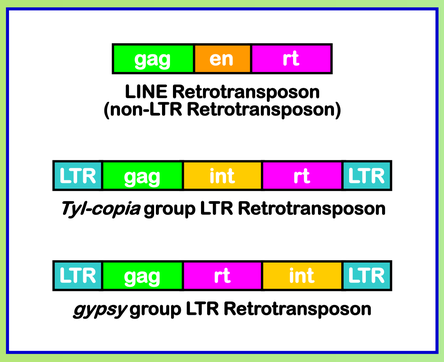

Figure 1 - Structural domains: **gag**, core particule components; **en**, endonuclease; **rt**, reverse transcriptase; **LTR**, long terminal repeat; **int**, integrase.

As observed for LTR RTns, it also seems that LINES are an universal genomic feature because they have been discovered in all eukaryotic genomes (but not yet in cork oak trees) where they have been found either in the flanking regions of normal genes or specifically in basal ones. This is the case of the R4 clade LINEs that only insert into the 26S nuclear rDNA (26S nrDNA) genes of insects and nematodes [1] after recognition of a well conserved, specific sequence (Figure 2), from which the result it is the corresponding gene inactivation.

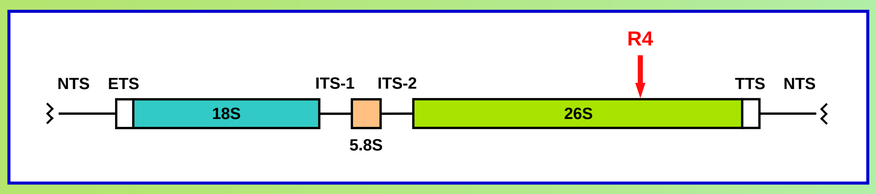

Figure 2 - The organization of higher plant ribosomal DNA: **NTS**, nontranscribed spacer; **ETS**, external transcribed spacer; **18S**, small-subunit rDNA; **ITS-1**, first internal transcribed spacer; **5.8S**, rDNA; **ITS-2**, second internal transcribed spacer; **26S**, large-subunit rDNA; **TTS**, transcription termination site; **R4**, localization of R4 clade LINE insertions.

As a consequence of these observations, we have performed a PCR screening programme for detection of R4-like LINE insertions on *Quercus suber* 26S nrDNA genes in order to elucidate any of the already observed genetic variabilities of the cork trees. The preliminary results of these research, that are here reported are to our knowledge demonstrated for the first time: *i) Quercus suber* genetic polymorphisms due to R4-like LINE insertions into some of the 26S nrDNA genes, and *ii)* the discovery of 2 new LTR RTn elements, at least.

Materials and Methods

Total genomic DNA was extracted from young leaves (15 days old), collected in Spring, from each of 40 cork oaks selected at the Herdade do Monte Fava, situated in Ermidas-Sado, using Qiagen Maxi and Mini DNAeasy kits according to the manufacturer's instructions. PCR amplifications were carried out with a Hybaid Express thermocycler, using Qiagen Taq PCR Master Mix kit (except otherwise indicated), PCR conditions and primer pairs described bellow. All the resulting PCR samples (10 µL for each one) were simultaneously resolved with the 123 bp DNA Ladder and/or High DNA Mass Ladder (Life Technologies, USA) by agarose gel electrophoresis (AGE) in TBE buffer according to standard methods. For purification and cloning of some PCR products, the following Qiagen kits were used: QIAquick–DNA extraction kit, PCR cloning ligation kit and PCR cloning transformation kit. Both strands of some cloned PCR products were sequenced by the sequencing services of either "4base lab" (Reutlinger, Germain) or "MWG Biotech AG" (Ebersberg, Germain).

Results and Discussion

1. Detection of genetic polymorphisms due to R4-like LINE insertions

In order to detect R4-like LINE insertions on *Quercus suber* 26S nrDNA genes, total genomic DNA was isolated from each one of 40 cork oaks above mentioned, and directly amplified by PCR using the following conditions: an initial denaturation step of 3 minutes at 94°C, 30 cycles of 1 minute at 94°C, 1 minute at 50°C and 3 minutes at 72°C, and an elongation step of 10 minutes at 72°C. A specific primer to highly conserved region in the R4 reverse transcriptase domain (primer R4 RTase, 5'-TTCTACATGGACGACGT-3') was used in combination with a primer complementary to the 26S nrDNA gene located about 77 bp at the right of the R4 insertion site (primer 26S-+77bp, 5'-GCCAGATTAGAGTCAAGCTC-3') [1].

The results showed multiple products on each PCR profile, instead of an unique one of approximately 1.8 kb in length as expected [1], and a clear presence of this 1.8 kb PCR product in only about half of the analysed DNA samples as exemplified in Figure 3. On the other hand, visual comparison of all PCR profiles, also provided evidence for other major and minor differences between them, suggesting that these polymorphisms could be due to amplifications of either truncated forms of R4-like insertions on 26S nrDNA genes or other unknown DNA sequences.

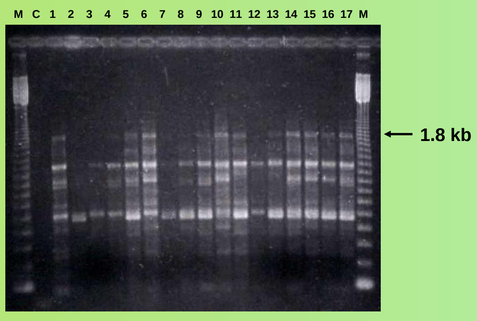

Figure 3 - PCR profiles obtained with primers R4 RTase and 26S-+77bp, simultaneously fraccionated with 123 bp DNA Ladder markers (M) and control PCR mixture without DNA (C) by 0.8% AGE. The 17 amplified DNA samples are indicated by numbers and the position of the expected 1.8 kb PCR product by an arrow.

2. PCR amplification of 26S nrDNA region containing R4 target site

Consequently, in order to confirm or not the presence of R4 target site on 26S nrDNA genes, all DNA samples were amplified by PCR, using the same conditions as above, with a set of specific 26S nrDNA primers [1], one located approximately 100 bp upstream of the R4 LINE insertion site (primer 26SF, 5'-CTGCCCAGTGCTCTGAATGTC-3') and another about 700 bp dowstream (primer 26SR, 5'-AAGAGCCGACATCGAAGGATC-3').

In all studies, the results have demonstrated for all DNAs a major abundant PCR product of approximately 0.7 Kb in length as expected [1]. However, 4 different types in size of a few longer PCR products were again observed in only about half of the samples as exemplified in Figure 4A. These products could be truncated R4-like inserts on some 26S nrDNA genes since one, although very faint, of about 5.5 kb corresponding to a full-length R4 clade LINE insertion containing approximately 4.7 kb [1] was also present in both PCR profiles numbers 1 and 10. These last results were confirmed by repeated PCR amplifications of both relevant cork tree DNAs as shown in Figure 4B.

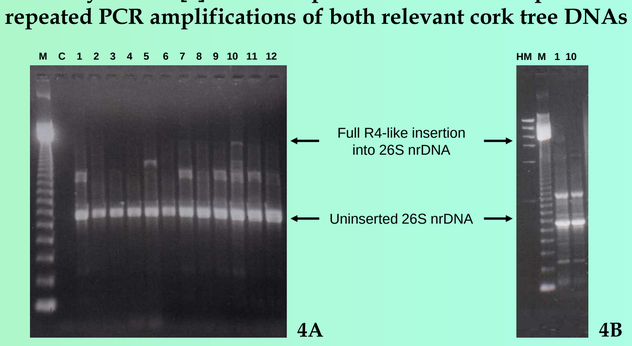

Figure 4 - PCR profiles obtained with primers 26SF and 26SR, simultaneously fraccionated with 123 bp DNA Ladder (M), High DNA Mass Ladder (HM) and control PCR mixture without DNA (C) by 0.8% AGE. **Figures 4A and 4B**: 12 and 2 amplified DNA samples are respectively indicated by numbers, and the positions of the PCR products containing about 0,7 kb (uninserted 26S nrDNA) and 5.5 kb (inserted 26S nrDNA) by arrows.

Consequently, the sequences of uninserted 26S nrDNA region as well as of some remaining PCR products were determined, and deposited in GenBank database under the following accessions numbers: AY094177 (for 26S rDNA partial sequence), AY099465 (for Melmoth retrotransposon-like RNAse H protein gene, partial sequence), AF512588 (for UGY-G1 reverse transcriptase pseudogene, partial sequence), and AF510494 (for putative lipase gene, partial sequence).These results clearly revealed the presence of the expected target site for R4 clade LINE insertions into *Quercus suber* 26S nrDNA genes as well as of, at least, 2 new LTR RTns inserted on *Quercus suber* genome (one, Melmoth, being a *Ty1/copia*–type REL, and another, QSUGY-G1, belonging to the *Ty3/gypsy*-type RTns), thus lending support to the cork tree genetic variabilities as above shown (Figure 3).

3. PCR amplification of full-length 26S nrDNA gene

To elucidate whether *Quercus suber* 26S nrDNA genes contained or not more than one R4 clade target site, all DNA samples were again amplified by PCR but using another set of specific 26S nrDNA primers [2], one corresponding to the 5'-end (forward primer N-nc26S1, 5'-CGACCCCAGGTCAGGCG-3') and the other being complementary to the 3'-end (reverse primer 3331rev, 5'-ATCTCAGTGGATCGTGGCAG-3') of complete plant 26S nrDNA sequences reported so far. The PCR mixtures (final volume of 50 µL)contained the following components: 25 µL 2x Taq PCR Master mix (supplied with Taq DNA polymerase, 400 µM each dNTP and 3 mM $MgCl_2$) plus 1.5 µL 50 mM $MgCl_2$; 1.5 µL each 100 mM dNTP; 1.0 µL each 20 µM primer; 5 µL dimethylsulfoxide (DMSO); 6.5 µL sterile MilliQ water and 2 µL diluted total DNA extract (100 ng). PCR conditions also differed a little from those above descrided as follows: an initial denaturation step of 3 minutes at 94°C; 30 cycles of 1 minute at 94°C, 1 minute at 55°C, and 3.5 minutes at 72°C; and an elongation step for 5 minutes at 72°C.

The results showed a major PCR product of approximately 3.4 kb in length for all analysed DNA samples as expected (Fig. 5, uninserted 26S nrDNA). Thus, the major PCR product amplified from cork tree DNA number B7-5 (Lane 7) was recovered from the gel, purified, cloned and finally sequenced on both strands. The resulting 3,399 bp sequence, that was deposited in GenBank database under accession number AY428812, unmistakably corresponded to uninserted *Quercus suber* 26S nrDNA, and only possessed an unique target site (5'-GTGACGCGCATG-3') as reported [1] for R4 LINE insertion at the expected position located on a well specific, conserved 26S nrDNA region (Figure 2).

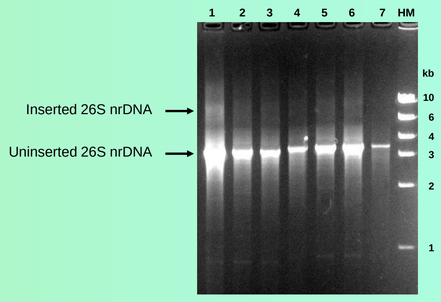

Figure 5 - PCR profiles obtained with primers N-nc26S1 and 3331rev, simultaneously fraccionated with High DNA Mass Ladder (HM) by 0.8% AGE. Amplified DNA samples are indicated by numbers, and the positions of the PCR products containing about 3.4 kb (uninserted 26S nrDNA) and approximately 8.1 kb (inserted 26S nrDNA) by arrows.

However, careful visual inspection of those profiles (Figure 5), also revealed that some cork tree DNA samples yields a larger PCR product of approximately 8.1 kb corresponding to insertion of a full-length 4.7 kb R4 clade LINE [1] into a few *Quercus suber* 26S rnDNA genes as previously detected (Figure 4).

The results obtained from densitometric analysis indicated an average proportion of 0.103/0.897 for those genes. In other words about 4.3% of of the *Quercus suber* 26S nrDNA genes were interrupted by a full-length 4.7 kb R4 clade-type LINE insertion. This percentage corresponds to that ones of approximately 5% and about 4% for full-length 4.7 kb R4 LINE insertion into nematode 26S nrDNA genes [1] and full-length 4.45 kb del2 LINE insertion into *Lilium speciosum* genome [3], respectively. On the other hand, these results also point out, at least, one *Quercus suber* nrDNA locus interrupted by a full-length R4 clade-type LINE insertion, assuming the maximum number of 24 for the cork tree rDNA loci per diploid genome. Therefore, based on these observations and yet that 3 divergent nrDNA families (up to 4.3% for each one) have been identified by 5.8S nrDNA and ITS2 analyses in both *Quercus petreae* and *Quercus robur* species [4], it is feasible to conclude that, at least, one of the *Quercus suber* nrDNA loci has diverged from the others due to this type of LINE insertions.

Conclusions

From these preliminary results concerning *Quercus suber* genetic variabilities obtained with 40 cork trees, the major conclusions are as follows:

1 – R4 clade-type LINE insertions and, at least, two LTR RTns (one of *Ty/copia*-type, and another of *T3/gypsy*-type) are respectively present on 26S nrDNA genes and other unknown genomic regions of some cork trees, this being the cause of the observed genetic diversities;

2 – only about half of the studied *Quercus suber* DNAs harbour truncated forms of the R4 clade-type LINE, this being also responsible for some of the observed genetic variabilities;

3 – only approximately 4.3% of the *Quercus suber* nrDNA loci are interrumpted by a full-length R4 clade-type LINE, this corresponding to a divergent 26S nrDNA locus.

References

[1] Burke, W.D., Muller,F. & Eickbush, T.H.. 1995. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acids Research.*,Vol 23, N° 22: 4628-4634.
[2] Kuzoff *et at.*.1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Mol. Biol. Evol.*, Vol 15, N° 3: 251-263.
[3] Leeton, P.R. & Smyth, D.R.. 1993. An abundant LINE-like element amplified in the genome of Lilium speciosum. *Mol. Gen. Genet.*, Vol 237, N° 1-2: 97-104.
[4] Muir G, Fleming C.C. & Schlotterer C..2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.. Mol. Biol. Evol., Vol 18, N°2: 112-119.

Portugal em Acção — Ministério da Agricultura, Desenvolvimento Rural e Pescas

This research was supported by: Project PAMAF 4003 "Contribution for cork oak breeding" and Project PIDDAC 210 "Characterization of genetic stability and variability in cork trees ans its correlation with cork quality"

E-mail: Filomena.Nobrega@efn.com.pt
Estação Florestal Nacional, Avenida da República, Quinta do Marquês
PT - 2780-159 OEIRAS
Tel.: (+351) 214463791; Fax: (+351) 214463702