

Comparison of Genetic Structure Assessed by Amplified Fragment Length Polymorphism and Retrotransposon-based Sequence-specific Amplification Polymorphism for Portuguese Populations of *Pinus pinea* L.

By I. EVARISTO^{1,*}, S. SANTOS¹, R. TENREIRO² and R. COSTA¹

(Received 31th January 2006)

Abstract

In order to assess genetic diversity within and among populations of *Pinus pinea* L. (stone pine), seven Portuguese populations originating from three Provenance Regions were selected and genotyped using two marker systems. We compared the genetic variation of these populations using retrotransposon-based sequence-specific amplification polymorphism (SSAP) and amplified fragment length polymorphism (AFLP). In total, 105 trees were screened with three primer enzyme combinations (PEC), producing 232 SSAP and 132 AFLP loci. Where SSAP yielded approximately twice the number of polymorphic fragments compared to AFLP. Differentiation was slightly higher for SSAP, than for AFLP ($F_{ST} = 0.105$ for SSAP and 0.074 for AFLP), and both significantly different from zero, $P < 0.01$. The levels of average genetic diversity within-population found with the two types of marker were not significantly different between SSAPs and AFLPs (26.6% and 22.8%, respectively). The populations that displayed the highest and lowest genetic diversity scores were the same for both markers, and only two populations had significantly different H_e estimates. The neighbor-joining tree based on the Nei's genetic distance displayed some geographic pattern. With the AFLP markers the populations grouped according to the provenance regions where they were sampled, resulting in one well supported cluster with the Southern populations, but with SSAP the pattern was not so coherent. In this study SSAP generated more polymorphic fragments and higher estimates of genetic diversity than AFLP did, due, probably, to the higher mutation rate of retrotransposition relative to base mutation. Nevertheless, congruence was found between estimates obtained with both markers, which is very interesting, for, in general, SSAP markers have lower costs compared to AFLPs, and they might be an interesting alternative marker system, when higher resolution is requested.

Key words: sequence-specific amplification polymorphism, amplified fragment length polymorphism, *Pinus pinea*, genetic structure.

Introduction

A novel group of molecular marker techniques, based on insertion polymorphism generated by retrotransposon (RTN) activity, has proven useful in developing mul-

tiplex DNA-based marker systems in plants (KUMAR, 1996; SCHULMAN *et al.*, 2004). Retrotransposons are an evolutionarily ancient class of mobile genetic elements that transpose replicatively within their host genomes via RNA intermediates. They are ubiquitous in plants, in many cases comprising more than half of the DNA nuclear content, and playing a major role in plant genome evolution (reviewed in KUMAR and HIROCHIKA, 2001; SCHULMAN *et al.*, 2004). The polymorphisms generated by the unique biological retrotranspositional process, result from new RTNs insertion sites, most of which are irreversible, without losing the parental copies. The replicative transposable mode can rapidly increase the copy numbers of elements and, thereby, greatly increase plant genome size (KUMAR, 1996; PEARCE *et al.*, 1996). This process provokes an alteration ranging from a few hundred base pairs to a few kilobases at the insertion site. There are two major groups of retrotransposons, those consisting of the long terminal repeat (LTR) sub-classified into Ty1-copia and the Ty3-gypsy groups, and the non-LTR retrotransposons or LINE elements. LTR retrotransposons present as large heterogeneous populations within plants and are highly plastic, showing great variation in copy number and genome location, even between closely related species (KUMAR *et al.*, 1997; PEARCE *et al.*, 1996). Retrotransposon-based markers require sequence information from the mobile element terminal region, but the recently developed rapid LTR RTN sequences isolation technique, together with easy access to transposon sequences from the vast amount of genomic sequence information, have extended these application to plants (PEARCE *et al.*, 1999).

The most popular transposon-based marker method is the sequence-specific amplification polymorphism (SSAP) approach (WAUGH *et al.*, 1997). It is a modified version of amplified fragment length polymorphism (AFLP), in which, one AFLP primer is replaced with an LTR primer. RTN-based markers are especially suitable for studying genetic diversity, within and between species. For example, active members of the RTN family produce new insertions in the genome, leading to polymorphism. The new insertions can then be detected and used to establish the temporal sequence of insertion events, helping to determine phylogenies. These genetic properties have recently been exploited to study biodiversity and phylogeny in different plant species (KUMAR and HIROCHIKA, 2001, and references therein). Some studies have directly compared retrotransposon marker methods with AFLP, and they revealed that retrotransposon markers are more polymorphic than AFLP (e.g.

¹) Estação Florestal Nacional, Laboratório de Biologia Molecular. Av. da República Quinta do Marquês, 2780-159 Oeiras, Portugal.

²) Faculdade de Ciências da Universidade de Lisboa. ICAT-Edifício ICAT, Campus da Faculdade de Ciências, 1749-016 Lisboa, Portugal.

*) Corresponding author: Isabel Evaristo. Phone: 214463700, Fax: 214463702, Email: isabel.evaristo@efn.com.pt.

ELLIS *et al.*, 1998; MAY TAM *et al.*, 2005; SYED *et al.*, 2005; QUEEN *et al.*, 2004; e.g. WAUGH *et al.*, 1997; YU and WISE, 2000). The time and cost efficiency, replicability and resolution of AFLPs are superior or equal to those of other markers, e.g. allozymes, random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), microsatellites (SSR). AFLP markers are, generally, dominant markers and, also, prone to size homoplasmy, even though, they have emerged as a genetic marker with broad application in population genetics, DNA fingerprinting and other fields of research (MUELLER and WOLFENBARGER, 1999; VEKEMANS *et al.*, 2002). One of the important features of those markers are resolution, because of the nearly unlimited number of markers that can be generated, using different primer combinations, and at least some AFLP markers will be located in variable regions and thus reveal genetic differences among organisms (MUELLER and WOLFENBARGER, 1999).

For neutral markers, genetic diversity and differentiation estimates depend mainly on demographic parameters (effective population size, migration rate, and others), the type of markers and their mutation rate, their mode of inheritance, and the population sampling (KREMER and MARIETTE, 2003).

Genetic diversity for *Pinus pinea* has previously been investigated by using different types of molecular markers; isozymes (FALLOUR *et al.*, 1997), RAPD (EVARISTO *et al.*, 2002) and chloroplast microsatellites (cpSSR) (GÓMEZ *et al.*, 2002). The genetic variability reported in the isozyme study showed that only two of 40 isozyme loci were polymorphic, and that the mean heterozygosity amounted to 1%, when all loci screened were considered, a very low value compared with the average, for the *Pinus* genus (HAMRICK *et al.*, 1992). In the GÓMEZ *et al.*, (2002) study, where only Spanish populations were screened, six out of the nine cpSSR were polymorphic and $H_s = 0.570$, but in a range-wide study of the species (G.G. Vendramin, pers. comm.) the mean heterozygosity was much lower $H_s = 0.181$. Nevertheless, RAPD detected high values of phenotypic diversity (Shannon index), but is difficult to compare with other markers estimates, due to the dominance effect (EVARISTO *et al.*, 2002).

In the present study, we aimed to screen 232 SSAP loci (derived from the LTR retrotransposons) and 132 AFLP loci for 105 randomly sampled trees from seven *P. pinea* populations for the following aims: i) to estimate the within and among population genetic variability, ii) to compare the estimates obtained with the SSAP

marker system, derived from the LTR retrotransposons, with AFLP, and iii) to infer population genetic similarity through cluster analysis.

Materials and Methods

Study Species

Pinus pinea is widely distributed around the Mediterranean Basin, and is useful for nut production, shelterbelts, erosion control, recreation, resin and wood production. This species had a long history of cultivation and diffusion by Man, since the Pre-Romanic age (MUTKE *et al.*, 2000), and it is practically impossible to distinguish natural from artificial stands, nowadays. Several authors define Anatolia (Turkey) as the putative area of origin for this species (MIROV, 1967; QUÉZEL, 1980). However, the numerous seeds found in archaeological settlements, dating from the upper Palaeolithic to Neolithic, suggest that stone pine is autochthonous in Spain (STEVENSON, 1985). Other putative native areas are proposed for eastern Mediterranean countries of Italy, Turkey and Lebanon, where *P. pinea* stands form a climax structure surrounded by characteristic native vegetation (MIROV, 1967). The stone pine ranges from Portugal to Syria, including the western coasts of Black Sea, and covers about 380,000 ha, standing at low and medium altitude. The largest stone pine natural forests in the Iberian Peninsula are found under diverse ecological conditions, where it reaches 1.200 m in altitude. In Portugal the stone pine spans all over the country, in pure or mixed stands with *Quercus suber* L. (cork oak) and *P. pinaster*, occupying an area of about 78.000 ha (CARNEIRO, 2002).

Plant material

Young needles collected from the 105 trees, which were randomly sampled from seven provenance stands, were stored at -80°C until DNA extraction. Details about the plant material analysed in this study is listed in *Table 1*.

For economically important species, such as *P. pinea*, according to the European legislation about forest reproductive material, and in order to organise its trade, a list of recommended provenances for use in each region was established in Portugal (CARDOSO and LOBO, 2001).

DNA isolation

DNA was extracted from needle tissue (100 mg fresh weight) using DNeasy Plant Mini Kit (QIAGEN) follow-

Table 1. – Geographical parameters of the studied *P. pinea* populations. See text for details.

Population name	N ^a	Provenance	Longitude (E)	Latitude (N)	Altitude (m)
Ponte Lima	12	I	8°30'	41°41' - 41°42'	290
Viscu	12	III	7°54'	40°41'	443
Pai Sobrado	12	V	8°31'24.8''	38°21'23.9''	47
Monte Novo	10	V	8°39'20''	38°2'42''	89
Mata Valverde	12	V	8°31'54.1''	38°1'37.6''	101
Quinta Sousa	30	V	8°3'15.3'' - 8°3'53''	38°33'33.7'' - 38°35'53''	45-71
Palma	17	V	8°34'54'' - 8°36'08''	38°26'58'' - 38°27'40''	12-84

^a Number of individuals.

Table 2. – SSAP and AFLP sequences used to screen the seven studied populations. N_f = number of scored fragments obtained for each primer combination, N_p = number of polymorphic loci (direct count), and P = proportion of polymorphic loci.

SSAP				AFLP			
PEC Sequences 5'→3'	N_f	N_p	P (%)	PEC Sequences 5'→3'	N_f	N_p	P (%)
GAAGGAGAGGGGCATGTTAGGT/M-CAA	38	28	73.7	E-ACT/M-CTG	38	22	57.9
GGGACAATAGGGGCACAACAA/M-CAG	111	85	76.6	E-AAG/M-CTA	39	20	51.3
GTTGTGCCCTATTGTCCCAT/M-CAC	83	62	74.7	E-AGG/M-CAC	55	33	60.0
Total	232	175	75.4	Total	132	75	56.8

ing the manufacturers protocol. DNA concentrations were determined relative to a molecular weight marker (High DNA Mass Ladder, Invitrogen) on 0.8% (w/v) agarose gel.

SSAP and AFLP analysis

The SSAP technique described by WAUGH *et al.*, (1997), is an anchored AFLP approach that amplifies the region between specific retrotransposon primers originated from the highly conserved LTR terminus, and a nearby cleaved restriction site, to which an oligonucleotide adaptor is added. The pre-amplification-PCR product is subsequently used as a template for the selective amplification with the *MseI* primer containing three selective bases at the 3' end, in combination with a labelled LTR-derived primer, based on the LTR sequences of *P. pinaster* Ty3 gypsy retrotransposons PpRT1 (ROCHETA *et al.*, 2006, Table 2), and Applied Biosystems AFLP protocol touch-down PCR conditions. The primers were designed using the program Primer-Select 5.00 (DNASStar Inc.) and synthesized by Applied Biosystems complementary to the LTR sequence and as close as possible to the polypurine tract (PPT), in order to minimize the expected size of polymorphic PCR products and labelled with a 6-FAM at the 5' end for detection in the ABI PRISM® 310 genetic analyser.

AFLP analysis, made with the Applied Biosystems Kit and protocol, and was based on a two-step amplification strategy with *EcoRI* and *MseI* primer adaptors (VOS *et al.*, 1995). In the first step, restriction fragments are produced by using two restriction endonucleases, and double stranded adaptors ligated to the DNA fragments ends, generating the template for subsequent polymerase chain reaction amplification. Pre-amplification consisted of a PCR reaction using primers complementary to the *EcoRI* and *MseI* adaptors with additional selective 3' nucleotides (*EcoRI* primer E_{+A} and *MseI* primer M_{+C}). Afterwards, diluted PCR products were used as templates for the second amplification using three primer combinations with three selective nucleotides where the *EcoRI* primer was fluorescently labelled.

Eight *EcoRI* +3 primers and *MseI*+ 3 primers were combined in various pairs in 37 PCR primer combinations, and tested on five samples from different populations. Three PECs were chosen for final analysis based on repeatability, polymorphism and for pines it is recom-

mended that for *EcoRI* primers extensions begin with -AC or -AG and for *MseI* the extension begins with -C (MYBURG and REMINGTON, 2000). Following selective amplification, AFLP fragments were denatured by adding deionised formamide, and run on GeneScan-500 (ROX) (Applied Biosystems) against a size standard. Fragment separation was performed by capillary electrophoresis using an automated sequencer ABI 310 (Applied Biosystems). In order to directly compare the SSAP with the AFLP marker system, the same pre-amplification PCR products produced with the *EcoRI* and *MseI* were used in the both experiments.

Data scoring and analysis

AFLP and SSAP patterns were visualized with Genescan 3.1 software (Applied Biosystems), and the presence (1)/absence (0) of AFLP and SSAP fragments was scored using the Genotyper 2.5 Software (Applied Biosystems). Only loci with clearly amplified fragments were used for data analysis.

Genetic diversity and differentiation were calculated using AFLPSURV v.1.0 (VEKEMANS *et al.*, 2002). Estimates of allelic frequencies at AFLP and SSAP loci were calculated using the Bayesian method with a uniform prior distribution of allele frequencies following ZHIVOTOVSKY (1999), without deviation from Hardy-Weinberg equilibrium. Gene diversity and population genetic structure were computed according to LYNCH and MILLIGAN (1994) for both marker types unbiased gene diversity within population, H_e ; average for all populations, H_S ; unbiased total gene diversity, H_T (NEI, 1987); and Wright's F_{ST} . For the two marker-systems, H_S was compared with a Z-test, and the variance associated with H_S included the component associated with the sampling of the populations.

Diversity assessed at different markers results from either 'marker specific' factors (e.g. mutation) or 'marker non-specific' factors (e.g. drift or migration). For example, a reduction in population size will decrease the level of diversity of any marker (SSAP or AFLP) despite differences in mutation rates. Consequently, one would expect that populations with different evolutionary histories display different levels of diversity, but follow a similar trend for both markers. For this reason, the comparative analysis of diversity between markers is not based on the comparison of levels of population diversity, but on the comparison of ranking of popula-

tions by correlation analysis (see MARIETTE *et al.*, 2001), by computing the Spearman's rank coefficient correlation (SOKAL and ROHLF, 1995), and congruence between markers would be supported by a significant correlation.

Additionally, the significance of genetic differentiation between populations was tested by comparison of the observed F_{ST} with a distribution of F_{ST} under the hypothesis of no genetic structure, obtained by using 1.000 random permutations of individuals among populations.

A principal component analysis (PCA) was performed based on the Nei's genetic distance after LYNCH and MILLIGAN (1994), using the NTSYS pc software (Version 2.1, (ROHLF, 1993)). A neighbour-joining consensus tree was computed using the PHYLIP v. 3.66 software (FELSENSTEIN, 2004), based on 1.000 bootstrap similarity matrices of Nei's genetic distance, after LYNCH and MILLIGAN (1994) obtained with the AFLPSURV software. The degree of relatedness between the similarity matrices was measured using a Mantel matrix-correspondence test (MANTEL, 1967). In addition, the association of the geographic and genetic distances with both markers was tested with the same method by using the NTSYS software v. 2.11 (ROHLF, 1993).

Results

SSAP and AFLP fragments

The size of SSAP amplified fragments varied from 36 to 492 bp. The three primer sets yielded a total of 232 scored fragments, 75.4% (175) of which were polymorphic (Table 2).

For AFLPs, 132 fragments were produced, varying from 38 to 55, depending on the PEC. Also the proportion of polymorphic fragments was mostly even across AFLP PECs, varying from 51.3 to 60% (Table 2). The three PECs produced a total of 75 (56.8%) polymorphic fragments, less than half the number of polymorphic fragments produced by SSAP. Finally, the size of AFLP amplified fragments ranged from 35 to 500 bp.

The selected primer combinations amplified, on average, about 111 SSAP and 49 AFLP fragments, per individual. Considering SSAP, Quinta de Sousa had the highest number of polymorphic loci (N_p), 147, and Pai do Sobrado had the lowest N_p (89), and for AFLP Quinta de Sousa also had the highest N_p (71) and Ponte de Lima the lowest (31) (Table 3).

Intra and inter-marker analysis

Within population diversity estimated for both marker systems was very similar, Monte Novo population exhibited the highest gene diversity levels, $H_e = 0.2802$ for SSAP and $H_e = 0.3270$ for AFLP, while Ponte Lima population displayed the lowest $H_e = 0.2571$ and $H_e = 0.2541$, respectively (Table 3), with the higher and lower values significantly different from each other for each marker system ($p < 0.05$, SSAP and $p < 0.001$, AFLP). When we compared diversity levels between markers, the values were not significantly different, except for the Monte Novo and Pai do Sobrado populations, with $p < 0.05$ and $p < 0.001$, respectively; also the average diversity was not significantly different for both markers (Table 3).

Table 4. – Comparison of the within population diversity with the AFLP and the SSAP markers in the seven *P. pinea* populations. Populations having the same letter do not show a significant difference in their level of diversity.

Populations	SSAP	Populations	AFLP
Monte Novo	0,28029 a	Monte Novo	0,32700 a
Viseu	0,27911 a	Pai Sobrado	0,29722 b
Pai Sobrado	0,26626 ab	Mata Valverde	0,27233 bc
Mata Valverde	0,26483 ab	Quinta Sousa	0,26661 bc
Palma	0,26164 ab	Palma	0,26618 bc
Quinta Sousa	0,25548 ab	Viseu	0,25893 c
Ponte Lima	0,25410 b	Ponte Lima	0,25710 c

Table 3. – Gene diversity estimates obtained with the SSAP and AFLP marker systems for the seven *P. pinea* populations, based on 132 AFLP and 232 SSAP markers. H_e = within population diversity, H_T = total gene diversity, F_{ST} = differentiation between populations, SE (H_e) = standard error associated to H_e , N_p = number of polymorphic loci (direct count) and P = proportion of polymorphic loci. Differences between both markers are indicated by ns = not significant, * = $p < 0.05$ and *** = $p < 0.001$ and NA = not available.

Populations	Code	SSAP				AFLP			
		N_p	P (%)	H_e	SE (H_e)	N_p	P (%)	H_e	SE (H_e)
Ponte Lima	1	115	49.6	0.2541 ^{ns}	0.0091	31	23.5	0.2571	0.0102
Viseu	2	118	50.9	0.2791 ^{ns}	0.0093	33	42.4	0.2599	0.0113
Pai Sobrado	3	89	38.4	0.2662 [*]	0.0098	56	42.4	0.2972	0.0009
Monte Novo	4	91	39.2	0.2802 ^{***}	0.0088	56	42.4	0.3270	0.0088
Mata Valverde	5	106	45.7	0.2648 ^{ns}	0.0087	48	36.4	0.2723	0.0093
Quinta Sousa	6	147	63.4	0.2554 ^{ns}	0.0102	71	53.8	0.2666	0.0124
Palma	7	113	48.7	0.2616 ^{ns}	0.0097	50	37.9	0.2662	0.0123
average		111	48	0.2660 ^{ns aj}	0.0039	49	37	0.2279 ^{aj}	0.0096
H_T				0.2970 ^{NA}				0.3004	
F_{ST}				0.1058 ^{NA}				0.0747	

a) H_s = mean gene diversity within populations.

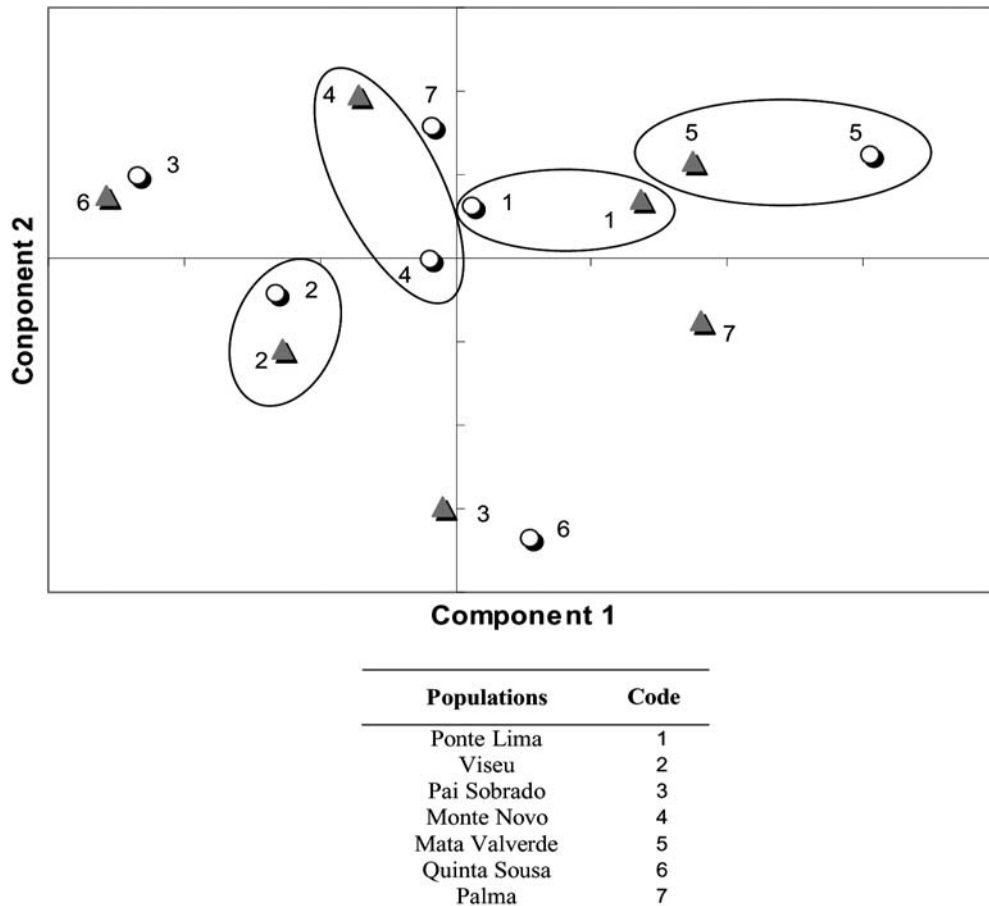


Figure 1. – Principal components analysis based on AFLP (triangles) and SSAP (circles) pairwise Nei's genetic distances of the seven studied populations (populations' code as in Table 3).

In Table 4, the multipopulation statistical comparison of within population diversity is displayed, showing that AFLPs detected more differences among populations than SSAP. With AFLPs, Monte Novo's genetic diversity was significantly different from all other populations' estimates, but not for SSAP where only Ponte de Lima was significantly different.

Additionally, population rankings for H_e calculated by AFLP were significantly correlated with the rankings obtained for SSAP (the Spearman correlation index was $r_s = 0.943$ and $p = 0.002$), when the outlier population was removed from the analysis (Viseu) (see populations sorted by diversity values in Table 4). In fact, the Viseu population was responsible for the biggest difference in population ranking for the two markers.

Genetic differentiation was slightly higher for SSAP ($F_{ST} = 0.105$) than for AFLP ($F_{ST} = 0.074$), and both were significantly different from zero, $p < 0.01$. Based on PCA analysis, the percentage of total variation explained by the two first axes was 37% and 23% (Figure 1). Population placement was similar for SSAP and AFLP, except for populations 3, 6 and 7 (see Figure 1).

Furthermore, a normalized Mantel statistic computed a correlation of $r = 0.71$ for the association of the genetic distances computed with both markers, with a probability of $P \leq 0.001$. This result indicates that both matrices are correlated and that they detected similar trends with respect to genetic distances.

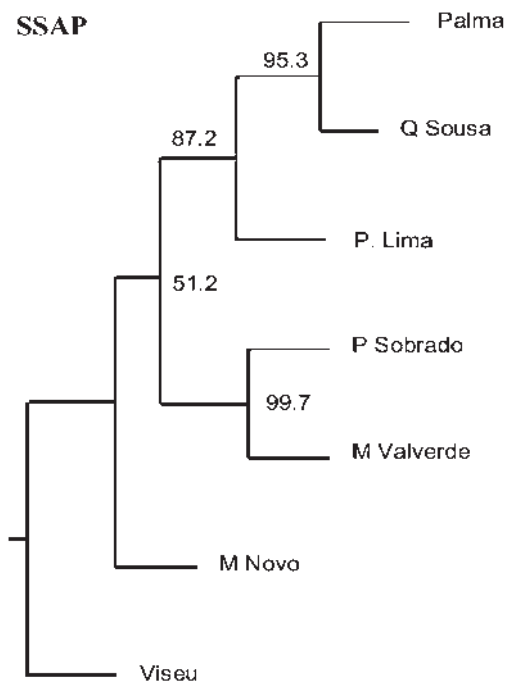


Figure 2. – Neighbour-joining consensus tree computed with the pairwise Nei's genetic distance matrix using 132 AFLP markers from the seven *P. pinea* populations. The numbers on the branches in the consensus tree indicate the percentage of times the partition of the populations into the two sets which are separated by that branch occurred among the trees, out of 1,000 trees.

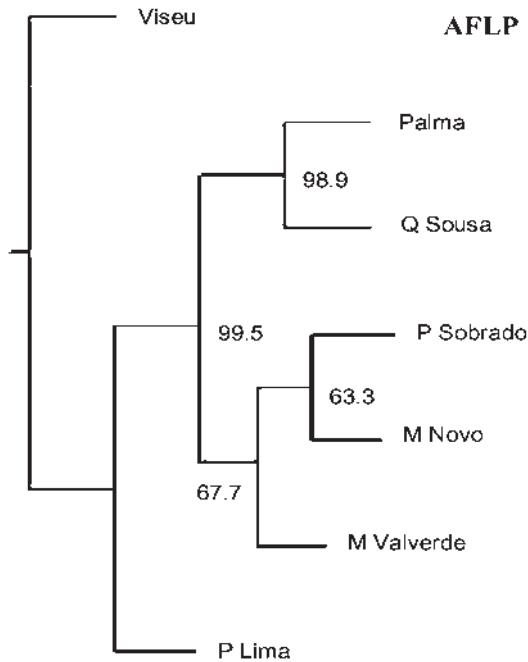


Figure 3. – Neighbour-joining consensus tree computed with the pairwise Nei's genetic distance matrix using 232 SSAP markers from the seven *P. pinea* populations. The numbers on the branches in the consensus tree indicate the percentage of times the partition of the populations into the two sets which are separated by that branch occurred among the trees, out of 1,000 trees.

The neighbor-joining trees based on the Nei's genetic distance are shown in Figure 2 (AFLP) and 3 (SSAP), and some geographical pattern is observed. With the AFLP markers the populations grouped according to provenance region from which they were sampled, resulting in one well supported node clustering the Southern populations, but with SSAP the pattern was not so coherent. Nevertheless, no association was found between genetic and geographical distances.

Discussion

Long terminal repeat (LTR) retrotransposons are mobile genetic elements that are ubiquitous in plants and represent a significant genomic fraction in plant species with large genomes, as is the case for *Pinaceae*. There are no studies regarding the potential influence of selection on the transposon-derived loci. The factors governing the activation and proliferation of these elements are poorly known, although increasing evidence suggests roles for natural agents such as interspecific hybridization and abiotic stress (M. C. Ungerer, pers. comm.). Future studies will indicate how plant genomes have evolved to utilize retrotransposon sequences (KUMAR and RZHETSKY, 1996).

SSAP vs. AFLP within population polymorphism

The same populations used in this study, were also analysed for RAPD and cpSSR, and, lower levels of genetic diversity were observed, in particular, with cpSSR, where only a single haplotype was obtained with six primers (EVARISTO *et al.*, 2002). In another region (Spain), only 9 haplotypes were found when 10 popula-

tions were screened with 6 cpSSR, (GÓMEZ *et al.*, 2002), while in other Pines higher numbers of haplotypes were found, 28 and 34 haplotypes, in 10 and 6 populations, respectively in *P. halepensis* (GÓMEZ, 1998) and in *P. pinaster* (VENDRAMIN *et al.*, 1998). Moreover, in a range wide study, including 34 populations, of *P. pinea* screened with twelve cpSSRs, only three were polymorphic, giving a total of only four different haplotypes. (G.G. Vendramin, pers. comm.). Therefore, the polymorphism obtained with AFLP and, especially, with SSAP, reveals important information for studies of diversity, conservation and breeding.

SSAP has been suggested as a suitable alternative to AFLP for assessing genetic diversity, differentiation, and polymorphism (SYED *et al.*, 2005). In this study SSAP generated more polymorphic fragments than AFLP. The number of polymorphic SSAP fragments was about one and a half times that at AFLP (Table 2). This result is in agreement with previous studies, for example, SSAP revealed two to threefold higher polymorphism per primer combination than AFLPs in *Hordeum vulgare* (WAUGH *et al.*, 1997), *Pisum* (ELLIS *et al.*, 1998), *Triticum aestivum* (QUEEN *et al.*, 2004), and *Anacardium occidentale* (SYED *et al.*, 2005). Moreover, SSAP was found to be the most informative system for estimating genetic diversity in tomato and pepper, in a study comparing SSAP with AFLP and SSR, producing about four- to nine-fold more polymorphic fragments than AFLP, and achieving the highest number of polymorphic bands per assay (MAY TAM *et al.*, 2005). In fact, one advantage of SSAP over AFLP is that higher levels of polymorphism are achieved (ELLIS *et al.*, 1998), and fewer experiments are needed to generate the desired numbers of markers. SSAP generates more polymorphic markers than AFLP, probably due to the higher retrotransposition mutation rate relative to base mutation.

Some marginal populations, Ponte Lima and Viseu, could be defined as being relatively isolated from the main area of species distribution and show a lower level of variability, particularly for AFLPs.

Congruence in SSAP vs. AFLP

Population diversity rankings obtained with both markers were similar. Population diversity estimates differ in similar ways, despite the fact that, in some cases, they display significantly different values of within-population diversity. However in a recent meta-analysis of different marker systems by KREMER and MARIETTE (2003), diversity estimates were significantly correlated in less than 12% of cases. MARIETTE *et al.*, (2002) using simulations, suggested that this lack of correlation may be due to different sampling strategies of markers within the genome, low number of loci used, low differentiation among populations, and non-equilibrium situations (recently created populations). In our case, at least the first condition was met (high number of loci), in addition differentiation was moderate, but there is not enough data on the equilibrium situation. However the correlation between SSAP and AFLP was significant only when the population responsible for the highest difference in ranking was removed from the analysis (Viseu).

Differentiation and populations clustering

The degree of differentiation among populations depends on the age of population separation. In theory, differentiation is relatively independent of marker system used, (KREMER and MARIETTE, 2003). Therefore, differentiation depends mainly on population size and rate of migration, factors acting with the same amplitude on different markers, which has been confirmed by experimental results comparing differentiation values obtained with different markers (HAMRICK *et al.*, 1992; NYBOM, 2004). In the present study, the differentiation values obtained with both markers were of similar magnitude, and the higher mutation rates associated with SSAP had no impact on the estimate.

The Mediterranean basin is characterized by intense human activities, which impact on various forest tree species, modifying genetic diversity and phylogeographic structure (LEDIG, 1992), which is reflected in the pattern of diversity and differentiation in *P. pinea* and other pine species of this region (e.g. *P. pinaster*, RIBEIRO *et al.*, 2001). A strong anthropogenic influence associated with extensive gene flow could explain these levels of differentiation among populations and the diversity found mainly within populations.

Comparing the genetic differentiation estimates from this study with ones obtained for other pines, reveals that the values within *P. pinea* are typical of outcrossing species $G_{ST}=0.065$, (HAMRICK *et al.*, 1992) and with *P. pinaster*, for the same markers, $G_{ST}=0.077$, (SALVADOR *et al.*, 2000); $G_{ST}=0.16$, (PETIT *et al.*, 1995).

In this study, SSAP generated more polymorphic fragments and similar estimates of genetic diversity than AFLP. Moreover, congruence was found between estimates obtained with both markers. In general, SSAP markers might be an interesting alternative marker system, when higher resolution is requested, but AFLP was more discriminating than SSAP.

Moreover, retrotransposon markers generate taxonomic data that are more consistent with morphological criteria than AFLP-based markers (ELLIS *et al.*, 1998). The reason for this is, yet, unclear, but according to QUEEN *et al.*, (2004) it may reflect two facts. (1) Retrotransposon insertion is a single type of biological event, whereas the different mutations detected as AFLP polymorphism can arise in any type of DNA sequence, and a polymorphism in a highly repeated sequence is unlikely to show the same phylogenetic properties of an insertion in a gene. (2) Retrotransposon insertions are irreversible, since the presence of an insertion (occupied site) is a derived site, while the absence of the insertion is the ancestral condition. Actually, SSAP bands might be lost by sequence change at restriction or primer sites but a component of the polymorphism is, nevertheless, irreversible and this may contribute to the enhanced performance of this method relative to AFLP. In addition, SSAP markers have lower costs compared to AFLPs, for only one LTR primer is fluorescently labelled in combination with standard adapter-based primers, and the number of assays for achieving the same number of polymorphic bands is lower. The disadvantage of SSAP for a newly studied species is the need to find retrotransposon sequences for primer design.

Despite comprising the majority of many plant genomes, and their role in the mutation and evolution of genes and genomes, only a handful of laboratories investigate these elements. Much future work is needed, and the tools are all in hand, so we can expect continuing great gains in our understanding of plant retrotransposons for the foreseeable future.

Acknowledgements

The authors thank to Dr. CÉLIA MIGUEL and JORGE CORDEIRO for providing LTR transposable element sequence from *P. pinaster*. We also thank Dr. MARGARIDA RIBEIRO for advice with data analysis and for the revision of the manuscript and to CARLA RIBEIRO for help with data analysis. This study was supported by Project PIDDAC-203: "Retrotransposons to assess genetic diversity in *Pinus pinea* L."

References

- CARDOSO, M. M. and P. A. LOBO (2001): Delimitação de pisos bioclimáticos e regiões de proveniência de pinheiro manso em Portugal, usando sistemas de informação geográfica, *Silva Lusitana* **9**: 93–108.
- CARNEIRO, A. I. L. N. (2002): Avaliação da variabilidade e controlo genético de características biométricas das pinhas e da semente da *Pinus pinea* L., Relatório do trabalho de fim de Curso de Engenharia Florestal, Universidade Técnica de Lisboa. Instituto Superior de Agronomia, Lisbon, 58 p.
- ELLIS, T. H. N., S. J. POYSER, M. R. KNOX, A. V. VERSHININ and M. J. AMBROSE (1998): Polymorphism of insertion sites of TY1-copia-class retrotransposon and its use for linkage and diversity analysis in pea, *Molecular General Genetics* **260**: 9–19.
- EVARISTO, I. N., R. SEABRA, J. BAETA and M. S. PAIS (2002): Caracterização molecular de proveniências de *Pinus pinea* L. por RAPD (Random Amplified Polymorphic DNA), *Silva Lusitana* **10**: 53–61.
- FALLOUR, D., B. FADY and F. LEFEVRE (1997): Study on isozyme variation in *Pinus pinea* L.: Evidence for low polymorphism, *Silvae Genetica* **46**: 201–207.
- FELSENSTEIN, J. (2004): PHYLIP (Phylogeny Inference Package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- GÓMEZ, A. (1998): Variabilidade genética de *Pinus halepensis* em Espanha mediante el uso de marcadores de ADN: RAPDs e Cp-microsatélites, PhD thesis, Universidad Politécnica de Madrid, Madrid, 162 p.
- GÓMEZ, A., E. AGUIRIANO, R. ALIÁ and M. A. BUENO (2002): Análisis de los recursos genéticos de *Pinus pinea* L. en España mediante microsatélites de cloroplasto, *Investigación Agraria: Sistemas y Recursos Forestales* **11**: 145–154.
- HAMRICK, J., M. GODT and S. SHERMAN-BROYLES (1992): [Factors influencing levels of genetic diversity in woody plant species](#), *New Forests* **6**: 95–124.
- KREMER, A. and S. MARIETTE (2003): Molecular markers in population genetics. In: DE VIENNE, D. (ed. *Molecular Markers in Plant Genetics and Biotechnology*, pp. 126–149. Science Publishers, Inc., New York.
- KUMAR, A. (1996): The adventures of Ty1-copia group of retrotransposons in plants, *Trends in Genetics* **12**, 41–43.

- KUMAR, A. and H. HIROCHIKA (2001): Application of retrotransposons as genetic tools in plant biology, *Trends in Plant Sciences* **6**: 127–113.
- KUMAR, A., S. R. PEARCE, K. MCLEAN, G. HARRISON, J. S. HESLOP-HARRISON, R. WAUGH and A. J. FLAVELL (1997): The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers, *Genetica* **100**: 205–217.
- KUMAR, S. and A. RZHETSKY (1996): Evolutionary relationships of eukaryotic kingdoms, *Journal of Molecular Evolution* **42**: 183–193.
- LEDIG, F. T. (1992): Human impacts on genetic diversity in forest ecosystems, *Oikos* **63**: 87–108.
- LYNCH, M. and B. G. MILLIGAN (1994): Analysis of population genetic structure with RAPD markers, *Molecular Ecology* **3**: 91–99.
- MANTEL, N. A. (1967): The detection of disease clustering and a generalized regression approach, *Cancer Research* **27**: 209–220.
- MARIETTE, S., D. CHAGNE, C. LEZIER, P. PASTUSZKA, A. RAFFIN, C. PLOMION, and A. KREMER (2001): Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers, *Heredity* **86**: 469–479.
- MARIETTE, S., V. LE CORRE, F. AUSTRERLITZ and A. KREMER (2002): Sampling within the genome for measuring within-population diversity: trade-offs between markers, *Molecular Ecology* **11**: 1145–1156.
- MAY TAM, S., C. MHIRI, A. VOGELAAR, M. KERVELD, S. R. PEARCE and M. A. GRANDBASTIEN (2005): Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR, *Theoretical and Applied Genetics* **110**: 819–831.
- MIROV, N. T. (1967): The genus *Pinus*. The Ronald Press Company, New York, 602 p.
- MUELLER, U. G. and L. L. WOLFENBARGER (1999): AFLP genotyping and fingerprinting, *Trends in Ecology and Evolution* **14**: 389–394.
- MUTKE, S., J. GORDO and L. GIL (2000): The stone pine (*Pinus pinea* L.) breeding programme in Castile-Leon (Central Spain), *FAO-CIHEAM NUCIS-Newsletter* **9**: 50–55.
- MYBURG, A. A. and D. L. REMINGTON (2000): Protocol for high throughput AFLP analysis using LI CORIR2 automated sequencer. In: NC State forest biotech AFLP protocol. North Carolina State University Forestry Department, Raleigh.
- NEI, M. (1987): *Molecular evolutionary genetics* Columbia University Press, New York, 512 p.
- NYBOM, H. (2004): Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants, *Molecular Ecology* **13**: 1143–1155.
- PEARCE, S. R., G. HARRISON, D. LI, J. S. HESLOP-HARRISON, A. KUMAR and A. J. FLAVELL (1996): The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation, *Molecular General Genetics* **250**: 305–315.
- PEARCE, S. R., C. M. STUART-ROGERS, A. KUMAR, T. H. N. ELLIS and A. J. FLAVELL (1999): Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies., *Plant Journal* **19**: 711–717.
- PETIT, R. J., N. BAHMAN and P. BARADAT (1995): Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci, *Heredity* **75**: 382–389.
- QUEEN, R. A., B. M. GRIBBON, C. JAMES, P. JACK and A. J. FLAVELL (2004): Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat, *Molecular Genetics and Genomics* **271**: 91–97.
- QUEZÉL, P. (1980): Biogéographie et écologie des conifères sur le pourtour méditerranéen. In: *Actualités d'écologie forestière*. Sol, flore, faune PESSON, P. (ed.), pp. 205–255. Gauthiers-Villars, Paris.
- RIBEIRO, M. M., C., PLOMION, R. PETIT, G. G. VENDRAMIN and A. E. SZMIDT (2001): Variation of chloroplast simple-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.), *Theoretical and Applied Genetics* **102**: 97–103.
- ROCHETA, M., J. CORDEIRO, M. OLIVEIRA and C. MIGUEL (2006): PpRT1 (*Pinus pinaster* retrotransposon 1): the first gypsy-like retrotransposon isolated in *Pinus pinaster*, *Planta* **225**: 551–562.
- ROHLF, F. J. (1993): *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System (V 2.1)*. Exeter Publisher, New York p.
- SALVADOR, L., R. ALÍA, D. AGÚNDEZ and L. GIL, (2000): Genetic variation and migration pathways of maritime pine (*Pinus pinaster* Ait.) in the Iberian Peninsula, *Theoretical and Applied Genetics* **100**: 89–95.
- SCHULMAN, A. H., A. J. FLAVELL and N. T. H. ELLIS (2004): The application of LTR retrotransposons as molecular markers in plants. In: MILLER, W. J., CAPY, P. (eds). *Mobile genetic elements protocols and genomic applications*, pp. 145–158. Humana Press Inc., Totowa.
- SOKAL, R. R. and F. J. ROHLF (1995): *Biometry*, 3rd ed. edn. W. H. FREEMAN and Co., San Francisco, 887 p.
- STEVENSON, A. C. (1985): Studies in the vegetational history of SW Spain. II. Palynological investigations at Laguna de Las Madres, SW Spain, *Journal of Biogeography* **12**: 293–314.
- SYED, N. H., M. J. SURESHSUNDAR, B. S. BHANU, J. J. V. CAVALCANTI and A. J. FLAVELL (2005): Ty1-copia retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.), *Theoretical and Applied Genetics* **110**: 1195–1202.
- VEKEMANS, X., T. BEAUWENS, M. LEMAIRE and I. ROLDAN-RUIZ (2002): Data from amplified fragment polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size, *Molecular Ecology* **11**: 139–151.
- VENDRAMIN, G. G., M. ANZIDEI, A. MADAGHIELE and G. BUCCI (1998): Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites, *Theoretical and Applied Genetics* **97**: 456–463.
- VOS, P., R. HOGERS, M. BLEEKER, M. REIJNS, T. VANDELEE, M. HOMES, A. FRIJTERS, J. POT, J. PELEMAN, M. KUIPER and M. ZABEAU (1995): AFLP: A new technique for DNA fingerprinting, *Nucleic Acids Research* **23**: 4407–4414.
- WAUGH, R., K. MCLEAN, A. J. FLAVELL, S. R. PEARCE, A. KUMAR, B. B. T. THOMAS and W. POWELL (1997): Genetic distribution of Bare-1-like retrotransposable elements in barley genome revealed by sequence specific amplification polymorphisms (SSAP), *Molecular General Genetics* **253**: 687–694.
- YU, G. X. and R. P. WISE (2000): An anchored AFLP- and retrotransposon-based map of diploid Avena., *Genome Research* **43**: 736–749.
- ZHIVOTOVSKY, L. A. (1999): Estimating population structure in diploids with multilocus dominant DNA markers, *Molecular Ecology* **8**: 907–913.