SSR Markers for Monitoring an *in vitro* Core Collection of *Populus tremula*

By A. Gómez[2], J. A. Manzanera[2], E. Aguiriano[1], J. M. Grau[1] and M. A. Bueno[1]

(Received 30th May 2003)

Abstract

Core collections efficiently preserve genetic diversity and facilitate collection management and "ex situ" conservation of genetic resources. *Populus tremula* (European aspen) is demanded by plant growers and breeders but there are very few data on the genetic variation in the geographic distribution of this species in Spain. Microsatellite markers are useful for the evaluation of genetic variability and identification due to their abundance and polymorphism. Six pairs of oligonucleotide primers designed for simple sequence repeats (SSR) of quaking aspen (*P. tremuloides*) were successfully used in European aspen. Number of alleles per locus were low in this species in comparison to outcrossed, long-lived woody perennials due to the importance of vegetative propagation by root suckers in aspen stands. Mean heterozygosity (0.475) was similar to that reported in quaking aspen. Cluster analysis suggests that postglacial migration of European aspen took place from the Pyrenees in two directions: westward and southward through mountain ranges of the Iberian Peninsula. To preserve the allelic richness of the *P. tremula in vitro* collection in Spain, a core collection of 26 clones is proposed.

Key words: allelic richness, core collection, genetic resources, microsatellites, *Populus tremula.*

Introduction

The endangerment of the European aspen populations from Spanish origin has urged the start of a strategy for the conservation of genetic resources. Collections via *"in vitro"* culture of forest trees are repositories of the biodiversity available for each species. They are also a valuable source of useful plants for preservation of biodiversity and genetic resources programs specially with species, which produce recalcitrant seeds, and those that are vegetatively propagated (Bueno et al., 2001; Thorpe and Malhotra, 1996). The production of "core" collections would be the solution to enhance efficiency in the preservation of genetic diversity and the management of collections. The representativeness of a core collection can be defined in two ways: how close it is to the full range of variation present in the whole collection (allelic diversity) and how closely the pattern of variation in the core collection resembles that in the whole one (pattern of allelic frequencies).

The core collection of *Populus tremula* L. would be the first one supplied in response to the demand by plant growers and breeders for material from the collection (Bueno et al., 2001). The "core" collection could also be the focus of a program of "ex situ" genetic resources conservation of *P. tremula.* Information about diversity within this collection can be obtained using molecular markers which afford many benefits for the identification of variation and the estimation of biological diversity.

Microsatellites have become the genetic markers of choice in mammalian and many plant systems because of their abundance, high degree of polymorphism and amenability to automation (Weber and May, 1989). Due to their co-dominant inheritance, simple sequence repeats (SSRs) have become the preferred tool for investigations of critical importance for germplasm managers, such as the establishment of unique genetic identities or fingerprints, determination of genetic relatedness between accessions, and the assessment of genetic diversity contained within a collection (Hokanson et al., 1998). For this purpose, SSR markers have been analysed in a number of species: sweet cherry (Schueler et al., 2003); cork oak (Gomez et al., 2001); soybean (Rongwen et al., 1995); grape (Lamboy and Alpha, 1998); apple (Guilford et al., 1997); shorgum (Brown et al., 1996); barley (Russell et al., 1997); avocado (Lavi et al., 1994), etc.

This work analyses the genetic diversity of the *in vitro* collection of *P. tremula* from Spanish origin and the allelic richness detected by SSR markers in order to determine the core collection of this species following the M strategy (Bataillon et al., 1996): the core collection includes trees from the different regions of origin and maximises allelic richness. This species does not form extended forests in Spain but appears in small groups integrated in forests of other species (Ruíz de la Torre, 1979). Distance between samples locations range from 2 km to 63 km inside each region of origin (Sánchez et al., 2000) and sample size has been chosen proportionally to the abundance of European aspen in each region.

Material and Methods

Plant material and DNA extraction

Eighty-two individuals of *Populus tremula* selected from 8 Spanish provinces representing the main area of distribution were used for the present study (*Figure 1*): One located in the Central Mountain Range (Segovia; 10 individuals); one in the Iberian Mountains (Cuenca; two individuals); two near the Pyrenees (Northeast: Huesca, 26 individuals; and Navarra, 12 individuals); and four in the Northwest (Lugo-Orense, five individuals; Zamora, five individuals, and León, 22 individuals).

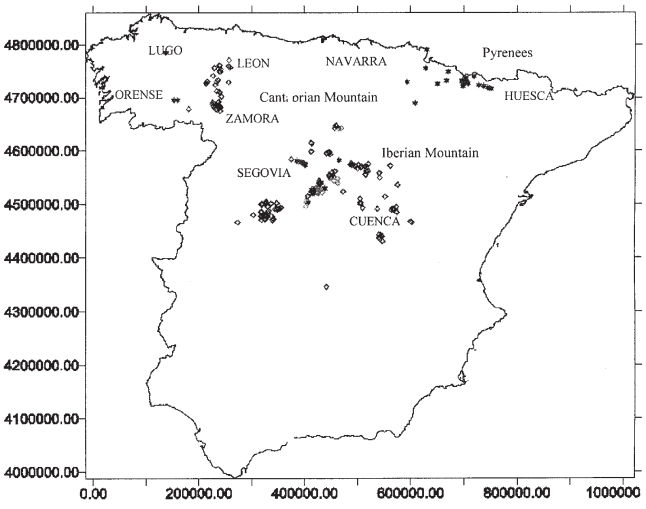

Figure 1. – Samples location of *Populus tremula* "in vitro" collection from Spain.

[1] INIA-CIFOR; Ctra. de La Coruña Km 7; E-28040 Madrid, Spain.

[2] IMIA, Apdo. 127, E-28800 Alcalá de Henares, Spain.

Communicating author: Bueno, M. A. E-mail: bueno@inia.es.
Telephone: 00 37 91 347 6863. Fax number: 00 37 91 357 2293

Table 1. – Repeat pattern and size range of alleles at 6 microsatellite loci (PTR1, PTR2, PTR3, PTR4, PTR5 and PTR6) in *Populus tremuloides*. (DAYANANDAN *et al.,* 1998 and RAHMAN *et al.,* 2000) and *Populus tremula* (this work).

Locus	*Primer sequence (5'→3')*	*Repeat*	*Alleles sizes (bp) In P. tremuloides*	*Alleles sizes (bp) In P. tremula*
PTR1	AGCGCGTGCGGATTGCCATT TTAGTTTCCCGTCACCTCCTGTTAT	$(GGT)_5N_{45}(AGG)_9$	254-272	254-260
PTR2	AAGAAGAACTCGAAGATGAAGAACT ACTGACAAAACCCCTAATCTAACAA	$(TGG)_8$	204-225	204-214
PTR3	CACTCGTGTTGTCCTTTTCTTTTCT AGGATCCCTTCCCTTTAGTAT	$(TC)_{11}$	190-266	216-238
PTR4	AATGTCGAGGCCTTTCTAAATGTCT GCTTGAGCAACAAACACACCAGATG	$(TC)_{17}$	200-230	200-202
PTR5	CTTCTCGAGTATAAATATAAAACACCA TCACATCACCCTCTCAGTTTCGC	$(TG)_7$	250-256	242-260
PTR6	AGAAAAGCAGATTGAGAAAAGAC CTAGTATAGAGAAAGAAGAAGCAGAAA	$(A'_{4\ 8}$	186-240	190-202

Approximate age of the donor trees ranged between 40 and 60 years.

Leaf samples from 51 trees were collected in 2000 from the *in vitro* collection of *P. tremula* located in the laboratory of CIFOR-INIA in Madrid (Spain), which was established between 1996 and 1999. Branch material from the remaining 31 clones were collected directly from the trees. These trees were collected from Orense (four individuals), Zamora and León.

DNA was extracted following the method described in DOYLE and DOYLE (1990).

DNA amplification

Six microsatellites were amplified using the polymerase chain reaction (PCR). Oligonucleotide primers complementary to flanking regions of SSR from *P. tremuloides* genome (*Table 1*) were designed by DAYANANDAN et al. (1998) and RAHMAN et al. (2000). Each 25 µl amplification reaction contained: 20 ng of total DNA, 0.2 µM of fluorescently labelled forward primer and unlabelled reverse primer (Progenetic), 200 µM each dNTPs, 50 mM KCl, 10 mM Tris-HCl (pH 9), 2.5 mM $MgCl_2$ and 0.5 U of Taq-DNA polymerase (Ecogen). Fluorescently labelled PCR products were separated and analysed on a semiautomated sequencer (ABI-Prism, Perkin-Elmer). Standards were used for length determination of alleles.

Amplifications were carried out in a Perkin Elmer 9600 thermocycler, with two profiles: for PTR1, PTR2, PTR3 and PTR4 *loci* 1 min at 94 °C, 5 cycles 1 min at 94 °C (denaturation), 1 min 55 °C (annealing) and 1 min 72 °C (extension) followed by 30 cycles 30 s at 94 °C, 30 s at 60 °C and 30 s at 72 °C. For PTR5 and PTR6 amplification was carried out following RAHMAN et al. (2000).

SSR analysis

An aliquot of the amplification reactions was visualised under UV light, after running in 1.5% agarose gels in 1x TAE buffer and ethidium bromide staining.

Another aliquot was analysed with an ABI-Prism automatic sequencer (Perkin Elmer). From each sample three SSRs were loaded and run together with internal molecular weight standards (Genescan/350 TAMRA). Resulting fragment sizes were calculated with the Genescan software (Perkin Elmer) by comparison with the standards.

Statistical analysis

Each genotype was defined as the combination of the 6 SSRs amplified fragments.

Heterozygosity was calculated using NEI's (1973) formula: $H = n/(n-1)^* (1-\Sigma p_i^2)$, where p_i is the frequency of allele i in the n analysed trees. Power of discrimination for each *locus* was calculated using the formula: $PD = 1-\Sigma P_i^2$ where P_i is the frequency of genotype i (KLOOSTERMAN et al., 1993). Effective alleles per *locus* were calculated according to WEIR (1990) with the formula $1/(1-H_{ep})$, were H_{ep}, the genetic diversity per *locus*, is equal to $1-\Sigma p_i^2$.

Genetic relationships among the 82 trees were investigated using an unweighted pair-group method (UPGMA) cluster analysis computed with the program NTSYS-pc (RHOLF, 1994).

The probability that two individuals drawn at random from a population have the same multilocus genotype, P_{ID}, was determined as follows (WAITS et al., 2001):

$$P_{ID} = \frac{n^3 (2a_2^2 - a_4) - 2n^2 (a_3 + 2a_2) + n(9a_2 + 2) - 6}{(n-1)(n-2)(n-3)}$$

Where a_k, with k being the power of the allelic frequencies as follows:

$$a_k = \sum^m_{i=1} p_i^k$$

Where m is the number of different alleles and p_i is the frequency of the ith allele in the population.

Strategy for core collection

The M strategy was selected (BATAILLON et al., 1996) to construct the core collection employing SSR markers. The M strategy includes the use of genetic markers to compute marker allelic richness of each putative core collection. The core collection is constructed under two constraints: first, the inclusion in each putative core collection of at least one accession per region and second, maximise the marker allelic richness of the core collection.

Results

All six SSR markers from *P. tremuloides* have been successfully amplified and sized in DNA from *P. tremula.* All six *loci* were polymorphic in the 82 trees analysed. Number of alleles per *locus* varied from 2 to 5 (*Table 2*), the average being 3.5 alleles per *locus*. The total number of detected alleles is 21.

Table 2 shows the Hardy-Weinberg expected heterozygosity (H), which ranges between 0.115 and 0.659 per *locus.* The average value of this index for all six *loci* studied was 0.475. The power of discrimination between genotypes was 99.2%. The low probability of identity ($P_{ID} = 2.7 * 10^{-4}$) allowed us to consider

individuals with the same multilocus genotype to be potential ramets of the same clone.

A total of 65 genotypes have been detected among the 82 sampled trees with the 6 SSRs analysed. Genotypes called No. 7 and No. 15 (*Table 3*) were the most abundant genotypes with 4 and 5 trees respectively. In general, when different trees have the same genotype, they are located at closer distances between them than other sampled trees. For instance, three trees from Segovia have the genotype No. 7 and trees with genotype No. 15 are from León and Zamora. Nearly a 66% of trees have a unique genotype.

Table 2. – Heterozygosity (H), Discrimination Power (DP), Number of alleles and Effective number of alleles of 6 SSR *loci* in *Populus tremula* from Spain.

	H	DP	Number of alleles	Effective number of alleles
PTR1	0.115	0.182	2	1.13
PTR2	0.535	0.718	3	2.13
PTR3	0.497	0.612	4	1.98
PTR4	0.468	0.647	4	1.87
PTR5	0.578	0.667	3	2.35
PTR6	0.659	0.808	5	2.89
Mean	0.475	0.606	3.5	2.06
Total		0.992	21	

Table 3. – *P. tremula* genotypes detected by SSR markers that belongs to more than one sampled tree.

Number	Genotype (sizes in bp)	Frequency
1	254 204/210 216/238 202 250/260 202	0.024
7	254 204 216 200 242 202	0.049
8	254 204 216 200 260 196/202	0.024
15	254 204 216 202 250 200	0.061
17	254 204 216 202 250 196/202	0.037
29	254 210 216/230 200 250 190/202	0.024
34	254 204/210 216 200 260 202	0.024
39	254 204/210 216 200/202 260 196/202	0.024
61	260 204/214 218 200 250/260 196	0.024
62	254/260 204/214 216/230 200 250 190/202	0.024
64	254/260 204/214 216/230 200/202 260 202	0.024

The phenogram generated from the UPGMA cluster analysis of the 82 trees resulted in several groupings which appear to be related with the geographical origin (*Figure 2*). The first group, which includes LU to LE19.1, are in general trees collected in the Northwest of the Iberian Peninsula. The second group includes trees from all locations but it is possible to distinguish inside this group a small sub-group called A which includes trees from Segovia and a subgroup called B that includes trees from León.

For the M strategy we analysed divisions at linkage distances 15, 10 and 5 in the phenogram (*Figure 2*) to select the individuals that will form the core collection. To obtain a handled "in vitro" core collection individuals have been selected in order to maximise allelic richness (*Table 4*). With division at linkage distance 5, the number of conserved alleles is the total number of alleles contained in the base collection (allelic richness).

The group at distance 15 includes individuals from the Northwest (two from Lugo-Orense and three from León) and the Pyrenees (two from Huesca and two from Navarra). The group at distance 10 includes six individuals from the Northwest (two from Lugo-Orense and four from León), six from the Pyrenees (three from Huesca and three from Navarra), one from the Central Mountain Range (Segovia), and one from the Iberian Mountains (Cuenca). The group at distance 5 includes ten individuals from the Northwest (three from Lugo-Orense, one from Zamora and six from León), eleven from the Pyrenees (six from Huesca and five from Navarra) and five from the Central area (two from Cuenca and three from Segovia).

Discussion

Six pairs of primers designed by DAYANANDAN et al. (1998) and RAHMAN et al. (2000) for microsatellite amplification in *P. tremuloides* were used for DNA analysis of *P. tremula*. Amplification profiles have been improved for PTR1, PTR2, PTR3 and PTR4 in order to obtain good amplification products in *P. tremula*. The use of microsatellite markers is generally restricted to species for which they are designed, due to the high degree of homology necessary between primers and sample DNA. The possibility of using four of those microsatellites as genetic markers in other species of the same genus was studied by DAYANANDAN et al. (1998) and only PTR2 and PTR4 had positive PCR amplifications in *P. deltoides, P. nigra, P. maximowiczii* and *P. x canadensis*. All six SSRs from *P. tremuloides* have resolution in *P. tremula* probably due to the close relationship between both species during evolution (PRIMMER et al., 1996; SUN and KIRKPATRICK, 1996; FIELDS and SCRIBNER, 1997; GÓMEZ et al., 2001). In fact, both quaking and European aspen have been placed in the subsection Trepidae of the section *Leuce* and hybridise very easily.

The mean number of alleles detected per *locus* in this study ranged from 2 to 5, which is lower than the mean number of alleles (3 to 11) detected by the same SSRs in *P. tremuloides* (DAYANANDAN et al., 1998; RAHMAN et al., 2000). These values are similar to those reported in plant literature for self-pollinating and/or annual crops but lower if they are compared with results reported for other outcrossed, long-lived woody perennials such as *Pinaceae* (Monterey pine, SMITH and DEVEY, 1994; eastern white pine, ECHT et al., 1996; Sitka spruce, VAN DE VEN and MCNICOL, 1996; Norway spruce, PFEIFFER et al., 1997) and *Fagaceae* (bur oak, DOW et al.. 1995 and cork oak, GÓMEZ et al., 2001). Aspen forests are often made up of adjacent stands of a

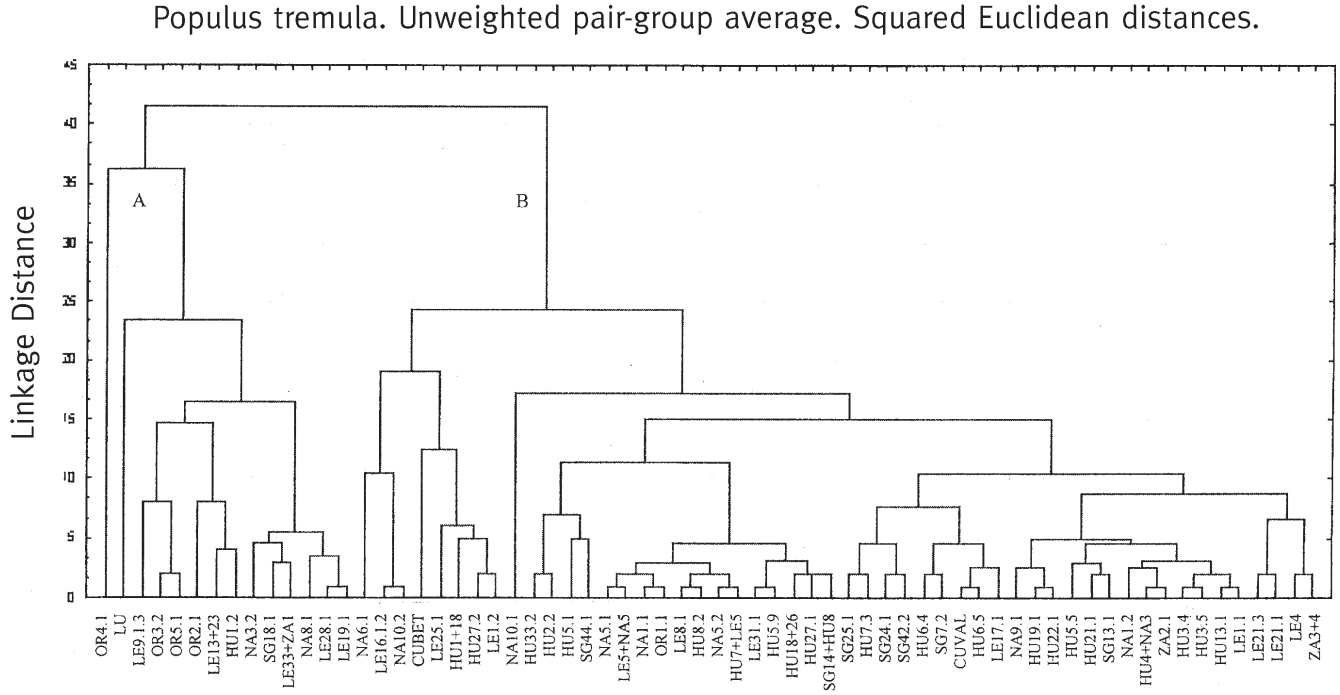

Figure 2. – *Populus tremula* "in vitro" Spanish collection UPGMA dendrogram by SSR markers.

Table 4. – Groups of clones at 5, 10 and 15 linkage distances from the UPGMA phenogram of *P. tremula* genotypes detected by SSR markers.

Clones		
Linkage distance		
15	10	5
LU	LU	LU
OR3.2	OR3.2	LE9.1
LE13	LE13	OR3.2
LE33	LE33	OR2.1
LE16	NA6.1	LE13
HU1	LE16	LE33
NA10.1	CUBET	NA8.1
NA5.1	LE25.1	NA6.1
HU4	HU1	LE16
	NA10.1	CUBET
	HU7	LE25.1
	NA5.1	HU1
	SG7.2	HU27.2
	HU4	NA10.1
		HU2.2
		HU5.1
		SG44.1
		SG24.1
		CUVAL
		NA9.1
		HU7
		NA5.1
		SG7.2
		HU4
		LE21.3
		ZA3
Number 9	14	26

dozen or more trees each, which are genetically identical within each stand but may vary greatly between stands (EINSPAHR and WINTON, 1977). The importance of vegetative propagation by root suckers in aspen over seed propagation is a reason for low variability.

The discrepancy in the estimates between *P. tremuloides* and *P. tremula* is very important considering that core subset must be designed to maximise allelic diversity. We assume that in this case the difference is not caused by sample size (36 individuals and 24 alleles in *P. tremuloides* DAYANANDAN et al. (1998) versus 82 individuals and 21 alleles in *P. tremula*). GUILFORD et al. (1997) and HOKANSON et al. (1998) assumed that the difference in the number of alleles found in apple is caused by sample size, but this is not the case in aspen. It is well known that North American plant species are more abundant and show a higher degree of variability than their European relatives (TAKHTAJAN, 1986). This fact has been explained by mountain range orientation. In North America, the North-South-oriented mountains facilitated migrations during the glaciations, while in Europe, the East-West-oriented mountains would have formed a barrier that caused population extinction. Furthermore, it has been suggested that the very widespread *P. tremuloides* has a limited geographic variability throughout its range when compared to most species. A possible cause is that seedling reproduction may be the exception at the present time, but the seed origin of most quaking aspen clonal stands today may date back to the close of the ninetieth century, when devastating fires and other catastrophes swept many areas of North America. Natural direct seeding occurred from survivors after the fires, followed by vegetative sucker propagation (EINSPAHR and WINTON, 1977). There are not enough data from truly in-depth provenance studies to support or refute the statement that this species has limited racial development (ZOBEL and TALBERT, 1988). Nevertheless, our results suggest that, in addition to this low variability in aspen, the glaciation bottle-neck has further reduced the genetic variability of the Spanish *P. tremula* populations.

The heterozygosities for individual *loci* ranged from 0.115 to 0.659 for the six *loci* in this study with a mean value for all loci of 0.475 (*Table 2*), quite similar to the 0.498 value reported by DAYANANDAN et al. (1998) and RAHMAN et al. (2000).

The combined discrimination power for all six loci was 0.992, with the probability of matching any two genotypes at all six loci in this study being 8 in one thousand. We were not able to differentiate genotypes No. 1, 7, 15, 29, 34, 61, 62 and 64, that belong to trees from the same geographical area. If we consider SSR markers robust enough in this respect (HOKANSON et al.,

1998), trees from the same stand and with the same genotype like both trees with genotype No. 62 and the two with genotype No. 64 could be ramets of a single clone propagated by root suckers that have been extended through the stand (HEYBROEK, 1984). Also, the low P_{ID} obtained ($2.7 * 10^{-4}$) suggests the clonal origin of equal multilocus genotypes, in agreement with the results obtained by SCHUELER et al. (2003) in wild cherry.

The UPGMA cluster analysis produced several groupings based on geographical origins. Samples from Huesca and Navarra may be found in all groups of the phenogram. The group 1 includes samples from the Northwest of the Iberian Peninsula whereas group 2 has two smaller groups, called A and B, formed by samples from Segovia and from León-Zamora respectively. These results agree with the hypothesis that post-glacial migration of *P. tremula* took place from the Pyrenean nucleus of Huesca-Navarra in two directions: the Central and Northwest mountains of the Iberian Peninsula. Our hypothesis is based on the study of hipervariable *loci* such as SSRs, with high mutation rates, present in populations separated by relatively large geographical distances that show significant effects of mutation (RAYBOULD et al., 1997), explaining the separation of groups 1 and 2. These results agree with previous studies based on RAPD markers that suggest that Northwestern populations have maintained a genetic contact exclusively with the Pyrenean populations through the Cantabrian mountain-range, and has been isolated from the Central populations by the central plateau during the warmer post-glacial period (SÁNCHEZ et al., 2000).

These results have implications for the genetic conservation of *P. tremula.* Several criteria needed to be taken into account to assist the selection of clones that should be conserved *in vitro.* A set of trees could be chosen so that all the SSR alleles were represented at least once, with division at linkage distance 5. Following this criterion, the number of alleles conserved is the total number of alleles present in the base collection allelic richness. This would help to maximise the genetic diversity available. Allelic richness is considered the most relevant criterion when studying population diversity (KREMER, 1994; EL MOUSADIK and PETIT, 1996; PETIT et al., 1998). Some authors considered that allele preservation is more important than the maintenance of allelic frequencies. Thus, although frequencies of group at distance 5 (*Fig. 2*) are different from those at the total collection, we consider more important that all alleles are represented in this core collection. At distance 5, individuals from different origins are present, thus coming from different ecological areas. Several trees must be added for aesthetic reasons (for example, some *P. tremula* trees show red coloured leaves in Autumn).

In general, allelic frequency was low and also was unbalanced in all *loci*, a certain allele being much more abundant than all other alleles of the *locus*. Because of their abundance, alleles with the highest frequency at each locus are unlikely to be lost during the creation of the core collection, regardless of which sampling method is used, whereas alleles with low frequency are expected to be the most vulnerable to loss (MARSHALL and BROWN, 1975).

Localised high-frequency and widespread, low-frequency alleles are more interesting in germplasm collections than localised low-frequency alleles (BROWN, 1978), because they are likely to be maintained by deleterious mutation-selection balance and are therefore of little interest in genetic conservation (MARSHALL and BROWN, 1975). In this way, the M strategy led to a more effective capture of interesting alleles (BATAILLON et al., 1996). From this point of view, the group of clones at distance 5 includes the more interesting alleles (*figure 2*) and, therefore, no advantages are found in conserving the total collection.

The *in vitro* core collection of *Populus tremula* from the present collection of Spanish origins could be composed of 26 clones, combining storage economy with allelic richness detected by SSR markers. To this core of 26 clones, other trees could be added for aesthetic reasons, ecological conditions of locations and studies based in more SSRs or/and other kinds of markers.

Acknowledgements

GÓMEZ was a recipient of a grant from the National Institute of Agronomic Research, INIA (Spain). AGUIRIANO had a practice contract during this work. We also thank Dr. N. ALBA and F. LALINDE for sample collection. This study was funded by INIA SC98-081 research project.

References

BATAILLON, T. M., DAVID, J. L. and SCHOEN, D. J.: Neutral genetic markers and conservation genetics: simulated germplasm collections. Genetics **144**: 409–417 (1996). — BROWN, A. H. D.: Enzyme polymorphisms in plant populations. Theor. Pop. Biol. **15**: 1–42 (1979). — BROWN, S. M., HOPKINS, M. S., MITCHELL, S. E., SENIOR, M. L., WANG, T. Y., DUNCAN, R. R., GONZÁLEZ-CANDELAS, F. and KRESOVICH, S.: Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in shorgum (*Shorgum bicolor* (L.) Moench). Theor. Appl. Genet. **93**: 190–198 (1996). — BUENO, M. A., MANZANERA, J. A., GRAU, J. M., SANCHEZ, N. and GÓMEZ, A.: Propagación *in vitro de Populus tremula y Populus alba.* Pages 53–81 in Populicultura en Castilla y Leon. Junta de Castilla y Leon. S-599-2001 (2001). — DAYANANDAN, S., RAJORA, O. P. and BAWA: Isolation and characterization of microsatellites in Trembling aspen (*Populus tremuloides*). Theor. Appl. Genet. **96**: 950–956 (1998). — DOYLE, J. J. and DOYLE, J. L.: Isolation of plant DNA from fresh tissue Focus **12**: 13–15 (1990). — DOW, B. D., ASHLEY, M. V. and HOWE, H. F.: Characterization of highly variable (GA7CT)n microsatellites in the bur oak (*Quercus macrocarpa*) Theor. Appl. Genet. **91**: 137–141 (1995). — ECHT, C. S., MAY-MARQUARDT, P., HSEIH, M. and ZAHORCHAK, R.: Characterization of microsatellite markers in eastern white pine. Genome **39**: 1102–1108 (1996). — EL MOUSADIK, A. and PETIT, R. J.: High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. Theor. Appl. Genet. **92**: 832–839 (1996). — EINSPAHR, D. W. and WINTON, L. L.: Genetics of quaking aspen. Pages 1–23 Forest Service USDA, US Govt. Printing Office, Washington (1977). — FIELDS, R. L. and SCRIBNER, K. T.: Isolation and characterisation of novel waterfowl microsatellite loci: cross-species comparisons and research application. Molecular Ecology **6**: 199–202 (1997). — GÓMEZ, A., PINTOS, B., AGUIRIANO, E., MANZANERA, J. A. and BUENO, M. A.: SSR markers for *Quercus suber* tree identification and embryo analysis. Journal of Heredity **92**: 292–295 (2001). — GUILFORD, P., PRAKASH, S., ZHU, J. M., RIKKERINK, E., GARDINER, S., BASSETT, H. and FORSTER, R.: Microsatellites in *Malus x domestica* (apple): abundance, polymorphism and cultivar identification. Theor. Appl. Genet. **94**: 249–254 (1997). — HEYBROEK, H. M.: Clones in forestry and in nature. Arboricultural Journal **8**: 275–286 (1984). — HOKANSON, S. C., SZEWC-MCFADDEN, A. K., LAMBOY, W. F. and MCFERSON, J. R.: Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. Theor. Appl. Genet. **97**: 671–683 (1998). — HUNTLEY, B. and BIRKS, H. J. B.: An atlas of past and present pollen maps for Europe, 0-13.000 Years ago. Cambridge University Press, Cambridge U.K. (1983). — KLOOSTERMAN, A. D., BUDOWLE, B. and DASELAAR, P.: PCR-amplification and detection of the human D1S80 VNTR *locus.* Amplification conditions, population genetics and application in forensic analysis. Legal Medicine **105**: 257–264 (1993). — KREMER, A.: Diversité génétique et variabilité de caractères phénotypiques chez les arbres forestiers, Genetics, Selection, Evolution, **26**: 105s-123s (1994). — LAMBOY, W. F. and ALPHA, C. G.: The utility of simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. J. Am. Soc. Hortic. Sci. **123**: 182–188 (1988). — LAVI, U., AKKAYA, M., BHAGWAT, A., LAHAV, E. and CREGAN, P. B.: Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). Euphytica **80**: 171–177 (1994). — MARSHALL, D. R. and BROWN, A. H. D.: Optimum sampling strategies in genetic conservation. Pages 53–80. *In:* Crop Genetics Resources for Today and Tomorrow. J. H. FRANKEL and J. G. HAWKES, eds. Cambridge University Press, London (1975). — MARSHALL, D.: Crop genetic resources: current and emerging issues. Pages 367–38. *In:* BROUN, A. H. D., CLEGG, M. H., KAHLER, A. L., WEIR, B. S. (eds). Plant population genetics, breeding, and genetic resources. John Wiley and Sons, Chichester, UK (1990). — NEI, M.: Analysis of gene

diversity in subdivided populations. Proc. Natl. Academy of Science U.S.A. **70**: 3321–3323 (1973). — PETIT, R. J., EL MOUSADIK, A. and PONS, O.: Identifying populations for conservation on the basis of genetic markers. Conservation Biology **4**: 844–855 (1998). — PFEIFFER, A., OLIVIERI, A. M. and MORGANTE, M.: Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). Genome **40**: 411–419 (1997). — PRIMMER, C. R., MOLLER, A. P. and ELLEGREN, H.: Polymorphisms revealed by simple sequence repeats. Trends in Plant Science **1**: 215–222 (1996). — RAHMAN, M. H., DAYANANDAN, S. and RAJORA, O.P.: Microsatellite DNA markers in *Populus tremuloides.* Genome **43**: 293–297 (2000). — RAYBOULD, A. F., MOGG, R. J., ALDAM, C., GLIDDON, C. J., THORPE, R. S. and CLARKE, R. T.: The genetic structure of sea beet (*Beta vulgaris ssp. maritima*) populations. III. Detection of isolation by distance at microsatellite loci. Heredity **80**: 127–132 (1997). — RHOLF, F. J.: NTSYS-pc, version 1.8. Exeter Software, Setauket, New York (1994). — RONGWEN, J., AKKAYA, M. S., BHAGWAT, A. A., LAVI, U. and CREGAN, P. B.: The use of microsatellite DNA makers for soybean genotype identification. Theor. Appl. Genet. **90**: 43–48 (1995). — RUÍZ DE LA TORRE, J.: Árboles y arbustos de la España Peninsular, ETSI Montes, 512 pp (1979). — RUSSELL, J., FULLER, J., YOUNG, G., THOMAS, B., TAMARINO, G., MACAULAY, M., WAUGH, R. and POWELL, W.: Discriminating between barley genotypes using microsatellite markers. Genome **40**: 442–450 (1997). — SÁNCHEZ, N., GRAU, J. M., ALBA, N., MANZANERA, J. A. and BUENO, M. A.: Genetic characterisation of *Populus tremula* Regions of Origin in Spain using RAPD fingerprints. Silvae Genetica **49**: 66–71 (2000). — SCHUELER, S., TUSCH, A., SCHUSTER, M. and ZIEGENHAGEN, B.: Characterization of microsatellites in wild and sweet cherry (*Prunus avium* L.) – markers for individual identification and reproductive processes. Genome **46**: 95–102 (2003). — SMITH, D.N. and DEVEY, M. E.: Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome **37**: 977–983 (1994). — SUN, H. S. and KIRKPATRICK, B. W.: Exploiting dinucleotide microsatellites conserved among mammalian species. Mammalian Genome **7**: 128–132 (1996). — TAKHTAJAN, A.: Floristic regions of the world, Univ. Calif. Press, Berkeley (1986). — THORPE, R. and MALHOTRA, A.: Molecular and morphological evolution within small islands. Phil. Trans. R. Soc. B. **351**: 815–822 (1996). — VAN DE VEN, W. T. G. and MCNICOL R. J.: Microsatellites as DNA markers in Sitka spruce. Theor. Appl. Genet. **93**: 613–617 (1996). — WAITS, L.P., LUIKART, G. and TABERLET, P.: Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. **10**: 249–256 (2001). — WEBER, J. L. and MAY, P. E.: Abundant class of human DNA polymorphysms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44**: 388–396 (1989). — WEIR, B. S.: Genetic Data Analysis. Sinauer Associates, Sunderland, M.A. (1990). — ZOBEL, B. and TALBERT, J.: Applied Forest Tree Improvement, John Wiley and Sons, New York, 505 pp (1988).

Forest Tree Transgenesis and Functional Genomics: From Fast Forward to Reverse Genetics

By S. KUMAR* and M. FLADUNG

BFH Institute for Forest Genetics and Forest Tree Breeding, Sieker Land Str. 2, D-22927 Grosshansdorf, Germany, (mfladung@uni-hamburg.de)

(Received 3rd June 2003)

Summary

Genomics has become an integral part of the forest tree improvement, and gene structural and expressional data are being produced at an unprecedented rate. However, biological resources in the form of tagged mutants are still lacking in forest trees, which at present is a missing part of tree genomics. The potential bottlenecks here are the steps involving plant transformation, which is instrumental both in reverse and forward genetics strategies aimed at to determine gene function. With few exceptions, genetic transformation is an obligatory final step by which traits are engineered into plants. For basic research transgenesis is the method of choice to confirm gene function, after deductions made through comparative genomics, expression profiles, and mutation analysis. The biological features of long-lived tree species create obstacles as well as provide opportunities to design new approaches to overcome the barriers associated with forest tree genomics.

To understand how a cell works we need to know the function of almost every gene in its genome. Genome sequencing provides a tremendous amount of information for the development of global approaches towards this goal, complementing and enhancing the more traditional (single-gene) approaches. Genome sequencing has been completed in model plant *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000) and in rice (GOFF et al., 2002; YU et al., 2002), a model cereal. Similarly, forest biologists have given enough justification to sequence *Populus* genome as a model for trees and woody perennials (WULLSCHLEGER et al., 2002; TAYLOR, 2002). A 6X coverage of the black cottonwood (*Populus trichocarpa*) genome as first tree genome will be available in public domain by the end of the year 2003 (International *Populus* Genome Consortium). Once whole-genome information is available for an organism, the challenge turns from identifying the parts to understanding their function as well as to improving genome structure, thus ushering in the 'post-genomic' era. In the short term, the first goal is to assign some element of function to each of the genes in an organism also referred to as 'functional genomics', and to do this with high-throughput, systematic approaches.

Major challenges of functional genomics in trees are to assess tree growth and wood yield. These parameters are important in terms of wood quality like strength and fibre length, and renewable energy resources (CHAFFEY et al., 2002; CAMPBELL et al., 2003; CONFALONIERI et al., 2003). With most of the *Populus* genome still to be assigned function, the notion of accumulating this information one gene at a time is hard to contemplate. This knowledge gap has been the crucial impetus for developing 'whole-genome' approaches that can acquire functional information, in the form of expression profiles, protein–protein interactions, computational approaches and the response to loss or gain of function by mutation.

Key words: Forest trees, genetic transformation, genomics, gene tagging, transposon, gene silencing, site-specific recombination.

Functional genomics of forest trees: A missing link?

Forest tree genomics is now undergoing a transition or expansion from the mapping and sequencing of genomes to an

* Present address: Department of Botany, North Carolina State University, Raleigh, NC 27695-7612. Phone: 919-515-7166. Fax: 919-515-3436. E.mail: skumar3@unity.ncsu.edu