Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species

Mauro Bardini, David Lee, Paolo Donini, Anna Mariani, Silvia Gianì, Marcello Toschi, Chris Lowe, and Diego Breviario

Abstract: TBP (tubulin-based polymorphism) is a new molecular marker based tool that relies on the presence of intron-specific DNA polymorphisms of the plant β-tubulin gene family. The multifunctional and essential role of the tubulin proteins is reflected in the conservation of regions within their primary amino acid sequence. The ubiquitous nature of this gene family can be exploited using primers that amplify the first intron of different β-tubulin isotypes, revealing specific fingerprints. The method is rapid, simple, and reliable and does not require preliminary sequence information of the plant genome of interest. The ability of TBP to discriminate between accessions and species in oilseed rape, coffee, and lotus is shown. In all cases, TBP was able to detect specific genetic polymorphisms in the context of a simplified and readily appreciable pattern of DNA amplification. The application of TBP for assessing genetic diversity and genome origins in disseminated plant landraces rather than in highly inbred cultivated species is also discussed.

Key words: β-tubulin, *Brassica napus*, *Coffea*, *Lotus*, SSRs.

Résumé : Le TBP (« tubulin-based polymorphism ») est un nouveau système de marquage moléculaire basé sur la présence de polymorphismes d'ADN spécifique des introns de la famille de gènes de la β-tubuline des plantes. Ces gènes codent pour des protéines de la tubuline dont le rôle multifonctionnel et essentiel est témoigné par la conservation de régions dans leur séquence primaire des acides aminés. L'ubiquité de cette famille de gènes peut être exploitée en utilizant des amorces qui amplifient le premier intron des divers isotypes de β-tubuline, faisant ainsi apparaître des empreintes spécifiques. La méthode est rapide, simple et fiable et n'exige pas d'information préliminaire sur la séquence du génome de la plante d'intérêt. La capacité du TBP de différencier les accessions et les espèces chez le colza, le café et le lotus est démontrée. Dans chaque cas, le TBP a été en mesure de détecter des polymorphismes génétiques spécifiques dans le contexte d'un modèle simlifié, et de lecture immédiate, d'amplification de l'ADN. Le travail traite également de l'application du TBP à la recherche de la diversité génétique et des origines du génome dans un ensemble très épars de races locales plutôt que chez des espèces cultvées fortement fixées.

Mots clés : β-tubuline, *Brassica napus*, *Coffea*, *Lotus*, SSRs.

Introduction

Assessment of genetic diversity is important for the study of biodiversity, population dynamics, and ecological relationships. Knowledge of available plant genetic resources is fundamental to support programs for the development of new cultivars, as well as to protect existing natural resources. Molecular markers are now the tools most widely used to assess genetic diversity (Koebner et al. 2001; Karp et al. 1998). Accordingly, they are used to characterize the genetic base of a given plant species and as tools to complement programs of breeding and biotechnology. Most of the current methods to determine genetic diversity are based on polymerase chain reaction (PCR), since this technique is simple and powerful (Powell et al. 1995). Concomitantly, most of the molecular markers that are used to assess ge-

Received 29 May 2003. Accepted 11 November 2003. Published on the NRC Research Press Web site at http://genome.nrc.ca on 5 March 2004.

Corresponding Editor: G.J. Scoles.

M. Bardini. Università degli Studi di Milano, Dipartimento di Biologia, Via Celoria 26, 20133 Milano, Italy.
D. Lee, P. Donini, and C. Lowe. National Institute of Agricultural Botany, Huntingdon Road, Cambridge, CB3 0LE, U.K.
A. Mariani.[1] Istituto di Ricerche sul Miglioramento Genetico delle Piante Foraggere, Via della Madonna Alta 130, 06128 Peurgia, Italy.
S. Gianì, M. Toschi, and D. Breviario.[2] Istituto Biologia e Biotecnologia Agraria–CNR, via Bassini 15, 20133 Milano, Italy.

[1]Present address: Instituo Genetica Vegetale – CNR, Sezione Perugia, Via della Madonna Alta 130, 06128 Perugia, Italy.
[2]Corresponding author:(e-mail: breviario@ibba.cnr.it).

 doi: 10.1139/G03-132

netic diversity correspond to anonymous DNA sequences that, although efficient in providing a vast range of natural sources for DNA polymorphisms, almost exclusively associate within regions of unknown function. In fact, most of the identified changes do not occur in functionally relevant regions of DNA and thus rarely lead to identification of variations occurring within genes.

In this regard, techniques based on potentially variable regions of specific genes or gene families permit the assessment of the genetic variation within DNA sequences with known function. Several examples of this have now been reported in many organisms, e.g., the *ADH* loci in cotton (Small et al. 1999) and grasses (Gaut et al. 1999) and the *GS* multigene family in *Brassica napus* L. (Ochs et al. 1999). DNA polymorphisms are frequently found outside the coding region of a gene, as shown for the regulatory region of *teosinte branched 1*, a gene that was important during maize domestication (Wang et al. 1999). In several cases, intergenic spacer regions have been used as the target sequences (Sallares and Brown 1999; Soranzo et al. 1999). Analysis of the intergenic spacer regions of nuclear ribosomal repeats in *Pyrenophora graminea* has revealed a much greater rate of evolution than that previously reported for the internal transcribed region of the rDNA (Pecchia et al. 1998). The sequences of the *NADH*, *nod*31, and *rps*12 genes have been used to characterize genetic variation in conifers (Soranzo et al. 1999). Intron sequences, which may be considered "intragenic" spacer regions, are also used to characterize varieties as for the *rbcS* gene in Triticae (Sasanuma and Miyashita 1998). In this respect, the family of plant β-tubulin genes may be regarded as a valuable functional DNA target amenable for diversity assessment.

The β-tubulin polypeptide, together with the α-tubulin chain, is the major constituent of microtubules, essential intracellular structures involved in fundamental mechanisms such as cell division, vesicular transport, cell wall deposition, signal propagation, and others (Nick 1998; Nogales 2000). The importance of this protein in cellular structures required for cell viability and its related multifunctional role implies a rather strict maintenance of much of the primary amino acid sequence. In accordance, plants contain a small number of β-tubulin genes encoding for slightly different proteins defined as the tubulin isotypes. The plant β-tubulin gene families of *Arabidopsis thaliana* and rice have been among the best characterized gene families with nine and eight members, respectively (Snustad et al. 1992; Gianì and Breviario 1996). The genomic organization of plant β-tubulin genes is typically conserved with two introns located at fixed positions within the coding region, with the exception of maize β-tubulin genes that contain just the first intron (Liaud et al. 1992).

Here, we report the development of a new technique to investigate genetic diversity that is based on the length of the first intron present in the coding region of plant β-tubulin genes. This technique, called TBP (tubulin-based polymorphism), could be particularly suitable for ecological studies where there is often a lack of genomic information for the plant species of interest. The rationale behind the TBP method is based on three simple facts.

The first is that all plant β-tubulin genomic DNA characterized so far reveals the presence of a first intron within the coding sequence that is located 396 nucleotides from the A residue of the ATG translation initiation codon. The second is that the length of this β-tubulin intron can vary between the different isotypes that define the β-tubulin gene family. The third is that the intron is flanked on both sides by coding nucleotide sequences that are fairly well conserved and allow the design of primers suitable for PCR amplification of the intervening sequence (Fig. 1). Given these facts, the first intron of plant β-tubulin may be the source for DNA polymorphisms defining a new molecular marker for genetic diversity. All of these assumptions have been preliminarily verified in rice, where TBP was originally developed (European Pending Patent No. 99963667.3). Primers designed on the boundaries of the β-tubulin intron were able to amplify bands of different sizes within any given rice cultivar. Some of these bands were isolated from the gel and sequenced to show that they were indeed DNA sequences corresponding to introns of different rice β-tubulin isotypes (European Bioinformatics Institute accession Nos. AJ243273 and AJ242868). In any case, the number of amplified bands never exceeded the reported number of rice β-tubulin alleles (Gianì and Breviario 1996; Gianì et al. 2003).

Although some polymorphic bands could be detected comparing the banding patterns obtained from different cultivars, rice is not the best plant to show the applicability of the TBP method because commercially available rice varieties are highly inbred and would require many accessions for meaningful statistical analysis. This is the reason why, with respect to statistics, we tested the reliability of the TBP method on oilseed rape.

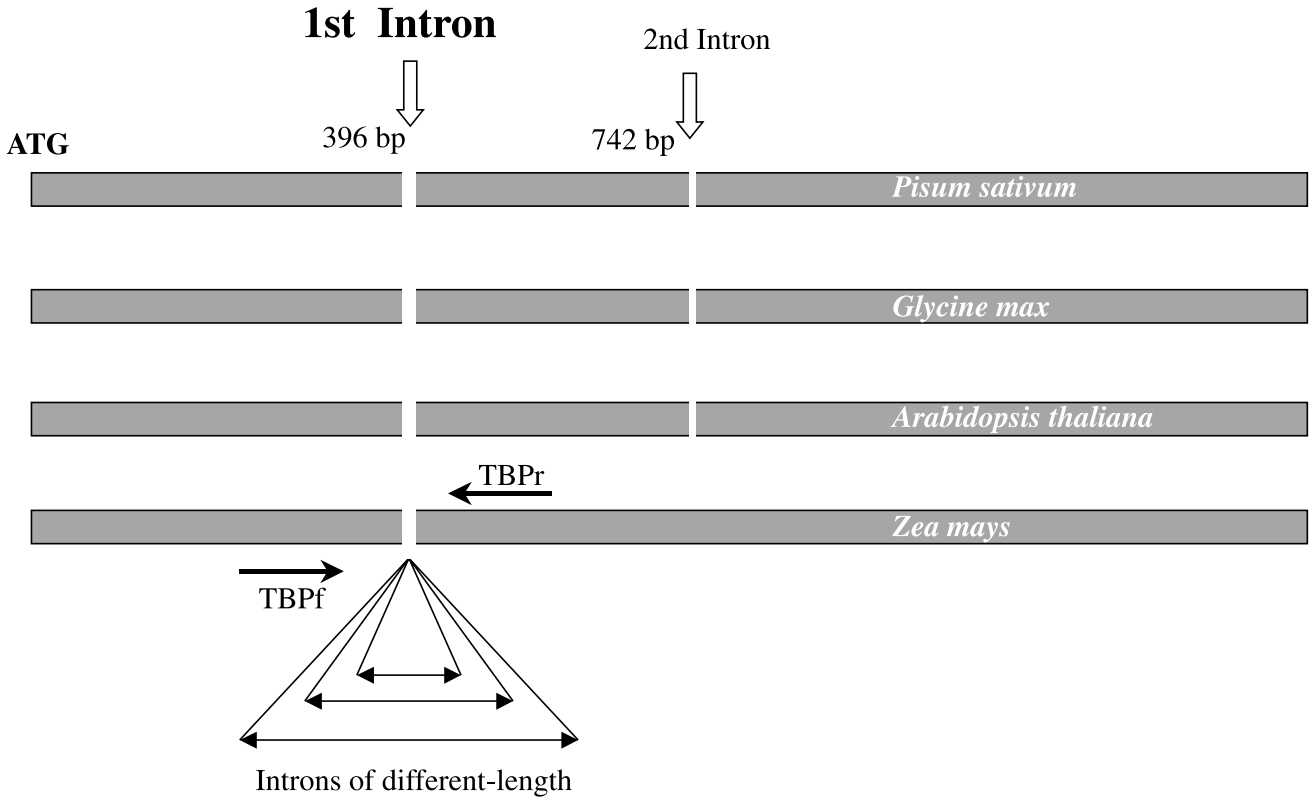
In addition, data on the use of TBP for fingerprinting different cultivars of coffee and lotus are also presented. Oilseed rape has been used to compare results and related statistical information gained from TBP and simple sequence repeat markers (SSRs) (Morgante and Olivieri 1993). Coffee has been used to identify the origin of cultivated *Coffea arabica* L. varieties and to assess the occurrence of polymorphisms within a plant species with a narrow genetic base (Lashermes et al. 1999). Ultimately, we applied TBP to study different species of *Lotus* in an attempt to elucidate the origin of the *Lotus corniculatus* L. complex, a still much debated issue (Roos and Jones 1985).

Materials and methods

Plant material

Oilseed rape (*B. napus*) samples were from the National Institute of Agricultural Botany U.K. collection (cultivars Bienvenue, Cobol, Express, Falcon, Hobson, Global, Marinka, Lirawell, and Nimbus). *Lotus* samples were obtained from the Perugia Consiglio Nazionale delle Ricerche collection (*L. corniculatus*, *Lotus alpinus* (DC) Schleich, *Lotus tenuis* Waldst et Kit, *Lotus pedunculatus* Cav., and *Lotus angustissimus* L.). *Coffea arabica* and *Coffea canephora* Pierre ex Froehner samples were provided by Kraft Foods and *Coffea eugenoides* was a gift from Dr. David Jaccoud Filho (UEPG, Ponta Grossa, Brazil). Seeds of oilseed rape and lotus were germinated in vitro and seedlings grown in either a greenhouse or a growth chamber until the fourth true-leaf stage. For oilseed rape, DNA was extracted from 44 or 45 individual plants of each cultivar. Fifty to 100 mg of

Fig. 1. Scheme of the TBP method. The bar represents a generic β-tubulin sequence with reference to some of the best characterized plant gene families. First introns are located at nucleotide position +396 from the ATG codon. TBPf and TBPr define the position of the forward and reverse primers relative to the first intron sequence. The position of a second intron, present in the vast majority of the plant β-tubulin genes but not maize, is also shown.



leaf discs from each plant were freeze-dried and DNA was extracted using a mixer mill (Qiagen, Crawley, U.K.) in conjunction with the DNeasy Plant extraction kit (Qiagen) following the manufacturer's instructions. For lotus, DNA was extracted from leaf discs of 15–20 individual plants depending on the different accession. For coffee, DNA was extracted from beans of 10 individuals for each cultivar. In each case, quality and quantity of DNA were assessed by agarose gel electrophoresis. Routinely, DNA extractions that resulted in poor quality or quantity were repeated. The final DNA concentration ranged from 10 to 100 ng/μL. Fifty nanograms of DNA was typically used in any single PCR.

Primer design and PCR amplification

The development of the microsatellite markers for *B. napus* has been described elsewhere (Tommasini et al. 2003). The position of the first intron of the plant β-tubulin gene is conserved and located after 132 amino acid codons in all of the plant species analyzed. The alignment of the DNA sequences using CLUSTAL W software (Thompson et al. 1994) showed conserved regions around the intron splice site. Several primers flanking the splice junction with different degrees of degeneration were tested for their ability to amplify a clear banding pattern. The best pair of primers was selected and used for all of the experiments discussed in this paper (TBPF1: 5′-GARGCYGARAAYTGYGAYTG-3′; TBPR1: 5′-RTCHGGRTAYTCYTCHCKRAT-3′).

PCRs (20 μL) were performed in 1× PCR buffer (10 mM Tris–HCl (pH 8.8), 50 mM KCl, and 0.1% v/v Triton X-100), 0.2 mM each dNTP, 2.5 mM $MgCl_2$, 50 ng of template DNA, 1 μM each primer, and 1 U *Taq* polymerase (Biogene, Englewood, N.J.) in a GeneAmp PCR System 9700 thermal cycler (PE Biosystems, Foster City, Calif.). Following the initial denaturation step at 94 °C for 3 min, the PCR consisted of 35 cycles of 94 °C for 30 s, 55 °C for 40 s, and 72 °C for 90 s. The reactions were held at 15 °C after a final extension at 72 °C for 8 min. The forward primer had been labelled with IRD700 dye (MWG-Biotech, Eberstburg bei München, Germany). These PCR conditions were the best that we found after testing different programs. After PCR amplification, the samples were denatured and loaded on a 6% w/v denaturing polyacrylamide gel in a LI-COR detection system, as described in Jackson and Matthews (2000).

Consistency of PCR-amplified products was checked by performing all reactions at least twice. Bands that were reproducible in successive amplifications were selected and marker bands were defined by their molecular sizes, estimated from mass standards.

Data collection and analysis

Bands generated by SSRs and by TBP were scored as either present (1) or absent (0) across all accessions to build separate binary data matrices. Using the binary data, we performed principal coordinates analysis (PCA) analysis and AMOVA (Peakall and Smouse 2001) to compare the ability of the two molecular marker techniques to distinguish among the six cultivars of oilseed rape. PCA was performed

using the DCENTER and EIGEN procedures in NTSYS-pc (Rohlf 1993).

Then, the relative frequency of each molecular marker was calculated for each cultivar to perform a correlation analysis between cultivars using the TBP and SSR data. Genetic similarity between cultivars was estimated using the formula for Euclidean distance:

$$\sum_{i=1}^{n} \frac{\sqrt{(X_i - Y_i)^2}}{n}$$

where n is the number of polymorphic markers generated by each technique, X_i is the value of the i-sim marker for the X cultivar, and Y_i the value of the i-sim marker for the Y cultivar. A Mantel test was performed using NTSYS to measure correspondence between sets of similarity data from SSR and TBP.

The mean polymorphism information content value for TBP markers was calculated using the following formula:

$$\text{PIC} = \sum_{i=1}^{n} \frac{(1 - f_{ai}^2 - f_{pi}^2)}{n}$$

where n is the total number of the scored polymorphic TBP markers, f_{ai} is the frequency of lines in which the ith fragment was absent (frequency of the null allele) and f_{pi} is the frequency of the lines in which the ith fragment was present (frequency of the amplified allele; Hongtrakul et al. 1997).

Results

TBP on oilseed rape

Figure 2 shows images of PCR-amplified products obtained from individual oilseed rape plants with either the SSR or the TBP method. Figure 2A shows the detail of the LI-COR gel image reporting products amplified within the 106- to 275-bp range with the use of SSR markers on the Global cultivar. Similar analyses have been performed on the other five different oilseed rape cultivars studied in this paper (data not shown). Figure 2B shows the detail (205- to 275-bp interval) of a representative LI-COR gel image of the PCR-amplified products obtained with TBP from DNA extracted from individual oilseed rape plants representing all six different cultivars. As shown, a strict association between specific banding pattern and individuals of the same cultivar was found. Both the SSR and the TBP methods revealed the presence of intra- and intercultivar polymorphisms, e.g., TBP was able to score 13 polymorphic bands (six are shown in Fig. 2B) exploiting just one PCR reaction, and SSRs reactions were perfomed on the six cultivars using 13 primer pairs, multiplexed in three reactions, producing 59 different alleles.

A total of 266 samples corresponding to individuals from six different oilseed rape cultivars (Falcon, Hobson, Global, Lirawell, Marinka, and Nimbus) were fingerprinted with the TBP method. This allowed a comparison with SSR data (Tommasini et al. 2003). Bands obtained with the TBP and SSR approaches were scored as present or absent from each accession and data referring to polymorphic markers were subjected to PCA. Samples were plotted by their first two principal coordinates, which accounted for 51.5% and 46.6% (SSRs and TBP, respectively) of the total variance (Fig. 3). As shown, individuals of the six different cultivars group in substantially similar clusters, although a tighter association among individuals of the same cultivar can be observed with the SSR analysis. This is easily seen for individuals of the Hobson and Falcon varieties, all alike with TBP and closely clustered with SSR (Fig. 3). This varietal separation is due to a considerably higher number of markers that originate by SSR compared with TBP. Thus, PCA indicates that the TBP method provides information on genetic similarity that is, to the first approximation, similar to that obtained using SSR markers with the difference that the results obtained with TBP come from a single PCR.

Next, we analyzed the data with AMOVA to verify to which extent the two methods were able to discriminate among and within the oilseed rape cultivars. The TBP method was capable of detecting 66% of the variance among varieties and 34% of the variance within varieties. These values are not that different from those obtained with SSR, 80% and 20%, respectively, considering that TBP relies on less than a quarter of the marker information detected by SSR analysis. This further confirms the reliability of the TBP method. The mean polymorphic information content score calculated for the TBP markers was 0.33, a high value considering 0.5 as the maximum polymorphism information content score for any biallelic marker.

Ultimately, we have built a new set of data where, for each cultivar, the relative frequency of every single marker has been calculated and measured for all individuals of the same cultivar. These data have been used to obtain two new matrices (one for TBP and one for SSR) where the six varieties are defined by the relative frequency of any given marker.

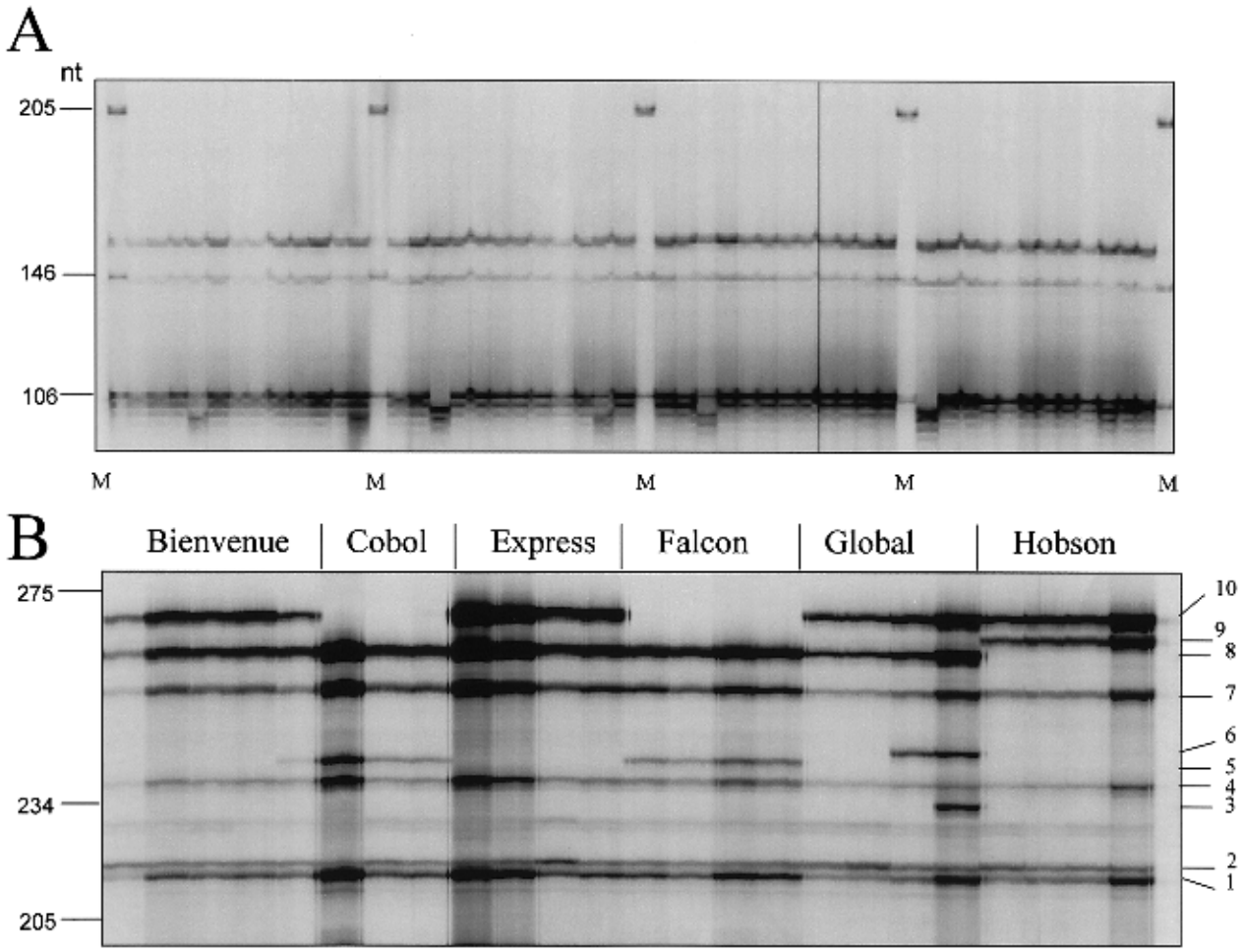
Using Euclidean distance, we built two different similarity matrices. Finally, to compare the ability of the two methods in defining similarity between cultivars, we performed a Mantel test and plotted the results in a two Cartesian axis graph reporting TBP and SSR similarity values for each comparison (Fig. 4). In the graph, the dashed line represents the ideal situation where the correlation value between the two sets of data is 1. A correlation value of 0.65 ($P > 0.001$) was obtained from the comparison of pairwise genetic distance estimates based on TBP and SSR data.

TBP on coffee

TBP was applied to different domesticated South American accessions of *C. arabica* to verify the tecnique's suitability to detect polymorphisms in a plant species known to have a narrow genetic base. *Coffea arabica* ($2n = 4x = 44$) is an amphidiploid putatively derived from hybridization between the diploid ($2n = 2x = 22$) species *C. eugenoides* and *C. canephora*. Molecular evidence for this came from restriction fragment length polymorphism and genomic in situ hybridization analyses (Lashermes et al. 1999). *Coffea arabica* is the most commercially valuable species, accounting for about 70% of world production. The remainder is mainly *C. canephora*.

As shown in Fig. 5, the use of TBP on coffee DNA revealed intraspecific polymorphisms in *C. canephora*, while no polymorphism was detected among the *C. arabica* samples. The overall pattern of DNA amplification is simple, largely restricted to four bands for each individual, two of

Fig. 2. TBP on oilseed rape. Details of representative denaturing polyacrylamide gel electrophoresis of PCR products amplified from genomic oilseed rape DNA with (A) SSRs or (B) TBP. (A) Forty-eight individuals from the Global cultivar were analyzed. (B) Leaf discs from individual seedlings corresponding to six different oilseed rape varieties are shown. Cultivar names are at the top. M denotes molecular size standards expressed in nucleotides shown on the left of both gels. Numbers on the right side of Fig. 2B refer to bands amplified by TBP.

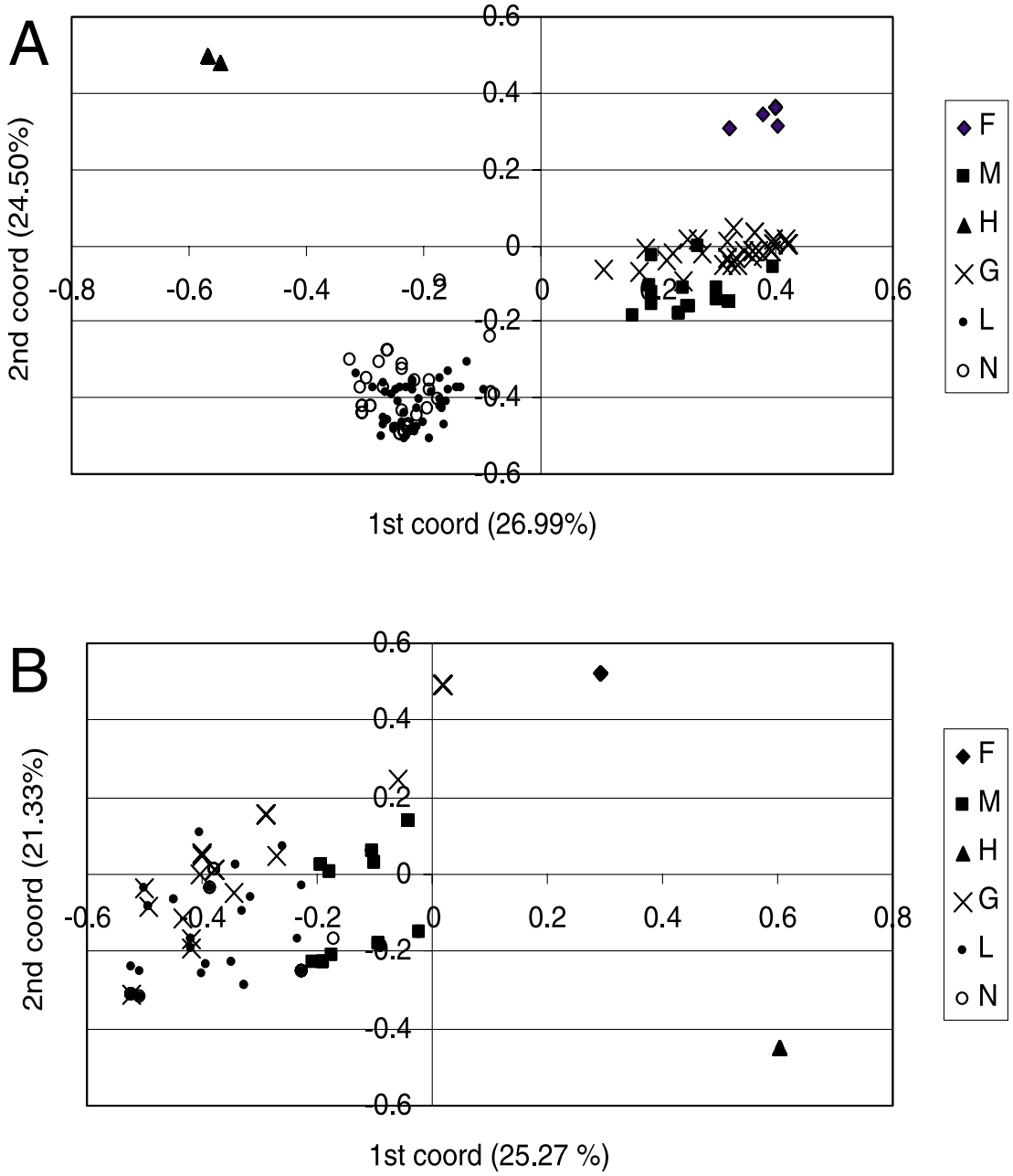


which are in common in the *C. arabica* and *C. canephora* samples (Fig. 5, bands 2 and 6). In fact, all of the *C. arabica* samples show the same four bands irrespective of their origin Conversely, three bands differentiate *C. canephora* from *C. arabica*, one of which (Fig. 5, band 3) is solely, although not always, present in the Madagascar accession compared with the Vietnam accession. Overall, the low level of polymorphism that we found is consistent with the history of coffee dissemination with its narrow genetic base and agrees with the findings obtained by using amplified fragment length polymorphism and SSR microsatellite markers (Anthony et al. 2002). Interspecific polymorphism was found that could be exploited to develop a test for species identification in coffee samples. One band is unique to *C. arabica* and is not found in *C. canephora*. It is also amplified in *C. eugenoides* (Fig. 5, band 7), supporting the hypothesis that *C. eugenoides* is a progenitor of *C. arabica*. Thus, TBP was able to differentiate among coffee species, despite the low number of markers (bands) obtained.

TBP on lotus

TBP was also used to investigate forage legumes belonging to the *Lotus* genus (family Fabaceae), some of which are of agricultural importance such as the birdsfoot trefoil (*L. corniculatus*) (Steiner 1999). The high degree of polymorphisms within birdsfoot has made resolving its ancestry with any single descriptor (cytological marker, plant secondary products, isoenzymes, seed proteins, or random amplified polymorphic DNA) difficult and not definitive. *Lotus corniculatus* is considered a complex that encompasses as many as 12 diploid and tetraploid species (Ball and Chrtkova-Zertova 1968). Among them, in addition to *L. corniculatus*, are *L. alpinus*, *Lotus glaber* Mill (*L. tenuis*), and *Lotus uliginosus* Schkuhr (*L. pedunculatus*). With the use of TBP, we have genotyped 84 individuals from five species of *Lotus*. Results of this analysis are shown in Fig. 6. We have studied only those species (*L. alpinus*, *L. tenuis*, and *L. pedunculatus*) that have been widely reported as putative progenitors of *L. corniculatus*. Individuals of *L. angustissimus* were also analyzed. All species are shown with their diploid chromosome number ($2n$ = 12) except *L. corniculatus*, whose accessions were all tetraploid. In addition, *L. alpinus* and *L. pedunculatus* also have individuals that are tetraploids ($2n = 4x = 24$). The

Fig. 3. PCA of pairwise genetic distances of six different varieties of oilseed rape. (A) PCA based on SSR data. (B) PCA based on TBP data. F, Falcon; M, Marinka; H, Hobson; G, Global; L, Lirawell; N, Nimbus.



ploidy levels of the samples were identified by cytogenetic analyses (data not shown). The scorable bands ranged from 200 to 800 bp and they varied from six in some *L. pedunculatus* samples to 16 in some *L. corniculatus* individuals. Figure 6 shows bands amplified within the 205- to 275-bp range. The overall banding pattern obtained from individuals of *L. corniculatus* is extremely polymorphic with profiles that differ even within the same accession. This is a further confirmation of the high genetic variability of this species with polymorphisms occurring at the level of β-tubulin introns. Even so, some of the bands found in *L. corniculatus* are common to the tetraploid forms of *L. alpinus* as well as to bands of the diploid *L. tenuis* and *L. angustissimus* (accession No. 22). No comigrating bands with either the diploid or tetraploid form of *L. pedunculatus* were found. This excludes this species as one of the possible progenitors, in agreement with the results of Campos et al. (1994). In this regard, the TBP profiles obtained from the diploid and tetraploid forms of *L. pedunculatus* show no difference, suggesting that the tetraploid forms are autopolyploids. On the contrary, whereas all of the diploid *L. alpinus* accessions showed a single unique and specific pattern, the tetraploid *L. alpinus* accessions have more bands and extremely variable profiles even within samples from the same accession. This rules out the possibility that *L. alpinus* tetraploids originated from an autotetraploidization event: diploid and

Fig. 4. Comparison of pairwise genetic distance between cultivars calculated by SSR and TBP frequency data. The dotted line indicates an ideal correlation value of 1 for full overlapping.

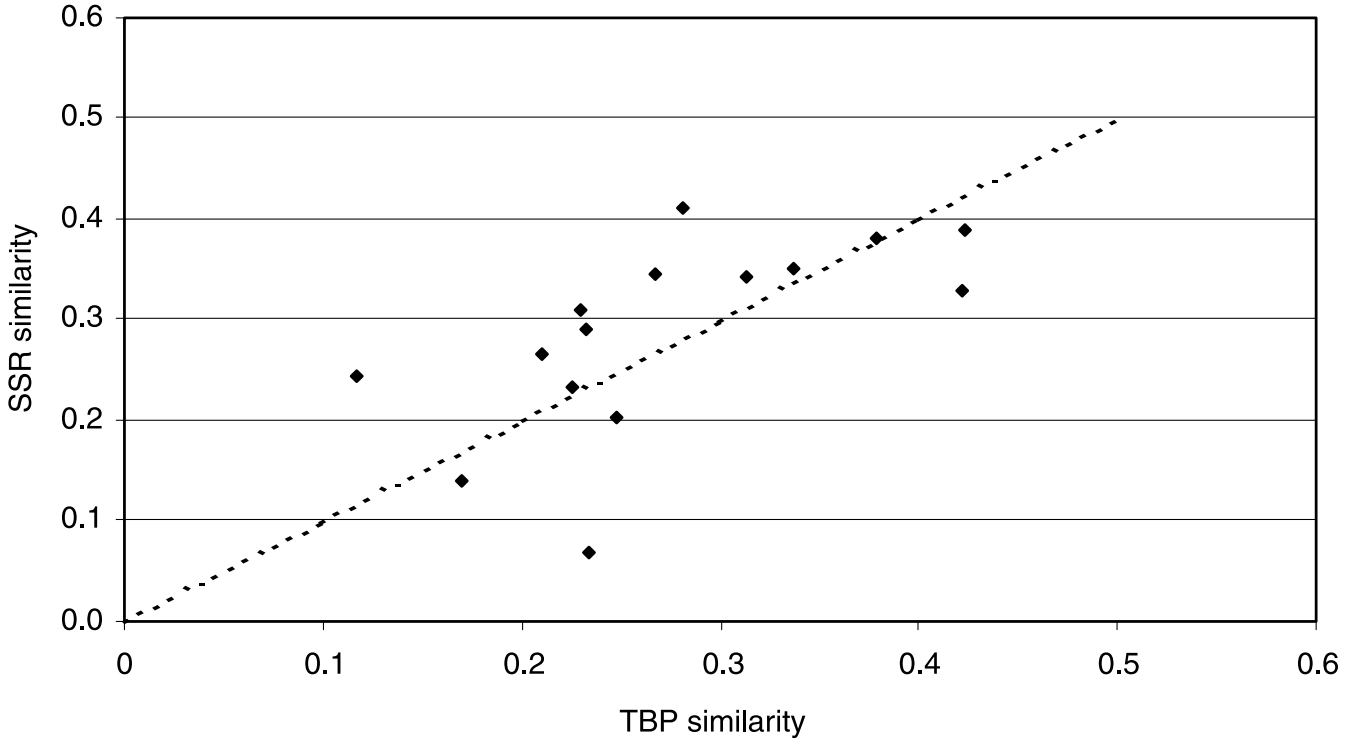
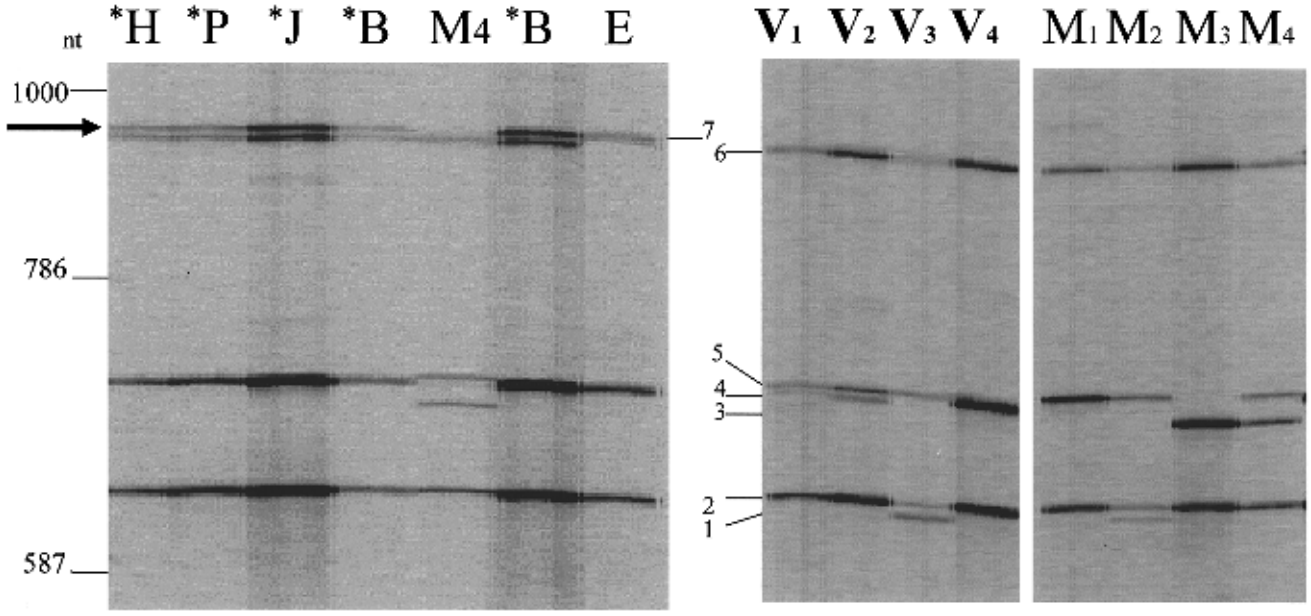
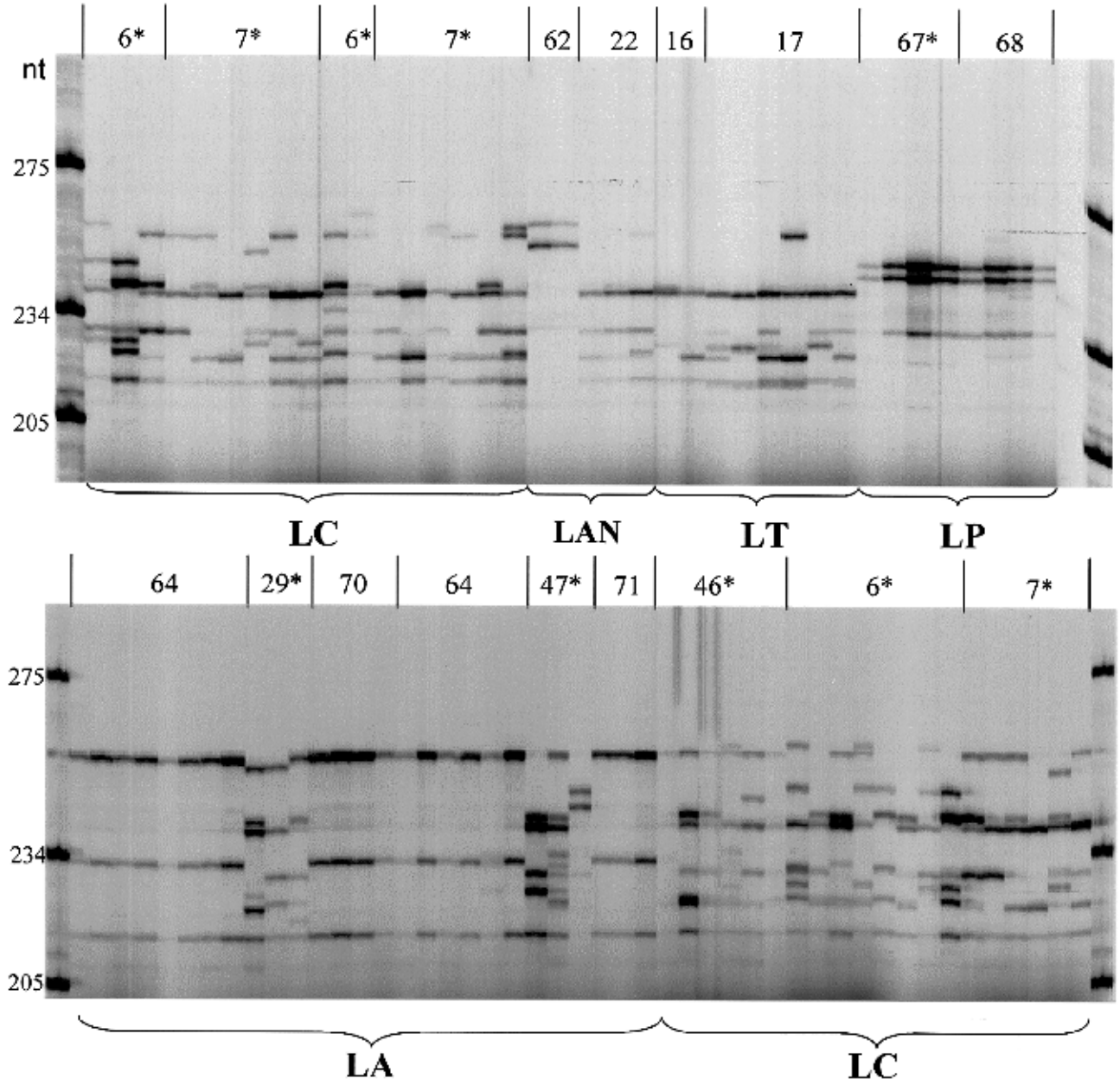


Fig. 5. TBP on coffee. Separation of coffee TBP bands (587–1000 nucleotides) on denaturing polyacrylamide gels. The asterisk denotes *C. arabica* samples and H, P, J, and B denote the origin of the single bean samples from Hawaii, Peru, Jamaican Blue, and Bahia, respectively. V and M denote *C. canephora* from Vietnam and Madagascar, respectively, and E represents a *C. eugenoides* sample. The arrow indicates a species-specific marker (band 7) for discriminating between *C. arabica* and *C. canephora*. The numbers 1–6 are the other TBP-amplified bands.



tetraploid individual fingerprints suggest the possibility that their progenitors (at least one) are from a different species or are due to fertilization from 2*n* gametes. These conclusions are supported by the dendogram derived from TBP data (not shown) that, although not fully supported by bootstrap values because of a paucity of markers, clearly separates *L. pedunculatus* from the list of putative progenitors of *L. corniculatus*.

Fig. 6. TBP on lotus. Separation of lotus TBP bands on denaturing polyacrylamide gels: magnification of the portion between 200 and 275 nucleotides (numbers on the left side). LA, *Lotus alpinus*; LC, *Lotus corniculatus*; LAN, *Lotus angustissimus*; LT, *Lotus tenuis*; LP, *Lotus pedunculatus*. The accession numbers of the samples are shown. An asterisk denotes tetraploid ($2n = 4x = 24$) samples.



Discussion

In the present work, we describe the development of a new tool for estimating genetic diversity that is based on functional sequences, not anonymous DNA markers. The latter, although useful, may not provide information that is strictly associated with important plant features. This technique contributes to bridging the gap between genomics (molecular biology) and phenomics (agrobiology traits).

The new tool that we introduce is called TBP (tubulin-based polymorphism), and this paper describes its development. TBP bases its capacity of detecting genetic diversity on gene intron sequences taken as the source of naturally occurring variability. Introns can be considered as functional sequences for two reasons. First, they are part of the genomic organization of a gene. Second, there is now mounting evidence about plant introns playing an important regulatory role in the control of gene expression (Rose 2002; Gianì et al. 2003).

Compared with current methods, which are mostly based on anonymous DNA markers, TBP shows limits and advantages. On the negative side, the TBP method can only detect a limited set of markers, e.g., the genes that make up the β-tubulin gene family, compared with SSRs, amplified fragment length polymorphism, or other methods and this could make it difficult to derive statistically meaningful phylogen-

etic trees. On the other hand, the use of TBP does not require any previous information about the genomes of interest and, as demonstrated, TBP can be done using a pair of generic primers. Even though a limited smaller numbers of bands are utilized, TBP can help in the characterization of different species and can detect inter- and intravarietal variation, as shown for coffee, oilseed rape, and lotus.

Compared with the few published methods relying on plant intron sequences as the source for DNA polymorphism, TBP has some notable features. It is based on tubulin genes, functionally relevant DNA sequences, dispersed throughout the genome. As such, TBP provides a multiloci approach not previously reported for any intron-based polymorphism detection. Members of other plant gene families are often closely arranged in the genome, providing an almost single-locus target, as was demonstrated studying polymorphism based on the organization of legumine genes in pea (Turner et al. 1993).

TBP is based on intron length polymorphism, readily detectable by gel electrophoresis without the need for additional analysis of the amplified bands (sequencing or restriction cutting). Length polymorphism is rarely detected when intron sequences of plant organellar genomes (group II introns) are used (Laroche and Bousquet 1999). Also, compared with analyses based on organellar DNA intron sequences, TBP does not suffer from functional/evolutionary sequence constraints (Kelchner 2002). Compared with mithocondrial- or chloroplast-located molecular markers (Grivet et al. 1999), TBP is based on the Mendelian inheritance of nuclear DNA sequences. This makes TBP less unpredictable and more suitable for speciation studies resulting from high gene flow levels, as may have been the case for *L. corniculatus*. Ultimately, in comparison with other plant intron sequences, tubulin introns are reported to be controlling elements of tubulin gene expression. This functional feature associates with a high level of nucleotide sequence variability found among introns of different tubulin isotypes (Jeon et al. 2000; Morello et al. 2002).

TBP may not be the method of choice to characterize low levels of genetic diversity in highly inbred cultivated crops, but it may be useful for ecological studies aimed at the characterization of the variability of plant resources typical of different niches and geographical areas. TBP may be of help for long-distance evolutionary scale studies or for discovering the progenitors of a species resulting from hybridization.

In its current state, TBP exploits genetic variations in the first intron of the coding sequence of members of the plant β-tubulin genes as the source for DNA polymorphism, but new developments, still based on this functionally important gene family, can be easily foreseen. For instance, the regulatory sequence 5′ upstream of the ATG could be exploited (Geuna et al. 1997). Rice has already been shown to contain polymorphic DNA in two different varieties when promoter sequences of the same β-tubulin isotype (*OsTub*16) were compared (M. Bardini, unpublished observations). The exploitation of genes coding for cytoskeletal proteins brings the analysis of DNA polymorphism into closer association with important functional aspects, as well exemplified by recent data reporting relevant morphological changes in plant right–left symmetry resulting from minor changes in the nucleotide sequences of tubulin genes (Thitamadee et al. 2002).

The reliability of TBP was tested by comparing the data with those obtained using SSR markers, a more classical and informative but also more laborious approach, in individuals of six varieties. The data that we present demonstrate a good level of agreement between the two methods, as ascertained with both PCA and AMOVA. Furthermore, comparison of the two indices of similarity also showed a good value of correlation (0.65) considering that the TBP data were obtained with a maximum of only 13 markers.

When applied to coffee, the low level of genetic diversity present in the cultivated accession of *C. arabica* agrees with the genetic bottleneck caused by the dissemination of this crop to South America. The data also confirmed the results obtained by others that *C. eugenoides* and *C. canephora* were the progenitors of *C. arabica*. A DNA band that differentiates *C. arabica* from *C. canephora*, likely to come from *C. eugenoides*, could be exploited as a *C. arabica* specific marker for coffee testing. In addition, TBP has also identified intraspecific polymorphisms within the accessions of the diploid *C. canephora*. This is consistent with the known allogamy of this species (Herrera et al. 2002).

In *Lotus*, the known complexity of the genus and, more specifically, the high level of genetic diversity characteristic of *L. corniculatus* have been challenged with TBP. Quite consistently with the complexity of the *Lotus* genus, the variability detected at the level of β-tubulin introns was high. Similarly, a high degree of variability was previously detected by an random amplified polymorphic DNA approach (Campos et al. 1994) and by internal transcribed spacer analysis (A. Mariani, unpublished observations). Although our data are not conclusive, the results show that the high genetic diversity present in *L. corniculatus* is further confirmed and that *L. pedunculatus* should not be considered as a progenitor of *L. corniculatus*. This is in accordance with Campos et al. (1994). However, the data indicate that *L. tenuis* and the diploid *L. alpinus* could be considered as progenitors of *L. corniculatus* because of the presence of similarly comigrating TBP markers. This is also consistent with what was reported by Roos and Jones (1985). Ultimately, our data indicate that tetraploid 4*x L. alpinus* did not originate as an autotetraploid from diploid 2*x L. alpinus* but, more likely, from crosses between two diploid species, one of which could be 2*x L. alpinus*, followed by chromosome doubling or via 2*n* gametes. With regard to the latter, TBP may become useful in the characterization and identification of different types of *L. alpinus* that, although similar in their morphological traits, could actually be different as a result of diversification from the same species in the course of speciation and may even be distinct species.

TBP works also in barley, rice, and wheat (D. Breviario, unpublished observations), but polymorphisms were limited and the system is not competitive with other molecular tools available for these species where much genomic information is known. The low level of polymorphism may reflect the high level of inbreeding in these agronomically important crops. Although the primers described have been successfully applied to a wide range of plant species, they were not successful in amplifying from members of the Asteraceae

family. Modification of the primer sequences has proved successful in amplifying bands from chrysanthemum and argyranthemum samples (D. Lee, unpublished observations), presumably reflecting some evolutionary event specific to this family.

In conclusion, TBP was shown to be successfully applied to a wide range of species spanning mocots and dicots. The data obtained from TBP on oilseed rape are consistent with the genotyping carried out with the use of SSR markers. TBP on *Lotus* shows that the current method of species identification using morphological descriptors must be questioned. In coffee, the data are consistent with the findings of other studies. TPB is rapid, inexpensive, simple, and reliable and does not require much preliminary information about the genome of the plant of interest. These characteristics make the method suitable for assessing genetic diversity and genome origin with preference for ecological studies where often little genetic information is available.

Acknowledgements

We thank the Italian Ministry of Research (MIUR-Agrobiotecnologie-progetto SCRIGNO) and the Department of Environment, Food, and Rural Affairs for financial support of this work. M. Bardini was funded by a Marie Curie Training Site Fellowship (contract No. QLK5-1999-50753)

References

Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor. Appl. Genet. **104**: 894–900.

Ball, P.W., and Chrtkova-Zertova, A. 1968. *Lotus corniculatus group*. *In* Flora Europea. Vol. 2. *Edited by* T.G. Tutin et al. University Press, Cambridge, U.K. pp. 173–176.

Campos, L.P., Raelson, J.V., and Grant, W.F. 1994. Genome relationships among *Lotus* species based on random amplified polymorphic DNA (RAPD). Theor. Appl. Genet. **88**: 417–422.

Gaut, B.S., Peek, A.S., Morton, B.R., and Clegg, M.T. 1999. Patterns of genetic diversification within the *Adh* gene family in grasses (Poaceae). Mol. Biol. Evol. **16**: 1086–1097.

Geuna, F., Hartings, H., and Scienza, A. 1997. Discrimination between cultivars of *Vitis vinifera* based on molecular variability concerning 5′ untranslated regions of the StSy-CHS genes. Theor. Appl. Genet. **95**: 375–383.

Gianì, S., and Breviario, D. 1996. Rice β-tubulin mRNA levels are modulated during flower development and in response to external stimuli. Plant Sci. **116**: 147–157.

Gianì, S., Morello, L., Bardini, M., and Breviario, D. 2003. Tubulin intron sequences: multi-functional tools. Cell Biol. Int. **27**: 203–205.

Grivet, D., Jeandroz, S., and Favre, J.M. 1999. *Nad1* b/c intron polymorphism reveals maternal inheritance of the mitochondrial genome in *Picea abies*. Theor. Appl. Genet. **99**: 346–349.

Herrera, J.C., Combes, M.C., Anthony, F., Charrier, A., and Lashermes, P. 2002. Introgression into the allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* × *C. canephora*). Theor. Appl. Genet. **104**: 661–668.

Hongtrakul, V., Huestis, G.M., and Knapp, S.J. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor. Appl. Genet. **95**: 400–407.

Jackson, J.A., and Matthews, D. 2000. A modified inter-SSR protocol for use in conjunction with the LI-COR Gene ReadIR 4200 DNA analyser. Biotechniques, **28**: 914–920.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.J., Kim, C., and An, G. 2000. Tissue-preferential expression of a rice α-tubulin gene, *OsTubA*1, mediated by the first intron. Plant Physiol. **123**: 1005–1014.

Karp, A., Peter, G., Ingram, I., and Ingram, D. 1998. Molecular tools for screening biodiversity. Chapman and Hall, London, U.K.

Kelchner, S.A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am. J. Bot. **89**: 1651–1669.

Koebner, R.M.D., Powell, W., and Donini, P. 2001. Contributions of DNA molecular marker technologies to the genetics and breeding of wheat and barley. Plant Breed. Rev. **21**: 181–220.

Laroche, J., and Bousquet, J. 1999. Evolution of the mitochondrial *rps*3 intron in perennial and annual angiosperms and homology to *nad*5 intron1. Mol. Biol. Evol. **16**: 441–452.

Lashermes, P., Combes, M.C., Robert, J., Trouslot, P., D'Hort, A., Anthony, F., and Charrier, A. 1999. Molecular characterization and origin of the *Coffea arabica* L. genome. Mol. Gen. Genet. **261**: 259–266.

Liaud, M.F., Brinkmann, H., and Cerff, R. 1992. The beta-tubulin gene family of pea: primary structure, genomic organization and intron-dependent evolution of genes. Plant Mol. Biol. **18**: 639–651.

Morello, L., Bardini, M., Sala, F., and Breviario, D. 2002. A long leader intron of the *OsTub*16 rice beta-tubulin gene is required for high level gene expression and can autonomously promote transcription both in vivo and in vitro. Plant J. **29**: 33–44.

Morgante, M., and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**: 175–182.

Nick, P. 1998. Signaling to the microtubular cytoskeleton in plants. Int. Rev Cytol. **184**: 33–80.

Nogales, E. 2000. Structural insights into microtubule function. Annu. Rev. Biochem. **69**: 277–302.

Ochs, G., Schock, G., Trischler, M., Kosemund, K., and Wild, A. 1999. Complexity and expression of the glutamine synthase multigene family in the amphidiploid crop *Brassica napus*. Plant Mol. Biol. **39**: 395–405.

Peakall, R., and Smouse, P.E. 2001. GenAlEx: genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia. Available from http://www.anu.edu.au/BoZo/GenAlEx/ [accessed April 2003].

Pecchia, S., Mercatelli, E., and Vannacci, G. 1998. PCR amplification and characterization of the intergenic spacer region of the ribosomal DNA in *Pyrenophora graminea*. FEMS Microbiol. Lett. **166**: 21–27.

Powell, W., Orozco-Castillo, C., Chalmers, K.J., Provan, J., and Waugh, R. 1995. Polymerase chain reaction-based assays for the characterization of plant genetic resources. Electrophoresis, **16**: 1726–1730.

Rohlf, F.J. 1993. Numerical taxonomy and multivariate analysis system (NTSYS-pc). Exeter Software, Setauket, N.Y.

Roos, M.D., and Jones, W.T. 1985. The origin of *Lotus corniculatus*. Theor. Appl. Genet. **71**: 284–288.

Rose, A.B. 2002. Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. RNA, **8**: 1444–1453.

Sallares, R., and Brown, T.A. 1999. PCR-based analysis of the intergenic spacers of the *Nor* loci on the A genomes of *Triticum* diploids and polyploids. Genome, **42**: 116–128.

Sasanuma, T., and Miyashita, N.T. 1998. Subfamily divergence in

the multigene family of ribulose-1,5-biphosphate carboxylase/oxygenase (rbcS) in Triticae and its relative. Genes Genet. Syst. **73**: 297–309.

Small, R.I., Ryburn, J.A., and Wendel, J.F. 1999. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton. Mol. Biol. Evol. **16**: 491–501.

Snustad, D.P., Haos, N.A., Kopczack, S.D., and Silflow, C.D. 1992. The small genome of *Arabidopsis* contains at least nine expressed β-tubulin genes. Plant Cell, **4**: 549–556.

Soranzo, N., Provan, J., and Powell, W. 1999. An example of microsatellite length variation in the mitochondrial genome of conifers. Genome, **42**: 158–161.

Steiner, J.J. 1999. Birdsfoot trefoil origins and germplasm diversity. *In* Trefoil: the science and technology of the *Lotus*. *Edited by* P.R. Beuselinck. CSSA Special Publications No. 28.

Thitamadee, S., Tuchihara, K., and Hashimoto, T. 2002. Microtubule basis for left-handed helical growth in *Arabidopsis*. Nature (London), **417**: 193–196.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673–4680.

Tommasini, L., Batley, J., Arnold, G.M., Cooke, R.J., Donini, P., Lee, D., Law, J.R., Lowe, C., Moule, C., Trick, M., and Edwards, K.J. 2003. The development of multiplex simple sequence repeat (SSR) markers to complement distinctness, uniformity and stability testing of rape (*Brassica napus* L.) varieties. Theor. Appl. Genet. **106**: 1091–1101.

Turner, L., Hellens, R.P., Lee, D., and Ellis, T.H.N. 1993 Genetic aspects of the organisation of legumin genes in pea. Plant Mol. Biol. **22**: 101–112.

Wang, R.I., Stec, A., Hey, J., Lukens, L., and Doebley, J. 1999. The limits of selection during maize domestication. Nature (London), **398**: 236–239.