Mol Breeding (2007) 20:249–259
DOI 10.1007/s11032-007-9087-9

High polymorphism and resolution in targeted fingerprinting with combined β-tubulin introns

Diego Breviario · Wm. Vance Baird · Shail Sangoi · Khidir Hilu · Pietro Blumetti · Silvia Gianì

Received: 23 November 2006 / Accepted: 15 February 2007 / Published online: 10 March 2007

Abstract Tubulin-Based-Polymorphism (TBP) was originally introduced as a novel method for assaying genetic diversity in plants. TBP is based on polymorphism resulting from the PCR-mediated amplification of the first intron in the coding region of the β-tubulin gene family. Although, the method was successful in genetic assessment of some plant species and varieties, it suffered from low number of molecular markers due to limited variation in the first intron of β-tubulin gene family. We have now rectified this limitation by introducing the second intron of the β-tubulin genes as a valuable source of molecular markers. We show that the combined use of the two introns substantially increases the number of molecular markers and results in a reliable assessment of species/varieties relationships. After a preliminary validation on *Brassica*, this new combinatorial method was tested on species of *Eleusine* and *Arachis*. For both, reliable assessment of species relationships were obtained that were consistent with recently published studies resulting from more elaborated methods including DNA sequencing. Combinatorial TBP is a reliable, reproducible, simple, fast, and easy to score method that is very useful for breeding programs and species and variety assessments.

Keywords *Arachis* · *Eleusine* · Intron · Marker assisted selection · TBP · Variety assessment

D. Breviario (✉) · P. Blumetti · S. Gianì
Istituto Biologia e Biotecnologia Agraria
IBBA-CNR, Via Bassini 15, 20133 Milano, Italy
e-mail: breviario@ibba.cnr.it

Wm.V. Baird · S. Sangoi
Department of Horticulture, Clemson University,
Clemson, SC 29634, USA

K. Hilu
Department of Biological Sciences, Virginia Tech,
Blacksburg, VA 24061, USA

Introduction

Various PCR-based techniques may be used to identify cultivars, clones, breeding lines and even individuals, as well as for estimating genetic diversity (Powell et al. 1995). Depending on the classification purpose, functional as opposed to non-functional DNA sequences may be preferred because the former should be more closely related to actual physiological and/or structural differences (Rafalski and Tingey 1993; Koebner et al. 2001; Martin-Lopes et al. 2001). Depending on the depth of phylogenetic analysis, taxon relatedness and degree of resolution may be reduced or lost with the use of slowly evolving DNA sequences/genes, whereas it may be artificially increased using fast-evolving sequences. For example, microsatellite/SSR analysis, although efficient for discriminating between closely

related individuals and material with a narrow genetic base, may be less effective in determining relations at higher taxonomic level (Yang et al. 1994). Introns may be considered as moderate-evolving sequences and their use as markers to assess genetic diversity has met with variable success. Originally thought of as junk DNA, and thereof presumed to evolve under minimal constraints in a fashion consistent with the neutral theory of sequence evolution (Kimura 1983), introns have now become of increasing interest as sources of DNA polymorphism. Recent reports have shown intron length polymorphism to be a convenient and reliable molecular marker with high interspecies transferability and it can be exploited for the construction of genetic maps because it directly reflects variations occurring within genes (Wang et al. 2005).

The results obtained from the application of the original Tubulin Based Polymorphism (TBP) method (Bardini et al. 2004) clearly contradicts any concept of co-evolution of introns and coding sequences. This was evident in several plant species as revealed by the readily detectable length polymorphism occurring in plant β-tubulin genes using PCR primers designed for well-conserved coding sequences flanking the first intron (Bardini et al. 2004). β-tubulins, encoded by multiple genes in plants and most other eukaryotes, clearly show that their exonic coding sequences and intronic non-coding sequences underwent different rates of evolutionary change. Coding sequences experienced a higher level of constraint, and thus lower rates of change, as a consequence of the key functions of tubulin in cell division and chromosome separation. In contrast, tubulin intron evolution has been more relaxed, although some constraint associated with their regulatory role in gene expression may exist (Jeon et al. 2000; Morello et al. 2002; Fiume et al. 2004). The rationale for the TBP method lies in the fact that intron positions in β-tubulin genes from across the plant kingdom occur in clusters, which are regularly spaced throughout the sequence (Liaud et al. 1992). This provides the opportunity to amplify any β-tubulin intron by designing PCR primers homologous to sequences in the flanking exons.

Thus, TBP can provide a faster and easier method to acquire preliminary information for identification and classification purposes of almost any plant species, when compared to other frequently used approaches such as the analysis of AFLP and SSR polymorphisms (Ridou and Donini 1999). These latter methods are more laborious to set up and, although they reveal a large amount of information about the genomes under investigation, in some instances that volume of data may be unnecessary for the identification purpose being pursued. Nonetheless, one limitation of the original TBP method, which exclusively relies on polymorphisms detected within the first intron of β-tubulin genes, was the small number of molecular markers that were obtained in some plant species. This limitation could hinder the accurate assessment of taxonomic relatedness and measurement of genetic distances, particularly when the analysis is performed at the lowest taxonomic levels.

To compensate for this shortcoming, we successfully investigated the use of a second intron in β-tubulin genes as an additional target for PCR-mediated amplification and polymorphism detection. The second intron is present in the vast majority of plant species, maize being the only reported exception (Liaud et al. 1992). Therefore, a combinatorial TBP (cTBP) that uses both intron I and intron II of the β-tubulin genes as the source for genomic polymorphism is expected to produce a greater number of molecular markers associated with each original gene locus. To test this possibility, a new set of primers were designed that could anneal to the coding sequences flanking the second intron of the β-tubulin genes. We examined species from three genera belonging to different monocot and dicot plant families.

We verified the feasibility of this improved approach on *Brassica* (oilseed rape, Brassicaceae), *Eleusine* (finger millet genus, Poaceae) and *Arachis* (peanut genus, Fabaceae). Oilseed rape was utilized to cross-reference data with those produced using the original TBP method (Bardini et al. 2004). The genus *Eleusine* was used to clarify the degree of relatedness among various species, particularly the relationship of the wild type *E. coracana* ssp. *africana* to the domesticated *E. coracana* ssp. *coracana* (finger millet) (Werth et al. 1994). Different *Arachis* landraces and

species were analyzed and classified according to this new, combinatorial TBP assay focusing on section *Arachis* that contains the peanut crop and related wild species. The results indicate that cTBP can be effective for establishing relationships from the species to the cultivar level, for assessing genetic diversity and for assisting in breeding programs.

Materials and methods

Plant material

Oilseed rape (*Brassica napus*) genomic DNAs, extracted from the cultivars Apex, Spok, Linewell, Logo, and Shogun were provided by Dr. D. Lee (National Institute of Agricultural Botany, Cambridge, UK).

A total of 23 accessions from nine species of *Eleusine* were analyzed, including an enigmatic *Eleusine* taxon collected from soybean and cotton fields in Calhoun County, South Carolina, USA (Neves et al. 2005). Genomic DNA was isolated from fresh leaves collected from greenhouse grown plants. Total DNA was prepared using a DNAeasy Plant Mini Kit (Qiagen, Valencia, CA) or by the isolation protocol of M'Ribu and Hilu (1996).

Different accessions from nine species of the section *Arachis* were analyzed. Genomic DNA was extracted from fresh leaf material following Johnson et al. (1995) and Milla et al. (2005). Quality and quantity of genomic DNA preparations were assessed by agarose gel electrophoresis. The final DNA concentration was adjusted to be between 20 and 100 ng/μl. DNA samples were stored at –80°C. Typically 50 ng was used in each PCR reaction.

Primer design and PCR amplification

The pair of primers previously selected (Bardini et al. 2004) for the amplification of sequences surrounding β-tubulin intron I were further modified into the following new oligonucleotides for this study: TBPfex1: 5′-AACTGGGCBAARGGNCAYTAYAC-3′ and TBPrex1: 5′-ACCATRCAYTCRTCDGCRTTYTC-3′. Primers flanking intron II were designed using multiple alignments obtained from CLUSTAL W (Thompson et al. 1994) for the identification of the conserved regions surrounding the intron splice site. Then, several combinations were designed by visual inspection and tested for their ability to amplify a distinct and consistent banding pattern. Eventually the following pair of primers for β-tubulin intron II were selected, TBPfin2: 5′-GARAAYGCHGAYGARTGYATG-3′ and TBPrin2: 5′-CRAAVCCBACCATGAARAARTG-3′.

PCR reactions were performed in a 20 μl volume with the use of an Eppendorf Mastercycler gradient thermal-cycler. Reactions were performed in 1× PCR buffer (10 mM Tris–HCl pH 8.8, 50 mM KCl), 0.1 mM each dNTP, 2.5 mM $Mg(CH_3C0_2)_2$, 50 ng of DNA template, 1.25 μM primer and 0.5 U Taq DNA Polymerase (Eppendorf HQ, Hamburg, Germany). A PCR programme (TBPTDex) capable of performing well with either pair of primers (for intron I and for intron II) was eventually optimized. Following initial denaturation at 94°C for 3 min, the PCR consisted of 14 cycles of 94°C for 30 s, 65°C for 45 s and 72°C for 120 s with a decrease of 0.7°C of the annealing temperature (originally set at 65°C) carried out in each cycle, followed by 15 additional cycles of 94°C for 30 s, 55°C for 45 s and 72°C for 120 s. The reactions were held at 4°C after a final extension at 72°C for 8 min. After PCR amplification, half of each reaction was loaded on a sequencing sizes 6% w/v polyacrylamide gel and run in 1× TBE (90 mM Tris–base, 90 mM boric acid, 2 mM EDTA pH 8.0) overnight at a constant voltage 250 V DC. At the termination of electrophoresis, the gel was silver stained according to a published protocol (Caetano-Anolles and Gresshoff 1994). After staining, the banding patterns were scanned; data collected from reproducible and successful amplification were stored and band sizes were estimated by comparison to molecular mass standards included in each gel. Consistency of TBP-PCR products was checked by performing all reactions at least twice. Control reactions with no primers or single primers were also run.

Data collection and analysis

Fragments were scored as a binary unit character (present = 1; absent = 0). Only unambiguously resolved and reproducible bands were scored. The reproducibility of banding patterns was confirmed by running many duplicate experiments with samples randomly loaded. The mean reproducibility, calculated as identical bands in two repeats, was between 97 and 100%. A genetic similarity matrix using simple matching coefficient was generated using NTSYSpc2.1. A dendrogram was computed using the unweighted pair-group method with arithmetic averages (UPGMA) clustering method (Rohlf 1998). The mean Polymorphism Information Content (PIC) value for TBP markers was calculated using the following formula:

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

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

Results

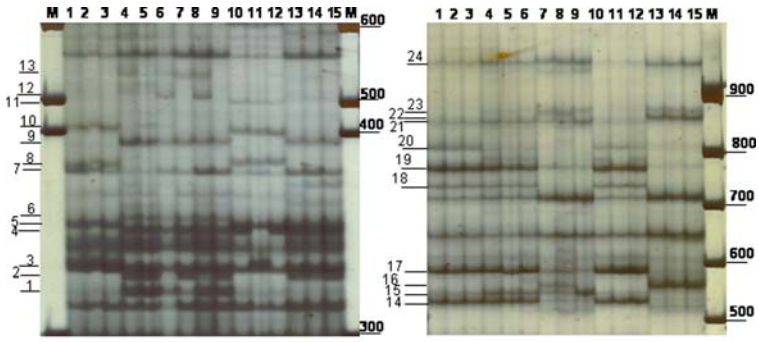
TBP: evaluating the second intron of plant β-tubulin genes as an additional source of DNA polymorphism

The rationale for the cTBP method is based on the concomitant amplification of two introns present at fixed positions, nucleotides 396 and 672 downstream from the adenine of the ATG translational initiation codon, within the coding sequences of the vast majority of plant β-tubulin genes. Preliminary experiments performed on individuals of different plant species such as bean (*Phaseolus vulgaris*), tomato (*Lycopersicon esculentum*), tobacco (*Nicotiana tabaccum*) and rice (*Oryza sativa*) confirmed that the second intron of the β-tubulin genes also could be a reliable source for DNA polymorphism (D. Breviario, data not shown). Both intron I and intron II can then be simultaneously used to set up a combinatorial analysis capable of providing a larger number of diagnostic markers.

The potential effectiveness of combinatorial intron analysis was then assessed in oilseed rape, the species used previously to verify the practicality and reliability of the original TBP method (Bardini et al. 2004). In the present study, we analyzed the banding patterns obtained for both the first and the second intron from five varieties of oilseed rape (Fig. 1). In both gels, polymorphic bands were numbered and the data cataloged by cultivar (Table of Fig. 1). A total of 13 polymorphic bands were scored from amplifications of intron I, and 11 were scored from intron II. The combined patterns from both introns revealed banding patterns specific for each variety (Fig. 1). For example: bands amplified from intron I for cv. Apex and cv. Linwell are almost all polymorphic (only band No. 3 is in common), whereas the pattern of amplification from intron II is substantially the same for both cultivars, with just one difference (presence of band No. 20 in cv. Apex; table of Fig. 1). At the same time, cv. Linwell is very similar to cv. Logo for intron I-derived pattern (band No. 7 being the single difference), but the pattern of intron II is remarkably different between the two, both in number and specificity (only 2 of 12 bands are in common; table of Fig. 1). Similar conclusions can be made for each pair of cultivar comparisons, and the table clearly summarizes these findings.

cTBP analysis of *Eleusine*

Because cTBP was able to differentiate among cultivars of oilseed rape, we tested its applicability to phylogenetic analysis using a problematic monocotyledon grass genus, *Eleusine*. We used TBP, initially the first intron alone, in an attempt to classify individuals representing the nine species of that genus (Fig. 2A). It is clear from the first intron fingerprints that combined data from the two introns were not required to differentiate among the species. The 28 polymorphic markers were sufficient to distinguish unambiguously the



	Intron I	Intron II
1-3. Apex	3, 5,7,8,10	14,15,17,18,19,20,21,24
4-6. Linwell	1,2,3,4,6,9,12,13	14,15,17,18,19,21,24
7-9. Logo	1,2,3,4,6,7,9,12,13	15,16, 21, 23, 24
10-12.Shogun	3,5, 8,10, 11	14,17,18,19,20
13-15. Spok	2,3,4, 5,7,9	16,22,23,24

Fig. 1 cTBP on oilseed rape. Genomic DNA extracted from three individuals for each of five different oilseed rape cultivars (Apex, Linwell, Logo, Shogun and Spok) are analyzed. The gels on the left and on the right show the products amplified from β-tubulin intron I and intron II, respectively. The table at the bottom summarizes the data. Each reproducible amplified band has been numbered on the left side in a progressive order from the bottom of intron I to the top of intron II. Numbers on the right side of each gel refer to molecular sizes standards expressed in nucleotides (M)

nine species in a phenetic analysis (Fig. 2A). It is worthwhile to note that such analysis is accomplished in a single gel, and requires only 1 day to complete. Furthermore, not only are differences easily observed among species, but they also are detectable among accessions of a single species. For example, in the case of individuals representing *E. coracana* ssp. *africana* (i.e., samples 16 through 21 of Fig. 2A) the presence or absence of bands No. 21 or 22 varies among the six accessions. Similarly for the two individuals representing *E. coracana* ssp. *coracana*, band No. 3 distinguishes sample 1 (absent) from sample 2 (present).

It is relevant to note that the number of amplified bands correlates with the ploidy level of each taxon. For example, the three tetraploids *E. coracana* ssp. *coracana*, *E. coracana* ssp. *africana*, and *E. kigeziensis* show a consistently greater number of markers than that produced for each diploid species. Again, the banding patterns produced are characteristic for each species, even among accessions of the conspecific and closely related *E. coracana* ssp. *coracana* and *E. coracana* ssp. *africana*. Figure 2A shows clear difference between these two tetraploid subspecies, the latter being characterized by a doublet (bands No. 8 and No. 9) at about 400 bp and distinctive bands grouping between 750 and 1,000 bp.

Fingerprint data from the first intron (i.e., band presence and absence) were used to perform both a cluster analysis (UPGMA, Fig. 2B) and a principal co-ordinates analysis (PCA, D. Breviario data not shown). The two different analyses provide a substantially similar pattern of distribution among the *Eleusine* species as represented by the specific genetic distances independently calculated in the two analyses. In the UPGMA dendrogram, genetic distances are represented by branch length (coefficient of similarity). Data reported in Fig. 2 clearly indicate a greater degree of similarity between individuals of the tetraploid *Eleusine* taxon ''Calhoun'' and *E. coracana* ssp. *coracana*, than that observed between them and individuals of the tetraploid *E. coracana* ssp. *africana*. This result was unexpected because phenotypic characters such as plant habit (i.e., prostrate vs. upright), seed germination (i.e., requiring scarification and elevated temperatures), panicle and seedhead morphology (i.e.,

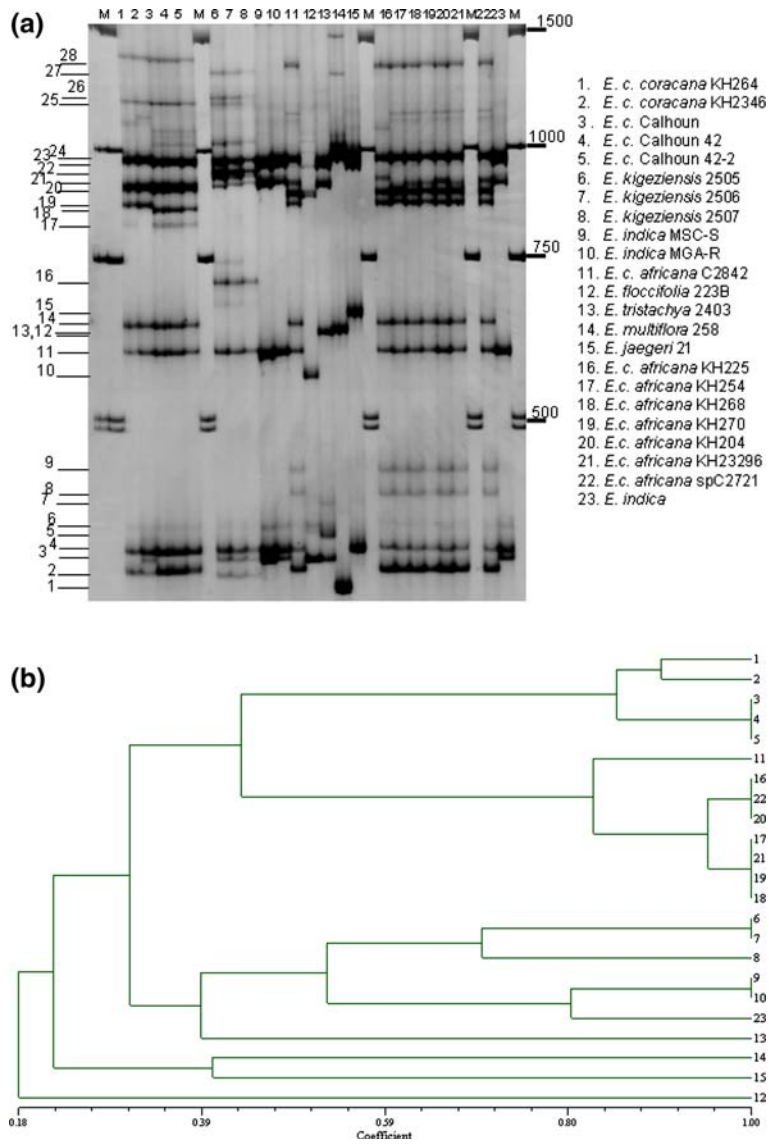
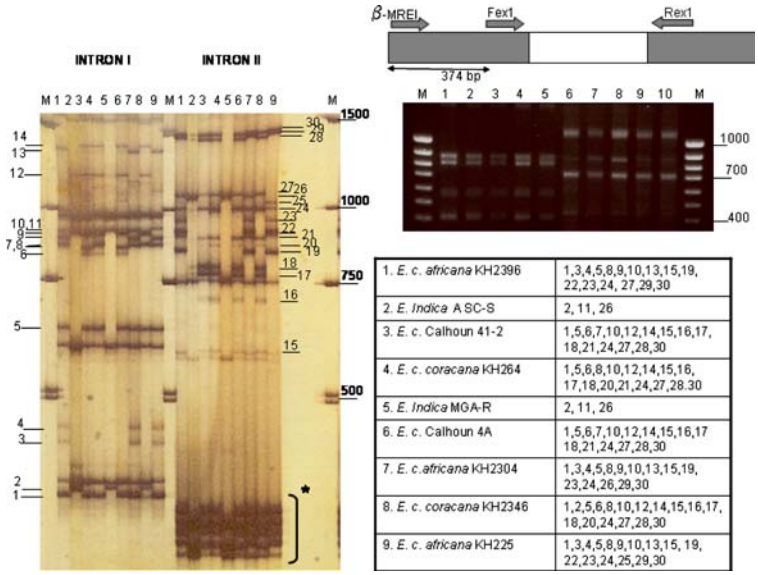


Fig. 2 TBP on *Eleusine*. (**a**) PCR-amplified products from intron I of β-tubulin genes of 23 different individuals belonging to nine different *Eleusine* species. A total of 28 molecular markers was detected (numbers on the left side of the gel). Numbers on the right refer to molecular sizes standards expressed in nucleotides (M). (**b**) Dendrogram generated by UPGMA showing clustering of the nine *Eleusine* taxa

few near-horizontal branches vs. many acute branches) and seed sizes (small and textured vs. large and smooth) show ''Calhoun'' to be nearly identical to *E. coracana* ssp. *africana* (V. Baird, data not shown).

Figure 3 provides an additional comparison between individuals of ''Calhoun'', *E. coracana* ssp. *coracana*, *E. coracana* ssp. *africana* and the diploid *E. indica* (goosegrass, $2n = 18$), which is the accepted maternal progenitor of these tetraploids ($2n = 36$) (Hilu 1995; Neves et al. 2005). In this experiment, further analysis was performed using results from both the first and the second intron of β-tubulin as sources of DNA polymorphism. The Table of Fig. 3 reports the number of each informative polymorphic band that is associated with each individual. Numbering is progressively assigned from the shortest fragment of the first intron (band No. 1) to the largest fragment of the second intron (band No. 30). As can be seen, *E. coracana* ssp. *africana* has eight unique bands (No. 3, 4, 9, 13,19, 22, 23, and 29), whereas *E. coracana* ssp. *coracana* and *E.* sp. ''Calhoun'' differ by only three bands (No. 8, and 20 unique to *E. coracana* ssp. *coracana*, and No. 7 unique to *E.* sp. ''Calhoun''; gel and Table in



1. *E. c. africana* KH2396	1,3,4,5,8,9,10,13,15,19, 22,23,24, 27,29,30
2. *E. Indica* A SC-S	2, 11, 26
3. *E. c.* Calhoun 41-2	1,5,6,7,10,12,14,15,16,17, 18,21,24,27,28,30
4. *E. c. coracana* KH264	1,5,6,8,10,12,14,15,16, 17,18,20,21,24,27,28.30
5. *E. Indica* MGA-R	2, 11, 26
6. *E. c.* Calhoun 4A	1,5,6,7,10,12,14,15,16,17 18,21,24,27,28,30
7. *E. c.africana* KH2304	1,3,4,5,8,9,10,13,15,19, 23,24,26,29,30
8. *E. c. coracana* KH2346	1,2,5,6,8,10,12,14,15,16,17, 18,20,24,27,28,30
9. *E. c. africana* KH225	1,3,4,5,8,9,10,13,15, 19, 22,23,24,25,29,30

Fig. 3 cTBP for *Eleusine* taxa. Genomic DNA extracted from nine individuals representing four closely related *Eleusine* taxa (*E. c.* ssp. *coracana*, *E. c.* ssp. *africana*, *E. c.* sp. *Calhoun* and *E. indica*) was analyzed with respect to polymorphism associated with intron I or intron II of their *β*-tubulin genes. The table summarizes the data. Each reproducible amplified polymorphic band is numbered on the gel and reported in the table. Numbers in bold on the far right side of the gel refer to molecular size standards expressed in nucleotides (M). *Indicates a poorly resolved area that has been ignored for the purpose of the analysis presented here. The diagram and the corresponding EtBr gel (top right) show the shift in molecular sizes of the intron I PCR-amplified products when a forward primer designed to the MREI sequence was used

Fig. 3). It is also evident that the number of bands amplified from the diploid *E. indica* (goosegrass) is significantly smaller than that from the tetraploids. The mean PIC value calculated for the 30 cTBP markers was 0.384, a high value considering 0.5 as the maximum PIC score for any biallelic marker.

Additional cTBP analyses performed on numerous individuals of *E. coracana* ssp. *coracana*, *E. coracana* sp. ''Calhoun'' and *E. coracana* ssp. *africana* further substantiated the close genetic similarity between *E. coracana* ssp. *coracana* and *E. sp*. ''Calhoun'' (D. Breviario, data not shown).

Tracing the *β*-tubulin origin of amplified bands

Evidence supporting the *β*-tubulin origin of the amplified bands is presented in Fig. 3. The ethidium bromide-stained agarose gel shows the shift in sizes of the amplification products from a PCR reaction targeting intron I that are expected by moving the position of the forward primer 300 bp upstream. This control experiment was made possible because the availability of nucleotide sequence for four *Eleusine* *β*-tubulin genes (Yamamoto and Baird 1999) allowed for the design of an N-terminal specific forward primer. To identify the origin of the amplification products, randomly chosen bands were cloned and sequenced from *E. indica*, *E. coracana* ssp. *africana*, *E. coracana* ssp. *coracana* and *E. multiflora*. Although the majority of the DNA sequence of each clone will represent intronic sequence, correctly targeted amplification products will contain sufficient coding sequence to compare to *β*-tubulin cDNA or genomic sequences in the public database. This is because the primer-pairs anneal to exonic sequences (i.e., for intron I primers 97 bp of exon I and 208 bp of exon II). Sequence analysis showed that all clones (five each from *E. coracana* ssp. *Coracana* and *E.coracana* ssp. *africana*, and three each from *E. indica* and *E. multiflora*) were derived from

β-tubulin (i.e., 77–100% nucleotide identity with *E. indica*, data not shown). Not surprisingly, the size-corresponding clones from *E. coracana* ssp. *coracana* and *E. coracana* ssp. *africana* also showed 100% sequence identity. Furthermore, by comparison to the four genomic sequences available for *E. indica* (Yamamoto and Baird 1999), the specific β-tubulin gene from which a particular clone was amplified could be identified. Results showed that one of the three *E. indica* clones was identical (over 581 bp) to TUB1, the second *E. indica* clone was identical (over 888 bp) to TUB2 and the third clone (421 bp in length) showed a maximum of 85% identity to the four known β-tubulin gene sequences (i.e. TUB 1 = 85%, TUB2 = 78%, TUB3 = 81% and TUB4 = 81%). This clone likely comes from an as yet not fully characterized fifth β-tubulin gene in *E. indica*.

cTBP analysis of *Arachis*

The applicability and reliability of cTBP was evaluated further on different individuals from nine species of the genus *Arachis* (Fig. 4). A total of 45 unambiguously resolved, distinct and reproducible bands were scored from the amplification of *A. hypogaea* and eight additional diploid species of *Arachis*. The scored bands ranged in size from 300 to 1,000 bp and the number of polymorphic markers for each species ranged from 7 to 14 for the first intron, and 8 to 15 for the second intron. This number of informative polymorphic markers justifies proceeding with reliable statistical and phylogenetic analysis as documented in the UPGMA analysis (Fig. 5). The dendrogram groups *A. hypogaea* ssp. *hypogaea* in a tight cluster with *A. hypogaea* ssp. *fastigiata* and *A. monticola*, with similarity coefficients of 0.96–1.0. *Arachis ipaensis* emerged as the next most-closely related species to *A. hypogaea.*

Discussion

TBP initially was developed based on evidence first identified in humans indicating that the number and sizes of the coding blocks have been evolutionary conserved in the β-tubulin gene family (Lee et al. 1983). Equally important, the sizes of the positionally conserved intervening sequences can be highly variable. The original single-site TBP method has been improved by increasing the source of variability with the addition of the second

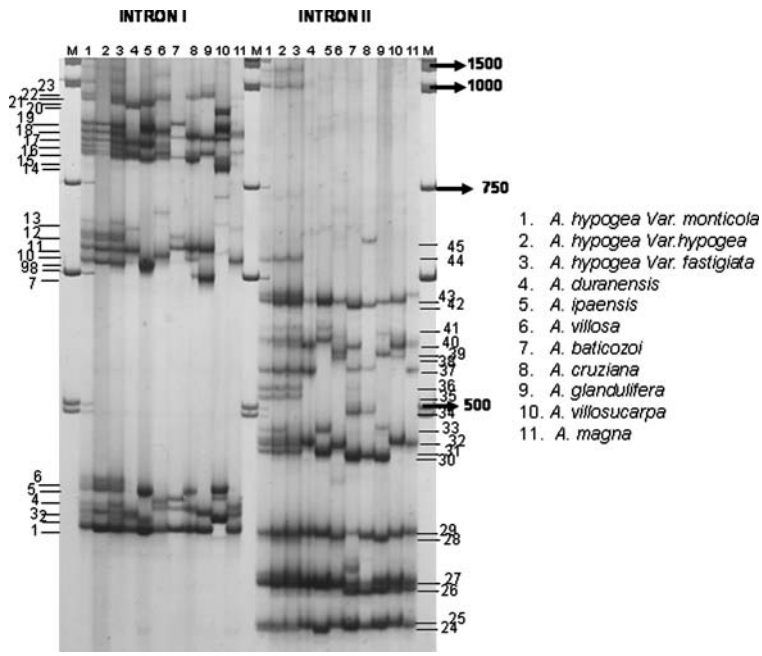


Fig. 4 cTBP of *Arachis* section *Arachis*. Genomic DNA extracted from 11 individuals representing nine *Arachis* species. Data were analyzed with respect to polymorphism derived from intron I or intron II of β-tubulin genes. Each reproducible amplified band has been numbered on the gel in a progressive order from the bottom of intron I to the top of intron II. Numbers in bold and indicated by an arrow on the right side of the gel refer to molecular sizes standards expressed in nucleotides (M)

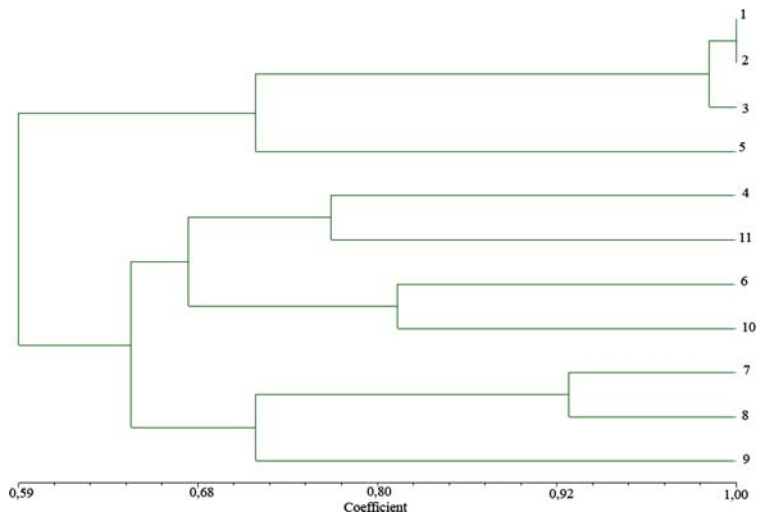


Fig. 5 UPGMA analysis of *Arachis* ssp. cTBP data shown in Fig. 4 displayed as a dendrogram (numbers refer to the 11 taxa investigated from section *Arachis*

intron present within the vast majority of plant β-tubulin genes. This improved, combinatorial version of TBP (cTBP) yields a higher number of molecular markers, allowing its application to a wide range of species and at various intraspecific levels. The choice of whether to use one or both introns as the target DNA sequences depends on the level of genetic characterization required (i.e., diversity assessment or varietal classification). Sizes of the amplified fragments are always larger than 300 bp, which is consistent with the positioning of the primer pairs within the coding sequence and with the 150 bp average size of plant introns (Ner-Gaon et al. 2004). The combinatorial pattern of bands obtained from the amplification of both introns I and II is specific to and unchanged within individuals. The importance of such a combinatorial analysis can be appreciated most in cases where only one of the two banding patterns is variety specific.

cTBP combines reliability with fast accessibility to the raw data and ease of scoring. It is useful for classification at lower taxonomic levels, either as the sole experimental method or as an initial assessment. cTBP has a further advantage that no preliminary information on specific sequence is required to target the β-tubulin introns. In addition, because it is based on length polymorphism, further biochemical or molecular manipulations are not necessary.

cTBP may be successfully applied to the comparison of hybrid varieties with their respective parental lines. It also may be useful for Distinctness, Uniformity, Stability (DUS) testing and to assess the genetic purity among varieties of hybrid crops. By comparison, other molecular methods such as SSR could be too sensitive for stability testing, because parental lines may not be homozygous at microsatellite loci. While β-tubulin data may be considered useful for DUS, their use for molecular phylogenetics is debatable, as is any single-gene based analysis (Rokas et al. 2003). However, this limitation may be overcome by the fact that cTBP analyzes more than one gene, actually up to eight or more, and the marker is derived from loci that are fundamental to plant viability.

Identity and specific gene origin of the amplified bands was examined in *Eleusine* by both an indirect method based on PCR-mediated DNA amplification and by direct DNA sequencing. This testing was possible because *Eleusine* β-tubulin gene sequences were available in public data banks (Yamamoto and Baird 1999). As a matter of fact, the TBP method identified a fifth β-tubulin isotype in *E. indica*, the presence of which was suspected but never confirmed (G. Swire-Clark and V. Baird, data not shown).

Assessment of species relationships in *Eleusine* addresses questions concerning the degree of relatedness between the wild species *E. coracana* ssp. *africana* and the domesticated *E. coracana* ssp. *coracana* (finger millet). This includes the identification of the paternal parent of these

tetraploid taxa, as there is strong evidence to support *E. indica* as the maternal nuclear and organelle genome donor for both (Hilu 1995; Bisht and Mukai 2000; Neves et al. 2005). Furthermore, with respect to these two tetraploids, there is a question as to whether the cultivated species arose via selection from the wild species or the two represent independent polyploidization events. Our results with *Eleusine* provide some evidence on subspecies evolution. As discussed earlier, cTBP data are able to discriminate among the species, and the number of the amplified bands (i.e., molecular markers) correlates very well with ploidy levels. Tetraploid *Eleusine* species (*E. coracana* ssp. *coracana*, *E. coracana* ssp. *africana*, *E. kigeziensis*) show significantly more bands than diploid species (*E. multiflora*, *E. indica*, *E. floccifolia*, *E. intermedia*, *E. tristachya*, *E. jaegeri*). The cTBP data not only confirm that *E. coracana* ssp. *africana* and *E. coracana* ssp. *coracana* are closely related sibling taxa having a common progenitor, but for the first time shows that the wild species is easily discernable from the cultivated (domesticated) species. Discrimination between *E. c.* ssp. *coracana* and *E. c.* ssp. *africana* has been difficult to attain. Therefore, it is more likely that the cultivated species, *E. c.* ssp. *coracana*, arose via selection from the wild *E. coracana* ssp. *africana* rather than the two arising from independent polyploidization events. This implies that *E. coracana* ssp. *coracana* had spread to the New World and our cTBP-based data on landrace Calhoun, found in soybean and cotton fields of Calhoun county South Carolina USA, support this hypothesis. In fact, despite the strong morphological similarity to *E. coracana* ssp. *africana*, the molecular data reported here showed quite clearly that *E. coracana* sp. Calhoun is much more similar to *E. coracana* ssp. *coracana* than to *E. coracana* ssp. *africana*. This is consistent with recent findings based on molecular analysis of internal transcribed spacers sequences (ITS) of the nuclear ribosomal DNA repeats (Neves et al. 2005). Therefore, there are two possible explanations: either Calhoun is closely related to the progenitor from which the cultivated *E. coracana* ssp. *coracana* was selected, or it represents a taxon that escaped from cultivation that has lost many characteristics related to domestication. Overall, the cTBP-data presented here are in agreement with previous analysis concerning the genetic affinity of the *Eleusine* species (Hilu 1995).

Similarly, cTBP data has proven effective in defining species relationships in *Arachis* sect. *Arachis* (Fabaceae). Each of the different species show a highly distinctive pattern of amplification that makes species identification very easy. Overall the cTBP-based data we collected are consistent with previous phylogenetic analyses providing further evidences that *A. ipaensis* is genetically different from the B genome species of *Arachis* such as *A. batizocoi* and *A. cruziana* (Tallury et al 2005). In *Arachis*, cTBP has been shown to be effective when used to discriminate among different landraces and at the individual plant level within them. In contrast, it did not identify variation at the individual plant level in peanut (*A. hypogaea*) (D. Breviario, data not shown). The lack of variation at the level of individual in peanut is in sharp contrast with that observed in the crop finger millet (*E. coracana* ssp. *coracana*) where cTBP detected variability at the individual plant level (Fig. 2A and data not shown). The situation in finger millet was also found in bean (*Phaseolus vulgaris*) ecotypes (D. Breviario, unpublished data) and *Lotus* genus (family Fabaceae) landraces (Bardini et al. 2004). However, TBP may fail to identify variability, even in combinatorial form, in highly inbred species that result from the dispersal of only a few or limited number of cultivars such as the case of peanut. This fits with the general features of the method, and calls for further improvements that are in progress and may result from the amplification of other regulatory β-tubulin sequences.

In conclusion, cTBP is shown here to be effective not only for discerning species relationships and evolution, but also for rapid and reliable identification of plant species and varieties.

Acknowledgements We acknowledge the assistance provided by Dr. Marcello Toschi and we thank Dr. David Lee for his generous gift of oilseed rape genomic DNA samples.

References

Bardini M, Lee D, Donini P, Mariani A, Gianì S, Toschi M, Lowe C, Breviario D (2004) Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species. Genome 47:281–291

Bisht MS, Mukai Y (2000) Mapping of rDNA on the chromosomes of Eleusine species by fluorescence in situ hybridization. Genes Genet Syst 75:343–348

Caetano-Anolles G, Gresshoff PM (1994) Staining nucleic acids with silver: an alternative to radioisotopic and fluorescent labeling. Promega Notes Magazine 45:13–19

Fiume E, Christou P, Gianì S, Breviario D (2004) Introns are key regulatory elements of rice tubulin expression. Planta 218:693–703

Hilu KW (1995) Evolution of finger millet: evidence from random amplified polymorphic DNA. Genome 38:232–238

Hongtrakul V, Huestis GM, Knapp SJ (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

Jeon JS, Lee S, Jung KH, Jun SJ, Kim C, An G (2000) Tissue preferential expression of a rice α-tubulin gene, *Ostub*A1, mediated by the first intron. Plant Physiol 123:1005–1014

Johnson ES, Miklas PN, Stavely JR, Martinez-Cruzado JC (1995) Coupling- and repulsion-phase RADPs for marker-assisted selection of PI181996 rust resistance in common bean. Theor Appl Genet 90:659–664

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

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

Lee MGS, Lewis SA, Wilde CD, Cowan NJ (1983) Evolutionary history of a multigene: an expressed human β-tubulin gene and three processed pseudogenes. Cell 33:477–487

Liaud MF, Brinkmann H, Cerff R (1992) The β-tubulin gene family of pea: primary structures genomic organization and intron-dependent evolution of genes. Plant Mol Biol 18:639–651

Martins-Lopes P, Zhang H, Koebner R (2001) Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels. Plant Mol Biol Rep 19:159–162

Milla SR, Isleib TG, Stalker HT (2005) Taxonomic relationships among Arachis section Arachis species as revealed by AFLP markers. Genome 48:1–11

Morello L, Bardini M, Sala F, Breviario D (2002) A long leader intron of the Ostub16 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

M'Ribu HK, Hilu KW (1996) Application of random amplified polymorphic DNA to study genetic diversity in *Paspalum scrobiculatum* L (Kodo millet, Poaceae). Genet Res Crop Evol 43:203–210

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Flurh R (2004) Intron retention is a major phenomenon in alternative splicing in Arabidopsis. Plant J 39:877–885

Neves SS, Swire-Clark G, Hilu KW, Baird WV (2005) Phylogeny of *Eleusine* (Poaceae: Chloridoideae) based on nuclear ITS and plastid trn-T-trnF sequences. Mol Phylogenet Evol 35:395–419

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

Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet 9:275–280

Ridout CJ, Donini P (1999) Use of AFLP in cereals research. Trends Plant Sci 4:76–79

Rohlf FJ (1998) NTSYS-pc, Numerical taxonomy and multivariate analysis system. Version 2.02iExeter software, Setauket, NY

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:798–804

Tallury SP, Hilu KW, Milla SR, Friend SA, Alsaghir M, Stalker HT, Quandt D (2005) Genomic affinities in *Arachis* section *Arachis* (Fabaceae): molecular and cytogenetic evidence. Theor Appl Genet 111:1229–1237

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acid Res 22:4673–4680

Wang X, Zhao X, Zhu J, Wu W (2005) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). DNA Res 12:417–427

Werth CR, Hilu KW, Langner CA (1994) Isozymes of *Eleusine* (Gramineae) and the origin of finger millet. Am J Bot 81:1186-1197

Yamamoto E, Baird WV (1999) Molecular characterization of four beta-tubulin genes from dinitroaniline susceptible and resistant biotypes of *Eleusine indica*. Plant Mol Biol 39:45–61

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 24:187–194