REVIEW ARTICLE

# MOLECULAR METHODS FOR PLANT TAXONOMY

**Vinay M. Raole* and Nagesh Chirumamilla**

*Department of Botany, Faculty of Science*
*The Maharaja Sayajirao University of Baroda, Vadodara-390002*
*Email: vinaysar@rediffmail.com*

*Received-02.08.2017, Revised-26.08.2017*

**Abstract:** Molecular systematics is the use of molecular genetics to study the evolution of relationships among individuals and species. The goal of systematic studies is to provide insight into the history of groups of organisms and the evolutionary processes that create diversity among species. There are two separate tasks to which DNA specificity is currently being applied. First one DNA data used to distinguish between species which is equivalent to species identification and the second one to discover new species. The aim of this review is to present the techniques that are available to a taxonomist to complement the conventional field methods of identification and delineation of plant species.

**Keywords**: Genetic diversity, PCR, AFLP, DNA Barcode

## INTRODUCTION

Taxonomy is a synthetic science which deals with identification, classification and nomenclature. So, taxonomy provides names, but it is not only a "biodiversity-naming" service: it is also a scientific discipline requiring theoretical, empirical, and epistemological rigor (Dayrat 2005). In the beginning of the molecular era of life sciences the biosystematics was the buzz word and researchers were trying to understand the interrelationship and phylogenetic considerations for explaining the evolutionary concept in plant sciences. The focus of this review is on the use of various nucleic acid techniques i.e. molecular characters for a reliable and efficient taxonomy.

Molecular taxonomy can be grouped into three general approaches referred to as DNA taxonomy, DNA barcoding and molecular operational taxonomic units (MOTU). The terms themselves sometimes lack a clear definition in the literature, and some confusion has arisen from their inconsistent application. A major distinction should be made between species identification, generally associated with the idea of molecular barcodes and species circumscription and delineation, broadly referred to as DNA taxonomy.

Most nuclear sequences targeted in molecular taxonomy experiments belong to the category of highly repetitive DNA. Nuclear ribosomal RNA genes (nrDNA) are tandemly (side by side) repeated and located at a few loci in plant genomes (Hamby and Zimmer 1992, Hayashi 1992, Hillis and Dixon 1992) . These, and particularly the ITS (internal transcribed spacers) (Alvarez and Wendel 2003, Poczoi and Hyvonen 2010), have long been widely used for resolving plant taxonomic issues, initially using restriction analysis and then sequencing.

Restriction fragment length polymorphism (RFLP) analysis was the first technology developed which enabled the detection of polymorphisms at the sequence level. The approach involves digesting DNA with restriction enzymes, separating the resultant DNA fragments by gel electrophoresis, blotting the fragments to a filter and hybridizing probes to the separated fragments. A probe is a short sequence of oligonucleotides which share homology and are thus able to hybridize, with a corresponding sequence or sequences in the genomic DNA. The sequence may be known (e.g. a cloned gene) or unknown (e.g. from random cDNA or genomic DNA clone). Specific probe/enzyme combinations give highly reproducible patterns for a given individual but variation in the restriction patterns between individuals can arise when mutations in the DNA sequence result in changed restriction sites. RFLP analysis is used extensively in the construction of genetic maps and has been successfully applied to genetic diversity assessments, particularly in cultivated plants (e.g. Deu *et al.* 1994; Jack *et al.* 1995) but also in populations and wild accessions (e.g. Besse *et al.* 1994; Bark and Harvey 1995). As a technique for diversity studies, there are three important advantages which should be considered. The first is that RFLPs are highly reproducible between laboratories and the diversity profiles generated can be reliably transferred. The second is that RFLPs are co-dominant markers, enabling heterozygotes to be distinguished from homozygotes. The third advantage is that no sequence-specific information is required and, provided suitable probes are available, the approach can be applied immediately for diversity screening in any system.

There are serious limitations, however, with the RFLP strategy with respect to wide-scale usage at the population level and particularly with regard to its potential for immediate application to any system. Firstly, a good supply of probes is needed that can reliably detect variation at the below species level. It may be possible to utilize (heterologous) probes from other related species, a possibility very much strengthened by syntenic relationships between

*Corresponding Author

related genera. RFLPs are time-consuming and they are not amenable to automation without considerable capital investment. Once probe/enzyme combinations have been selected, throughput will depend on the number of gels that can be run each day in the laboratory in question. To this must be added the factor of DNA extraction. RFLP analysis requires relatively large quantities of good quality DNA (e.g. 10µg per digestion). For some plant systems, where extraction is problematic because of the presence of polyphenols or polysaccharides which complex with the DNA, or where only very limited amounts of source material are available, this feature alone may preclude the choice of RFLP analysis for diversity screening.

Even in those systems where all the above considerations are optimal, the main problem faced may simply be that insufficient polymorphisms are detectable at the below species level by RFLP analysis. Taking all aspects of non PCR-based screening approaches into consideration, it is difficult to envisage that this would be the preferred choice today, given that alternative strategies are now available. When combined with PCR amplification of a specific locus, however, both VNTRs and standard RFLP probes have much to offer.

The development of the polymerase chain reaction (PCR) for amplifying DNA led to a revolution in the applicability of molecular methods and a range of new technologies were developed which can overcome many of the technical limitations of RFLPs. A subset of the latter involve the use of a single `arbitrary' primer, and result in the amplification of several discrete DNA products. Each product will be derived from a region of the genome containing two short segments with sequence similarity to the primer, on opposite strands and sufficiently close for the amplification to work. AP-PCR (arbitrary primed PCR) (Welsh and McClelland, 1990) and DAF (DNA amplification fingerprinting) (Caetano-Anolles, Bassam and Gresshoff, 1991) differ from RAPDs principally in primer length, primer to template ratio, the gel matrix used and in the visualization procedure. The enormous attractions of these arbitrary priming techniques are: (*a*) there is no requirement for DNA probes or sequence information for the design of specific primers; (*b*) since the procedure involves no blotting or hybridizing steps, it is quick, simple and automatable and; (*c*) very small amounts of DNA (10 ng per reaction) are required. The data derived from RAPDs (or AP-PCR and DAF) have their strength in distinguishing individuals, cultivars or accessions, although the difficulty of achieving robust profiles, particularly in RAPDs, makes their reliability for `typing' questionable

One important conclusion is that to achieve the same degree of statistical power using RAPDs (or any other dominant marker system), compared with co-dominant markers, two to ten times more individuals need to be sampled per locus and further, to avoid bias in parameter estimation, the marker alleles for most of these loci should be in relatively low frequency.

Keygene have developed a method which is equally applicable universally, which reveals very high levels of polymorphism and which is highly reproducible. This procedure, termed Amplified Fragment Length Polymorphism (AFLP) is essentially intermediate between RFLPs and RAPDs, in that the first step is restriction digestion of the genomic DNA but this is then followed by selective rounds of PCR amplification of the restricted fragments. The fragments are amplified by $P^{33}$- labelled primers designed to the sequence of the restriction site, plus one to three additional selected nucleotides. Only fragments containing the restriction site sequence plus the additional nucleotides will be amplified and the more selected nucleotides added on to the primer sequence (up to a maximum of three can be added at either site) the fewer the number of fragments amplified by PCR. This selection is necessary to achieve a total number of fragments within the range that can be resolved on a gel (approximately 150 to 200 fragments). The amplified products are normally separated on a sequencing gel and visualized after exposure to X-ray film. Recently, the technique has been automated, using fluorescent labelled primers and, therefore, high throughput can be achieved. Two different types of polymorphisms are detected: (1) point mutation in the restriction sites, or in the selective nucleotides of the primers which result in a signal in one case and absence of a band in the other; and (2) small insertions/deletions within the restriction fragment which results in different size bands.

AFLPs have proven to be extremely proficient in revealing diversity at below the species level and provide an effective means of covering large areas of the genome in a single assay. AFLPs, however, do run into the same problem as RAPDs regarding the type of data generated and the concomitant problems of data analysis for population genetic parameters.

Plants possess three different genomes and, therefore, three potential sources of sequences for a PCR-targeted approach. The chloroplast genome (cpDNA) is uniparentally (often maternally) inherited in plants. It is highly abundant in leaves and therefore amenable to isolation in large quantities. Primers are available that will work either directly, or with small alterations, across broad taxa e.g. across all green plants (Demesure et al. 1995). The majority of studies using sequence data from cpDNA have been phylogenetic ones and at fairly high taxonomic levels (intergeneric and above), although, recently, primer pairs for cpDNA have been used for population studies. In contrast, the mitochondrial genome (mtDNA) of plants is less abundant in leaves and more difficult to extract, there is less background knowledge, fewer primers are available and these

have been less well characterized. The high rates of structural rearrangements mean that mtDNA analysis using restriction site assays is of limited use at the interfamilial and interspecific taxonomic levels but it can be very useful at detecting variation at the intraspecific and population levels. Primer pairs for conserved regions of mtDNA sequences are available (Demesure *et al*., 1995) For assays of the nuclear genome, only the ribosomal RNA (rDNA) gene family has been widely used for diversity studies. rDNA genes are located at specific chromosomal loci (NOR, nucleolar organizing regions) where they are arranged in tandem repeats which can be reiterated up to thousands of times. Each repeat unit comprises a transcribed region separated from the next repeat by an intergenic spacer (IGS). The transcribed region comprises: an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer (ITS1), the 5±8S gene, a second internal transcribed spacer (ITS2) and the 26S gene. Primers pairs have been designed which will enable amplification of the different regions in a wide range of organisms. These regions evolve at different rates and can thus, in principle, be used at all taxonomic levels (Hillis and Dixon, 1991). ITS has proven to be a valuable tool for intergeneric studies in many organisms. Botanists, however, may experience difficulties in detecting sequence variations below the species level.

The advantages of PCR-targeted approaches are in the quality of the data and the information they provide. The fragment in which polymorphisms are studied is of known identity, therefore avoiding the ambiguities of analyzing RAPD and AFLP bands, or random RFLP probes. For population studies, the use of an organellar sequence in complementation with a nuclear sequence can provide particularly illuminating data with respect to mechanisms of differentiation, gene flow and dispersal. In contrast, the origin of RAPD (and AFLP) fragments with respect to the three genomes is generally unknown, although where the origin of the fragments has been studied, there is clear evidence that at least a proportion of RAPD fragments are of cpDNA or mtDNA origin. RFLPs, RAPDs and AFLPs provide indirect data that is only useful when converted into distance measures. This enables frequency data and distance measures to be determined for each genotype class but does not enable the classes to be ordered or grouped in any way. Data based on DNA sequences or restriction site mapping, on the other hand, provide the means of both classifying individuals into different classes and also of assessing relationships among the classes (Braslavsky *et al*. 2003).

There are clear disadvantages of the PCR-targeting approach, however. The first is that, unless the frequency of variants is high enough to be easily detected by PCR-RFLP, or other sensitive gel assays, sequencing will be required which, in turn, necessitates investment of adequate resources and experienced researchers. Another problem is that, although the quality of the data is high, because the approach is often resource-intensive the coverage of the genome is highly restricted, usually to only one sequence and, therefore, to one point of comparison. For the PCR-targeted strategy to be widely applicable, target sequences need to fit two specific criteria. They must contain regions where the sequence is sufficiently conserved such that primers designed for one organism will amplify the same region in a broad range of taxa. At the same time, they must contain regions where the sequence varies at a rate that is high enough for polymorphisms to occur at the population level. Ideally, this should be at a rate such that PCR-RFLP, or rapid assays such as SSCP and TGGE, would uncover sufficient polymorphism, although complete sequencing is the only method that will detect all the variation present. Fortunately, new investigators selecting this strategy do not have to start entirely from scratch, because regions of cpDNA and mtDNA that fit these criteria have been identified, as described above. At the present time, however, there is a dearth of nuclear genes that fit the bill. Furthermore, the rate at which sequences vary (and, therefore, the success of this strategy) appears to differ between genomes and, at present, the limited number of suitable sequences and the worry that those available may not be variable enough in the system under study, are the main reasons why this approach may not be the choice selected.

Most markers generated using RAPD or AFLP technology have been shown by genome mapping experiments to cluster around the centromeres of chromosomes (Saliba-Colombani *et al*. 2000, Qi *et al*. 1998, Saal and Wricke 2002, Young *et al*. 1999), a heterochromatin region with mainly noncoding sequences. Consequently, these markers often reveal an important amount of variation. The evolutionary rate of a molecule is also driven by its evolutionary mechanisms. Microsatellite markers are the most variable molecules known to date. They are mostly noncoding molecules and vary in length (due to the variation in the number of tandem repetitions or VNTR) due to replication slippage (SMM model (Shriver etal 1993), which occurs at a high frequency ($10^{-6}$ to $10^{-2}$) in plants (Bhargava and Fuentes 2010).

Microsatellites or simple sequence repeats (SSRs) are highly mutable loci and, as mentioned earlier, when used as RFLP probes are variable at the population level and can even distinguish individuals and assign parentage. The problems of using them as multi-locus probes, outlined earlier, arise because the repeat sequence may be present in many different regions of the genome. However, since the flanking sequences at each of these loci may be unique, if SSR loci are cloned and sequenced, primers to the flanking regions can be designed and used to amplify

only that single region containing the SSR, which is then referred to as a sequence-tagged microsatellite (or a sequence tagged SSR) (Morgante and Olivieri 1993).

There are several important advantages of choosing sequence-tagged SSRs for population genetic studies. They are usually single loci which, because of their high mutation rate, are often multi-allelic (Saghai-Maroof *et al.* 1994), they are co-dominant markers and they can be detected by a PCR (non-hybridization based) assay. They are very robust tools that can be exchanged between laboratories and their data is highly informative. As with conventional VNTRs, the variation at the SSR locus is caused by changes in the repeat length. Although many such changes can be resolved on agarose gels, it is common to run SSRs on sequencing gels where single repeat differences can be resolved and all possible alleles detected. The assay is relatively quick, but throughput can be increased by selecting a small number of different SSRs with alleles that have different non-overlapping size ranges and multiplexing either the PCR reactions, or the products of the separate reactions, so that all the alleles of the different loci can be run in a single lane on the gel. Multiplexed SSRs have been automated, in which case throughput can be increased further.

There are, nonetheless, some negative aspects of using sequence tagged SSRs. Although they are co-dominant markers, their mode of evolution is different from normal coding loci. Different SSR alleles are thought to arise by slippage or unequal crossing-over and their rate of mutation, and the possibility of deriving the same length alleles by multiple events, mean that it is difficult to use them to estimate relatedness beyond a few generations (Setogouchi *et al* 2009). This in turn means that the phylogenetic information cannot be derived from SSRs. Furthermore, the potentially infinite number of alleles possible at SSR loci make computation of allelic frequencies problematic. Both these features have been addressed by statisticians so that for important population genetic parameters such as $F_{ST}$ estimators for SSR loci ($R_{ST}$) have been derived, but phylogenetic inferences are still limited. Another major problem with choosing this strategy is that unless the investigator is extremely fortunate, sequence-tagged loci will not be available for the system that they wish to study. Although they are ubiquitous, retrieval of SSRs has not been easy in plants because of their relative low abundance compared with animal genomes (Varshney *et al.* 2007). Where they have been isolated, it has often been found that they show limited cross transferability to other genera and even to other species within the same genus.

DNA barcoding was proposed by Hebert *et al.* (2003) as a method for identifying unknown specimens. Short mitochondrial DNA (mtDNA) sequences of cox1 gene, chloroplast genes *rbc*L, *mat*K and nuclear internal transcribed spacer ITS2 are used to group unknown individuals with a priori defined taxonomic entities based on sequence similarity, deriving a species identification from DNA rather than morphological characters. As such, DNA barcoding is not predictive, i.e. it fails when an identical sequence is not available and a limit for admissible divergence has not been established. Hence, DNA barcoding is limited in its potential, as it requires a near complete database of vouchers against which individuals can be placed (Moritz and Cicero 2004, Will and Rubinoff 2004).

However, the extent of genome coverage by molecular markers is partly dependent on the molecular technique that is used, and there is often a trade-off between the possibility due to time and cost limitations of surveying numerous markers and the information content provided by each polymorphic marker such as RAPD or AFLP. An important parameter that is shared by "traditional" and molecular taxonomy studies is sampling strategy and sampling effort. Taxonomy is based on a comparative approach that requires the investigation of as many specimens/samples as possible in order to catch all the extent of natural variation.

It is also clear that the end users of taxonomy such as conservation planners need an operational, character-based, and cheap way to discriminate species (Savolainen *et al*. 2005, Dunn 2003, Alves and Machado 2007). This could tend to diminish the perceived potential of molecular taxonomy, but in this perspective and in spite of the shortcomings that we have just underlined, molecular taxonomy obviously has a great role to play. Current technological capacities do not allow the routine inclusion of the whole genome in taxonomic analyses, choices must be made on the genomic compartment(s) to survey nuclear, mitochondrial, or chloroplast the molecular technique(s) to use and the precise, individual marker(s) to be considered. Another limitation of molecular taxonomy is the possible lack of genetic divergence when sister species have very recent origins because they will share alleles due to recent ancestry and, if reproductive isolation is not complete, to ongoing gene flow—i.e., hybridization. Molecular markers can also suffer from homoplasy, i.e., markers can show similar character states that, however, do not derive from a common ancestor. In this case, they do not inform on the genealogy of taxa and, because they do not reflect a shared evolutionary history, they may be misleading on evolutionary and, as a consequence, on taxonomic relationships.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the support from DST FIST to the Department of Botany, The Maharaja Sayajirao University of Baroda, Vadodara-390002, Gujarat, India.

## REFERENCES

**Ali, I.F., D.B. Nelae and K.A. Marshall** (1991). Chloroplast DNA restriction fragment length polymorphisms in *Sequoia sempervirens* D. Don Endl., *Pseudotsuga menziesii* (Mirb). Franco, *Claocedrus decurrens* (Tott.) and *Pinus taeda* L. Theor. Appl. Genet. 81: 83-89.

**Alvarez , I.A. and Wendel, J.F.** (2003). Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol 29:417–434

**Alves, R.J.V. and Machado, M.D.** (2007). Is classical taxonomy obsolete? Taxon 56:287–288

**Bark, O.H. and M.J. Havey** (1995). Similarities and relationships among populations of the bulb onion as estimated by RFLPs. Theor. Appl. Genet. 90:407-414.

**Besse, P., M. Seguin, P. Lebrun, M.H. Chevallier, D. Nicolas and C. Lanaud** (1994). Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. Theor. Appl. Genet. 88:199-207.

**Bhargava, A., Fuentes, F.F.** (2010). Mutational dynamics of microsatellites. Mol Biotechnol 44:250–266

**Braslavsky, I.; Hebert, B.; Kartalov, E. and Quake, S.R.** (2003). Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA*, *100*, 3960–3964.

**Caetano-Anollés, G.** (1994). MAAP: a versatile and universal tool for genome analysis. Plant Mol. Biol. 25:1011-1026.

**Caetano-Anollés, G., B.J. Bassam, and P.M. Gresshoff** (1991). DNA amplification fingerpinting using very short arbitrary oligonucleotide primers. Bio/Technology 9:553-557.

**Dayrat, B.** (2005). towards integrative taxonomy. Biological Journal of the Linnean Society 85, 407–415.

**Demesure, B., Sodzi, N. and Petit, R.J.** (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast D NA in plants. *Molecular Ecology* 4: 129-131.

**Deu, M., D. Gonzalez-de-Leon, J.C. Glaszmann, I. Degremont, J. Chantereau, C. Lanaud and P. Hamon.** (1994). RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor. Appl. Genet. 88:838-844.

**Dong, J. and Wagner, D.B.** (1993). Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta*. Theor. Appl. Genet. 86:573-578.

**Dunn, C.P.** (2003). Keeping taxonomy based in morphology. Trends Ecol Evol 18:270–271

**Hamby, R.K. and Zimmer, E.A.** (1992). Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman & Hall, New York

**Hayashi, K.** (1992). PCR-SSCP: A method for detection of mutations. Genetic Analysis: Techniques and Applications 9:73-79.

**Hebert, P.D.N., Cywinska, A., Ball, S.L. and DeWaard, J.R.** (2003). Biological identifications through DNA barcodes. Proc R Soc B 270:313–321.

**Hillis, D.M. and Dixon, M.T.** (1991). Ribosomal DNA : molecular evolution and phylogenetic inference. Q Rev Biol 66:411–453

**Jack, P.L., Dimitrijevic, T.A.F. and Mayes, S.** (1995). Assessment of nuclear, mitochondrial and chloroplast RFLP markers in oil palm *(Elaeis guineensis Jacq.).* Theor. Appl. Genet. 90:643-649.

**Karvonen, P., Szmidt, A.E. and Savolainene, O.** (1994). Length variation in the internal transcribed spacers of ribosomal DNA in *Picea abies* and related species. Theor. Appl. Genet. 89:969-974.

**Lynch, M.** (1990). The similarity index and DNA fingerprinting. Mol. Biol. Evol. 7:478-484.

**Lynch, M. and Milligan, B.G.** (1994). Analysis of population structure with RAPD markers. Mol. Ecol. 3:91-99.

**Milligan, B.G.** (1991). Chloroplast DNA diversity within and among populations of *Trifolium pratense.* Current Genet. 19:411-416.

**Moritz, C. and Cicero, C.** (2004). DNA barcoding: promise and pitfalls. PLoS Biol 2:1529–1531.

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

**Poczai, P. and Hyvönen, J.** (2010). Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Mol Biol Rep 37:1897–1912.

**Qi, X., Stam, P. and Lindhout, P.** (1998). Use of locus specific AFLP markers to construct a high density molecular map in barley. Theor Appl Genet 96:376–384

**Saal, B. and Wricke, G.** (2002). Clustering of amplifi ed fragment length polymorphism markers in a linkage map of rye. Plant Breed 121:117–123

**Saghai-Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, R.** (1994). Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc. Nat. Acad. Sci. USA 91:5466-5470.

**Saliba-Colombani, V., Causse, M., Gervais, L. and Philouze, J.** (2000). Effi ciency of RFLP, RAPD, and AFLP markers for the construction of an intraspecifi c map of the tomato genome. Genome 43:29–40

**Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K. and Lane, R.** (2005). Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philos Trans R Soc B Biol Sci 360:1805–1811

**Setoguchi, H.; Mitsui, Y.; Ikeda, H.; Nomura, N.; Tamura, A.** (2009). Development and characterization of microsatellite loci in the

endangered *Tricyrtis ishiiana* (Convallariaceae), a local endemic plant in Japan. *Conserv. Genet.*, *10*, 705–707.

**Shriver, M.O., Jin, L., Chakraborty, R. and Boerwinkle, E.** (1993). VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993

**Varshney, R.K.; Chabane, K.; Hendre, P.S.; Aggarwal, R.K. and Graner, A.** (2007). Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Soc*. **2007**, *173*, 638–649.

**Welsh, J. and McClelland, M.** (1990). Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:6531-6535.

**Williams, J.G.K., Kubelik, A.R., Livak, K.J., Raflaski, J.A. and Tingey, S.V.** (1990). DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic. Acid Res. 18:6531-6535.

**Will, K.W. and Rubinoff, D.** (2004). Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55

**Young, W.P., Schuppet, J.M. and Keim, P.** (1999). DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99: 785–792.