Plant Syst Evol (2009) 282:257–280
DOI 10.1007/s00606-008-0031-1

REVIEW

Non-coding nuclear DNA markers in phylogenetic reconstruction

Miriam Calonje · Santiago Martín-Bravo · Christoph Dobeš · Wei Gong · Ingrid Jordon-Thaden · Christiane Kiefer · Markus Kiefer · Juraj Paule · Roswitha Schmickl · Marcus A. Koch

Received: 13 September 2007 / Accepted: 13 February 2008 / Published online: 21 May 2008

Abstract Molecular DNA based data sets are the most important resource for phylogenetic reconstruction. Among the various marker systems, which were introduced and optimized within the last decade, coding sequences played an important role, especially when molecular clock approaches and multi-gene datasets were assembled. However, non-coding DNA sequences do not only play a quantitatively dominant role, as demonstrated by the two examples nuclear ITS (Internal transcribed spacer regions of nuclear ribosomal DNA) and plastidic *trn*L-F region, but there is also a wide range of different marker systems that can be applied in different ways. Herein, we review the application of several non-coding nuclear DNA marker systems for phylogenetic reconstructions and summarize valuable information for future research.

Keywords ETS · Intron · ITS · Molecular marker · Phylogenetic reconstruction · Promoter · SSR · Transposon

Introduction

In autumn 2006 a special symposium was held in Bonn during the 16th conference of the Section "Biodiversity and Plant Systematics" (German Botanical Society) on various aspects of non-coding DNA in plant phylogenetics and evolution. Among the two most widely used non-coding DNA marker systems are the nuclear encoded ITS (Internal Transcribed Spacer of nuclear encoded RNA) and the plastidic *trn*LF region (*trn*L intron, *trn*LF intergenic spacer). Despite their frequent application throughout the plant kingdom, there is still limited knowledge on the molecular evolution of the marker system itself. In case of ITS the process of concerted evolution/non-concerted evolution and the problem of paralogues and duplicated and/or silenced copies is still only rarely analysed in detail (Wissemann 2000; Koch et al. 2003b; Volkov et al. 2007). Similarly, we still have to learn more about the *trn*LF and other non-coding plastid regions that can resolve deep nodes in angiosperms (Borsch et al. 2003) but is also a valuable source to resolve relationships among species.

In the following we concentrate on six different kinds of nuclear DNA marker systems: (1) ITS, (2) ETS—the external spacer separating ITS tandem copies, (3) SSRs—simple sequence repeats, (4) TE—transposable elements, (5) nuclear introns, and (6) promoter regions.

The internal transcribed spacer (ITS)

The major non-coding nuclear DNA region which has been extensively used for phylogenetic reconstruction at the generic and specific level (reviewed by Baldwin et al. 1995; Álvarez and Wendel 2003; Nieto Feliner and Rosselló 2007) is the biparentally inherited internal transcribed spacer (ITS) of 18S–5.8S–26S nuclear ribosomal DNA. As a part of a transcriptional unit of nrDNA, the ITS is present in virtually all organisms, excluding vertabrates. Thus the fast evolving ITS (in contrast with relatively slowly

Electronic supplementary material The online version of this article (doi:10.1007/s00606-008-0031-1) contains supplementary material, which is available to authorized users.

M. Calonje · S. Martín-Bravo · C. Dobeš · W. Gong · I. Jordon-Thaden · C. Kiefer · M. Kiefer · J. Paule · R. Schmickl · M. A. Koch (✉)
Heidelberg Institute of Plant Science, Biodiversity and Plant Systematics, Heidelberg University, Im Neuenheimer Feld 345, 69120 Heidelberg, Germany
e-mail: mkoch@hip.uni-heidelberg.de

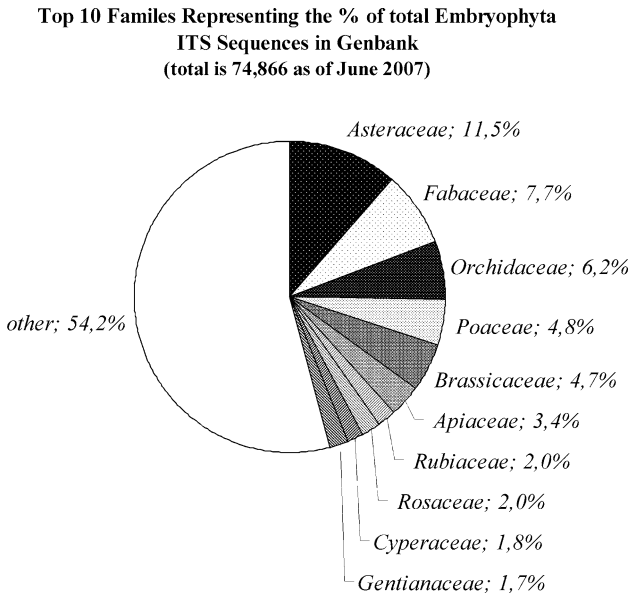

Fig. 1 Percentage of ITS sequences, for the top ten families in Embryophyta (land plants), that were in GenBank as of June 2007

evolving regions such as *rbc*L) has been heavily employed to address questions concerning phylogeny (i.e., evolution of related species), as well as in other applications, such as animal pathology, or epidemiology (e.g. Desquesnes and Davila 2002; Bhattacharya et al. 2007; Walton et al. 2007; Apprill and Gates 2007). Two advantages of ITS is its biparental inheritance in comparison to the maternally inherited plastid markers, and its feature of having high number of copies. These features are useful in studies aiming at unraveling patterns of reticulate evolution, hybrid formation and parentage of polyploids (e.g. Muir et al. 2001; Barkman and Simpson 2002; Koch et al. 2003a, b; Albach and Chase 2004; Fehrer et al. 2007). Even in the specialized study on the tribal classification of Brassicaceae, ITS has shed light at the tribal level between morphologically confusing groups to create a proposed total phylogeny for the entire family (Bailey et al. 2006; Warwick et al. 2006). The high copy numbers allow for highly reproducible lab results, as well as exciting new features involving concerted evolution that are discussed below. As ITS continues to be one of the most common regions for phylogenetic reconstructions, the number of publicly available ITS sequences has tripled since 2003. For Embryophyta, the number of sequences has gone from 23,937 in 2003 to 74,866 in 2007, and the number is increasing by more than 900 per month (numbers counted with script by Markus Kiefer). The top ten families with the most sequences as of June 2007 are represented in Fig. 1. Asteraceae is the highest, followed by Fabaceae, Orchideaceae, Poaceae, Brassicaceae and Apiaceae. The top ten families account for 34,256 (45.8%) of the total Embryophyta sequences. The genus with the most sequences is *Carex*, with 920.

The ITS region consists of three components: the ITS1 and ITS2 spacers and the highly conserved 5.8S rDNA exon located in between (Wheeler and Honeycutt 1988; Fig. 2). The total length is varying approximately from 500–700 bp in angiosperms (Baldwin et al. 1995) to 1500–3700 bp in some gymnosperms (Álvarez and Wendel 2003). Both spacers are not incorporated into mature ribosomes, but undergo a specific cleavage during the maturation of the ribosomal RNAs (Hadijolova et al. 1984; Venema and Tollervey 1999). Hence they may be considered a nearly neutrally evolving DNA marker; however, the spacers contain signals for processing the rRNA transcript and the whole process depends on secondary structure of ITS, implying some degree of conservation (Hillis and Dixon 1991), leading to its usability in phylogenetics.

High copy number and concerted evolution

High copy number of ITS is one of the main reasons for the wide application of ITS in molecular systematics. This fact can be advantageous or a hindrance, and is explained in the following section. The nuclear ribosomal array is a large gene family, characterized by the many individual 18S–5.8S–26S repeats, which are tandemly confined at one or more chromosomal loci in hundreds to thousands of copies in plant genomes (Rogers and Bendich 1987). These loci (regions) are often referred to as NORs (nucleolar organizer regions). According to the concept of concerted

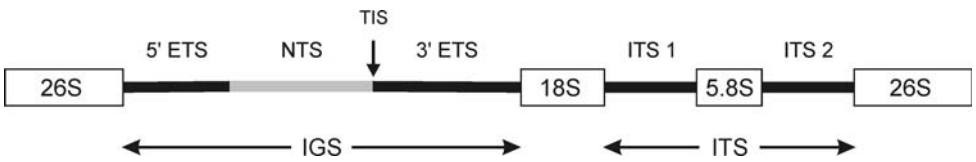

Fig. 2 Schematic presentation of the universal structure of the nuclear ribosomal DNA region showing the alternate position of the 18S, 5.8S and 26S ribosomal genes, which are separated by internal transcribed spacer 1 (ITS1) and internal transcribed spacer 2 (ITS2), respectively. Consecutive gene blocks are by themselves separated by an intergenic spacer (IGS) consisting of a 5′ end and 3′ end external transcribed spacer (ETS) and a central nontranscribed region (non-transcribed spacer = NTS). The 3′ ETS starts at the so-called transcription start site (TIS). The exact position of the transcription termination remains ambiguous in plants. Respective regions are not drawn to scale (modified from Starr et al. 2003)

evolution, these individual copies are homogenized to the same sequence type via mechanisms of unequal crossing-over and high-frequency gene conversion; however, their impact depends heavily on the rates of these processes relative to recombination and mutation (reviewed by Elder and Turner 1995). Processes of intergenic sequence homogenization promote the intragenomic uniformity and prevent the sequence divergence of individual repeat-units, in some cases even between rDNA loci on non-homologous chromosomes (e.g. Wendel et al. 1995a, b). In such a case, if concerted evolution is completed, a direct amplification (PCR) and consequent sequencing can easily be applied. The impact of concerted evolution has been documented in several studies (e.g. Sang et al. 1995; Fuertes Aguilar et al. 1999; Li and Zhang 2002; Koch et al. 2003b). As one can assume, the number of the copies is also partially correlated with the ploidy level as shown in the example of Mishima et al. (2002), which possibly can complicate phylogenies of polyploids. However, the relative contribution of concerted evolution to the reconstruction of phylogenies is not yet fully understood. Due to the lack of empirical data, the process has mainly been investigated on a theoretical level using mathematical models and computer simulations (Nagylaki and Petes 1982; Arnheim 1983; Ohta and Dover 1983, 1984; Walsh 1987; Stephan 1989).

It has also been shown that concerted evolution does not act immediately after organismal processes such as hybridization or polyploidization, or after genomic changes like gene and chromosome segment duplication, and various forms of homologous and non-homologous recombination. Hence, individual gene copies are not homogenized and multiple divergent rDNA copies can then be present, representing different orthologues and paralogues (also referred to as ribotypes). In this case, ITS can be utilized in detection of historical or recent hybridization or introgression (Emshwiller and Doyle 1998; Vargas et al. 1999; Widmer and Baltisberger 1999). Contrary to the classical concept of concerted evolution, there is an increasing evidence for the non-concerted evolution of ITS in specific lineages of plants (e.g. Ruggiero and Procaccini 2004; Harpke and Peterson 2006; Bayly and Ladiges 2007). Non-concerted evolution is charaterized by a presence of high numbers of rDNA copies, which are highly polymorphic within one organism. As it is assumed that not all copies are functional, but rather degenerate, they are often referred to as pseudogenes. These non-functional copies are expected to evolve at a higher rate and are characterized by increased substitution rates in conserved regions and reduced stability of secondary structure (Bailey et al. 2003). Studies regarding these alternative copy types are elucidating further inferences about the detailed history of a species or lineages, i.e. phylogeography.

There is a wealth of recent studies aimed at sampling ribotypes in order to elucidate phylogeographic history of hybridization or gene flow (note: not phylogeny). For example, Nieto Feliner et al. (2004) describe the possible evolutionary mechanisms of the ribotypes of *Armeria* (Plumbaginaceae) in the Spanish Sierra Nevada. Their results showed some correlation with geography, which indicates that the ribotypes found within a population can shed light on the history of hybridization and/or genetic drift in the area. More recently, despite the appearance of many ribotypes in a study of *Buxus* sp. in the Mediterranean basin, a phylogeographic split of western and eastern accessions was still able to be detected (Rosselló et al. 2007).

When dealing with orthologues, paralogues or pseudogenes in the laboratory, several important issues have to be considered. After detection of polymorphic sites in the sequence data, methods such as cloning or molecular cytogenetic techniques are often employed in order to recognize these different copies. Pseudogenes could be consequently identified by four main methods reviewed by Bailey et al. (2003). However, the current views on the psuedogene/paralogue sequences utilization in a phylogenetic analysis can either be complete exclusion (e.g. Yang et al. 1999) or inclusion (Buckler et al. (1997). On one hand the exclusion of polymorphic sites from phylogenetic reconstruction can eliminate the effect of random copy selection and specify the phylogeny as done by Woods et al. (2005). On the other hand the inclusion of these regions leads to an application of a "tree-based" approach for identifying pseudogenes, which provide more characters for increased statistical support. Finally, a recent review on ITS has diagrammed the necessary lab protocols for successful ITS sequencing from plants, including the identification of multiple copy types (Nieto Feliner and Rosselló 2007).

Recent advances

The following explanation describes how the secondary structure of the ITS2 RNA transcript has been proposed to be an ideal novel tool for phylogenetics (reviewed by Coleman 2003). Selection of the candidate region has been done by the level of sequence conservation within the ITS region. Compared to the 5.8S gene, ITS1 and ITS2 are considerably more variable in primary sequence. Moreover, ITS2 is slightly more conserved than ITS1 and can allow for alignments in ranks above the genus level (Hershkovitz and Lewis 1996). The helical structure that is created during the maturation of the rRNA is essential for the recognition and function of the complex to be converted into mature rRNAs, and therefore subject to selectional constraints (Hadjiolova et al. 1994; Lalev and

Nazar 1999; Venema and Tollervey 1999). The folding results in domains with both paired and unpaired bases, implying different evolutionary constraints. More variable segments appear (e.g. Schlötterer et al. 1994; Liu and Schardl 1994; Mai and Coleman 1997) to evolve at neutral rates, while the relatively conserved regions are obviously stabilized by selection. With respect to a phylogenetic analysis using the direct sequence, this divergence within the spacer sequence led to attempts of differential weighting of paired and unpaired nucleotide positions (e.g. Hillis and Dixon 1991; Springer et al. 1995). However, when reconstructing a secondary ITS2 structure, many different alternative structures can arise (Wolf et al. 2005) and it is possible the weighting could be misapplied.

A complete description of this new use of secondary structure for phylogenetics can be found within the online program called ITS2 (Schultz et al. 2006). This program uses the rRNA ITS2 sequences in the Genbank database and predicts a potential secondary structure for a sequence by comparing it to other sequences in the database. The graphical data is then put into another format to develop trees based only on those characters that differ in order to create the numerous secondary structures. It is important to note that the phylogenetic inferences made from these predicted secondary structures, is based on a completely different set of characters than with reconstructions of phylogeny with the raw sequences themselves.

As with other markers, determining substitution rates of ITS has been a focus of much research. However, substitution rates vary from species to species and cannot be applied universally to related groups of plants. Kay et al. (2006) collected 29 different ITS substitution rates from the literature. Their review included 21 different plant families for the 29 different rates reported. In several studies substitution rates for ITS ranged from 0.38×10^{-9} to 8.34×10^{-9} substitution/site/year. Within the families reported, they found no correlation between phylogenetic relatedness at the familial level, as expected due to the coding and non-coding portions of the ITS region. However, Kay et al. (2006) concluded that there is a positive correlation between growth habit and similar substitution rate. For herbaceous annual/perennial rates, the average was 4.13×10^{-9} substitution/site/year. For woody rates, the average was 2.15×10^{-9} substitution/site/year. With this knowledge in hand, it is preferable to determine the ITS substitution rate of each species studied, and not use a previously published rate for calculating the divergence times or the age of a species. In addition, when reconstructing a phylogeny of a genus with species that are perennials and others that are annuals, the two groups should be considered separately, for they may have radically different substitution rates. Other factors that are hypothesized to influence substitution rates are generation time, DNA repair systems, speciation rates, and population structure. All of which are highly variable within a genus.

Another use of the ITS is the recent attempt to determine a marker for DNA based identification by implementing a bar coding system for the plant kingdom as part of the Barcode of Life project. However, it has recently been proposed to use the plastid markers such us *rpo*C1, *rpo*B and *mat*K or *rpo*C1, *mat*K, and *psb*A-*trn*H (Chase et al. 2007). In this proposal for standardizing the protocol for barcoding, it is argued that the occurrence of multiple ITS copies occurs too often to be considered useful across the entire plant kingdom.

In summary, the use of ITS as a marker for evolutionary studies and phylogenetic reconstructions has been questioned, yet it can still be useful if proper considerations are made. The phylogenetic use of this biparentally inherited marker for studying neutral and reticulate evolution at the genus and/or species level is still valid. In addition, whether the system being studied obeys concerted evolution or not, ITS can elucidate complicated phylogeographic histories within and between populations. Luckily, the laboratory practices needed for ITS sequencing are relatively simple and becoming more automated, which will accelerate the acquisition of even more sequences. Recent advances involving the inclusion of the comparison of the predicted secondary structures, not comparing the sequences themselves, which are inferred from the sequences of the RNA of ITS2, show promising advances in phylogenetic reconstruction of species. ITS will continue to be a valuable nrDNA sequence for phylogenetic inference.

The external transcribed spacer (ETS)

The external transcribed spacer (ETS) of the ribosomal nuclear DNA was adopted as a phylogenetic marker by a number of studies since its use was promoted by Baldwin and Markos (1998). ETS is part of the intergenic spacer, which is between each of the repeated blocks of 18S, 5.8S and 26S ribosomal DNA genes. These genes are by themselves separated by ITS1 and ITS2, which were often analysed in combination with the ETS (Fig. 2).

The ETS as understood here constitutes the 3′ end part of the intergenic spacer of the ribosomal genes (rDNA IGS) and begins by definition with a motif associated with the start of the ribosomal transcriptional process, the so-called transcription start site (TIS; e.g. Bhatia et al. 1996; Volkov et al. 2003). We will focus on this 3′ end ETS because other DNAs present in the rDNA IGS like the 5′ end external transcribed spacer (downstream of and bordering the 26S rDNA gene) and the non-transcribed spacer (NTS) were only occasionally used in phylogenetic studies

Table 1 Statistics comparing sequence characteristics of ETS and ITS as extracted from 64 published studies using both regions to address the evolution of particular taxonomic groups

	Sequence length (bp)	Sequence divergence (%) uncorrected *P* values	Parsimony informative characters (*N*)	Consistency index (CI) of shortest tree (s)[a]
ITS	575 ± 50 (106; 375/814)	13.78 ± 8.54 (26; 2.8/39.7)	110 ± 75 (38; 7/291)	0.7 ± 0.14 (21; 0.432/0.923)
ETS	595 ± 327 (74; 314/2100)	16.53 ± 9.76 (25; 3.2/40.1)	136 ± 94 (39; 19/375)	0.73 ± 0.11 (21; 0.56/0.94)
ETS:ITS ratio	–	1.34 ± 0.52 (25; 0.65/3.15)		–

Measures of descriptive statistics are given in the following format within each cell: mean ± standard deviation, (*N*; min/max)

Apocynaceae, Araliaceae, Asteraceae, Burseraceae, Costaceae, Crassulaceae, Cyperaceae, Fabaceae, Malvaceae, Moraceae, Myrtaceae, Oleaceae, Onagraceae, Orchidaceae, Phrymaceae, Poaceae, Rosaceae, Rubiaceae, Saxifragaceae

[a] The CI excluding autapomorphies is usually used. However, in several studies no distinction between CIs excluding or CIs including autapomorphies was made by the respective authors

(Becerra 2003; see Starr et al. 2003 for more detailed introductions to the structure of the IGS).

The predominant application of the 3′ end ETS over other rDNA IGS components in evolutionary studies is due mainly to its higher degree of conservation, both at the structural as well as the sequence level. The rDNA IGS is known for a gradual decrease in sequence conservation upstream from the 18S gene to the center of the rDNA IGS that consists of repetitive elements (McIntyre et al. 1988; Perry and Palukaitis 1990; Takaiwa et al. 1990; Bhatia et al. 1996; Nickrent and Patrick 1998; Murakami 2001). The latter shows extremely high evolutionary dynamics (Bhatia et al. 1996; Fernández et al. 2000). These properties of the rDNA IGS region cause serious obstacles to the development of primers and its alignability apart from the ETS, even at low taxonomic levels. Both external transcribed spacers (the 3′ end and the 5′ end) share the advantage of being mostly single in sequence, i.e., containing no or few DNA-repeats only. The 5′ end ETS, however, seems to be considerably shorter and on average evolves faster than its 3′ end counterpart (Starr et al. 2003).

Prior to Baldwin and Markos' (1998) influential publication, data on the molecular properties and evolutionary dynamics of the ETS were almost exclusively available for various crop plants (cf. Bhatia et al. 1996; Nickrent and Patrick 1998), stemming from functional studies. Length of the nrDNA IGS varies between 1 and >12 kbp (Rogers and Bendich 1987) and, therefore, could not be sequenced without the use of internal primers. Nested primers of universal applicability cannot be developed for the nrDNA IGS due to the lack of suitable and sufficiently conserved sequence motifs. Baldwin and Markos (1998) made this region available to plant systematists by inventing a protocol for the design of ETS sequencing primers for specific taxonomic groups. That requires reverse sequencing of various IGS amplicons starting from the 18S gene and design of forward primers within possibly found (5′ end-most) conserved regions of the obtained ETS sequences. As a consequence, numerous additional family-to-genus-specific primers subsequently have been published, including more than 60 studies covering at least 19 plant families (Table 1; supplementary literature). Based on these data sets, several characteristics of the ETS with evolutionary significance became obvious. The four most important ones will be discussed in the following. This discussion is based mostly on a thorough evaluation of statistics and corresponding statements made by the respective authors of the original articles. No reanalyses of their data have been performed in order to reach these conclusions leaving the potential to deepen our understanding of evolutionary particularities and the performance of this marker system in phylogenetic reconstructions, especially in comparison to other marker systems.

First: the ETS evolves at exceptionally high rates

This property can be best demonstrated by comparing the most frequently used plant nuclear marker, the nrDNA ITS (Table 1). The nrDNA ITS exhibits a mean substitution rate of about 2.86×10^{-9} substitutions/site/year, which are already among the highest known from plants (Kay et al. 2006). While absolute substitution rates have not been calculated for ETS and may not be validly comparable to those of ITS, because existence of a universal substitution rate seems highly unlikely (Kay et al. 2006), it was shown that relative rates are 1.3–7 times higher in ETS than in ITS (Baldwin and Markos 1998; Bena et al. 1998; Linder et al. 2000). Expressed in terms of sequence divergence, meaning percentage of sites distinct between two given sequences, ETS was equal to or more variable than ITS in 22 out of 25 comparisons based on identical genomic DNA samples. These measures provide empirical evidence for higher ETS than ITS variability, i.e., higher variability per ETS site than ITS site. An analogous picture is obtained

when absolute numbers of parsimony informative characters available to a study are considered. Numbers of variable sites are directly dependent on sequence length, which may be another, at least, practical criterion for selection of markers for phylogenetic reconstructions. Although reliable data on the actual size of the ETS are published only for a limited number of taxa (most studies do not refer to the position of the 5′ end primer relative to the ITS), numbers of ETS sites available to address a particular evolutionary question are usually of a magnitude similar to ITS (575 ± 50 bp in ETS vs. 595 ± 327 bp in ITS), although variance of ETS-lengths is much higher than in ITS (Table 1). Several authors pointed also to the higher topological resolution of ETS versus ITS based phylogenetic trees in their particular studies. Although the ETS proved superior over ITS in terms of levels of sequence divergence, numbers of parsimony informative sites, and resolving power in the majority of cases, some exceptions with respect to at least one of these parameters became known (e.g. Ford et al. 2006; Plovanich and Panero 2004), precluding establishment of an absolute rule. Despite these obvious advantages of ETS in phylogenetic studies, homology of sequences usually cannot be assigned unambiguously at higher taxonomic levels (approximately from the tribe onward). For example, *Cynara* and *Onopordon* (Asteraceae) belong to the same tribe but showed only a 27% match of sequence sites for the ETS region about 450–400 bp left to the start of the 18S gene (Tucci et al. 1994).

Second: the phylogenetic signal of the ETS is mostly congruent with that of the nrDNA ITS

Even in the case that ETS provides less resolution to phylogenetic reconstructions than ITS, this region may still be valuable for this objective. ITS and ETS based trees were found to be congruent with each other in 25 out of 32 studies applying both markers. Based on this congruence it was further suggested that there is only low probability of recombination between these regions (Bena et al. 1998). Even when incongruence tests found discordance between data sets, according to observation conflicts were mostly restricted to single or few accessions only and/or reflected hybridization (e.g. Chan et al. 2002; Jousselin et al. 2003; Okuyama et al. 2005). The discordant placement of hybrids in the ITS and ETS trees is of special interest, for it might indicate that there was recombination within the ribosomal array, i.e., resulting in a linear arrangement of xenologous parts. Because it was repeatedly shown that combined analysis of ITS and ETS datasets usually results in higher support and/or resolution of trees, the addition of an ETS dataset appears as a good possibility to further improve an existing ITS-based phylogeny.

Third: concerted evolution is usually operative at high levels and outstrips rates of speciation

In the majority of cases, no or only marginal intra-individual ETS sequence polymorphism was observed. Hence, concerted evolution eliminating such polymorphisms should be effective in most systems. No multiple ETS types were found even in allotetraploid members of the genus *Stylosanthes* (Fabaceae; Stappen et al. 2003). Interestingly, higher levels of intra-individual polymorphism were discovered in ITS paralogues than in ETS paralogues in some studies, which seems to indicate that concerted evolution may operate at different rates in the ETS and ITS region (Plovanich and Panero 2004). Nevertheless, hybridization was found to result in significant numbers of ambiguous sites (Wichman et al. 2002; Andreasen and Baldwin 2003; Noyes 2006) and in one case depression of concerted evolution was suggested (Wichman et al. 2002). Although different ETS sequences may coexist for some time within single genomes, concerted evolution should eliminate the problem of divergent paralogues on the long run.

Theoretical modelling further suggests that with an elevated rate of concerted evolution the correct species tree may be inferred from sequence data for repetitive regions of the genome such as rDNA (Sanderson and Doyle 1992). However, concerted evolution also may cause rapid erosion of molecular evidence of hybridiziation and bias phylogenetic reconstructions towards false scenarios, especially if a particular progenitor sequence is favored above others (Feliner et al. 2001). Nevertheless, significant reduction of the paralogy problem in general can greatly facilitate reconstruction of species trees. Therefore, like in ITS, concerted evolution promotes the use of the ETS region for phylogenetic reconstructions. Nevertheless, concerted evolution may not operate at all levels of rDNA organization, which will be discussed in the following paragraph.

Fourth: the occurrence of divergent repeats sometimes requires assessment of orthology versus paralogy of copies

Although most of the ETS is single in sequence, repeated DNAs also occur regularly. Homogeneity of whole ETS copies is usually maintained within single rDNA clusters and even throughout genomes via the process of concerted evolution. However, this process may not be necessarily effective on the level of subrepeats, which are repeated regions within a repeat. Thus, Baldwin and Markos (1998) found little evidence of concerted evolution between two tandem subrepeats. In this particular case, only two identical mutations and two simultaneous changes to a common state from different ancestral states occurred in both copies, and blocks of consecutive variable sites showing similar

timing and direction of change were lacking. This example underscores the existence of paralogous copies at various levels in ETS sequences and the need for caution in interpreting rDNA variation for phylogenetic studies. Separate phylogenies of copies isolated from different individual organisms should be reconstructed in such cases: If copies coming from different accessions each cluster together, they already should have been in existence prior to the evolutionary diversification of the taxa under study, and copies within each of these clades may be treated like orthologues. However, evolution of repeats often may be complicated by processes like recombination or gene conversion. Therefore, it is highly recommended that rigorous tests of congruence of the phylogenetic signal contained in the repeat with that of regions outside are performed (see for instance Linder and Rieseberg 2004) before its inclusion into a given phylogenetic analysis. In contrast, grouping together of copies from a single accession indicates their more recent origin and little divergence between copies and such a repeat should in general be excluded from any phylogenetic study. Otherwise wrong homology assessments and propagation of identical phylogenetic signals may be introduced into the dataset. The evolutionary processes of duplication and deletion of repeats and different degrees of subsequent homogenization of copies by gene conversion or crossing over may result in the presence of orthologues (Bhatia et al. 1996; Nickrent and Patrick 1998), paralogues (Fernández et al. 2000) or both (Goertzen et al. 2002).

We can conclude that the ETS and ITS share different features such as small size, high sequence variation, conserved flanking regions (at least at lower taxonomic levels), rapid concerted evolution, and evolution under similar functional constraints. These properties, together with the availability of technical protocols, which allow for the development of sequencing primers for the ETS, promote a more universal use of this region in phylogenetic and evolutionary studies, even in still-unexplored taxonomic groups.

Simple-sequence repeats (SSRs)

Simple-sequence repeats (SSRs) or microsatellites have been used as a marker system in population genetics for nearly two decades with a large popularity due to their attributes of codominance and high polymorphism (Tautz 1989). They have been widely used to study the genetic relationships between genotypes at the intraspecific level but are rarely applied in phylogenetic inference. The major reasons for that are: Firstly, the properties of SSRs, such as a constraint on allele size range (Garza and Freimer 1996; Goldstein and Pollock 1997), high mutation rates and size homoplasy (Bruford and Wyne 1993) may hamper the reconstruction of phylogeny (Ochieng et al. 2007). Secondly, the lack of trans-specifity of SSRs and the low level of SSR conservation among taxa restrict its application to intra-specific level. Thirdly, for a well-resolved phylogenetic analysis between species or even genera, large numbers of different SSRs are needed (Takezaki and Nei 1996). In the phylogenetic studies of plants using microsatellites to date, only a small number of loci, in general less than ten, have been used.

Only when these problems are solved SSRs can be applied in phylogenetic analysis. In the following we give a brief overview of the various methods of microsatellite isolation, then we review SSR application in plant phylogenies and discuss several important issues concerning phylogenetic analysis using SSRs. We put forward the problems facing phylogenetic analysis using SSRs. At the end, we give a prospect for the application of SSRs in plant phylogenetic studies.

Increasing number of methods for microsatellite isolation

Traditional construction of genomic libraries is an intensive task and a costly procedure that often results in a low isolated microsatellite density. Kandpal et al. (1994) developed a method to construct microsatellite-enriched libraries. Compared to the traditional isolation methods, this approach can produce highly enriched SSR libraries and makes hybridization an efficient way to isolate microsatellites. This technique, which was subsequently modified by Glenn and Schable (2005), is the most widely used method in microsatellite isolation. It has been successfully applied for many plant species such as red clover (Kölliker et al. 2006), cotton (Kumpatla et al. 2004), red oak (Aldrich et al. 2002), ryegrass (Jones et al. 2001) and potato (Ashkenazi et al. 2001). Fisher et al. (1996) developed the 5′-anchoring procedure, and some modified techniques appeared later on, such as sequence-tagged microsatellite profiling/STMP (Hayden and Sharp 2001; Hayden et al. 2002), selective amplification of microsatellite polymorphic loci/SAMPLE and selective amplification of microsatellites/SAM (Hayden and Sharp 2001). These techniques have been used for SSR isolation of, e.g. *Anthyllis vulneraria* (Van Glabeke et al. 2007) and *Litchi* (Li et al. 2006). Additionally, inter simple sequence repeat (ISSR) and random amplified polymorphic DNA (RAPD) were also used for the isolation of microsatellites (Wu et al. 1994, Lian et al. 2001). Dual-suppression polymerase chain reaction (PCR) technique (Lian and Hogetsu 2002; Islam et al. 2004) is an alternative method, which has been successfully used in *Ginkgo biloba* (Yan et al. 2006), *Bruguiera gymnorrhiza* (Islam et al. 2006) and

Robinia pseudoacacia (Lian and Hogetsu 2002). Publicly available databases such as EMBL, GenBank, and DDBJ can be used for the search of SSR sequences in order to develop primers to create microsatellite libraries.

Plant phylogenetic studies utilizing SSRs

Plant phylogenetic studies using microsatellites are mostly based on the repeat numbers of microsatellite motifs as well as the sequences from either the repeated regions or flanking regions of the microsatellite motif. The repeat numbers of microsatellite motifs are one of the characteristics of SSRs that can be used for plant phylogenetic analysis. Usually high levels of polymorphism and a large number of unique alleles for the accessions were found using microsatellite repeat numbers (Álvarez et al. 2001; Zwettler et al. 2002; Tang and Knapp 2003). Although several phylogenetic analyses are based on the different numbers of SSR repeat units, sequenced SSRs are more frequently used in phylogenies. The length variations of those sequences are both results from insertions, deletions and SNPs within the SSR motif/repeat region, but also the sequences that are flanking the repeat regions. To what extent length variation in those two different regions of a microsatellite contributes to the SSRs' total length variation depends on the marker itself and the investigated species.

Several examples show how both the highly divergent flanking regions and core regions of SSRs have led to a better understanding of close relationships between genera, species or even hybrids. For example, taxonomic relationships between eight species out of four genera of Vitaceae were resolved (Rossetto et al. 2002) with three *Vitis vinifera* SSRs, especially because of the highly variable flanking regions with high substitution and indel rates. A combined analysis of ITS1 and *trn*L intron failed to resolve the relationships in this study. However, as the authors emphasize, "the resolving power of this study is clearly limited by the number of taxa used". For this study, it must be stressed that the SSRs were located in the coding region of the genome. Chen et al. (2002) reported that sequence polymorphism was mainly found in the flanking regions. In this work, 11 SSRs from *Oryza* were used to investigate sequence divergence in rice (*O. sativa*) and related *Oryza* species and cultivars. Here the inner SSR repeat motif showed most of the sequence variation (80%). Alternatively, Kutil and Williams (2001) showed in their comparative study of eight *Pinus* species with 15 SSRs from cpDNA that no variation at all occurred in the flanking regions. The SSR variation between the different species was concentrated on the SSR repeat motif in the form of length variation originating from repeat gains and/or losses.

Problems in the reconstruction of phylogenies using SSRs

When using microsatellites for phylogenetic reconstructions, attention should be paid to the following aspects: (1) Establishing the SSR bands homologous; (2) Choosing the appropriate genetic distance measures; (3) Overcoming the problem of low transferability among the taxa.

Establishment of the homology of microsatellite bands

Though the co-migrating fragments in microsatellite analyses are assumed to be homologous, non-homology does exist. The inclusion of non-homologous fragments in an analysis is likely to bias the results and break the assumptions of a phylogenetic analysis. The comparison of paralogous SSR alleles will produce an incorrect organismal phylogeny.

The homology problem occurs both in the SSR flanking regions and the repeat units. If two state identical fragments co-migrate, either of them could produce homoplasy. Homology depends on the mutation rate of the SSR motif and flanking regions: The higher the mutation rate, the higher the probability of co-migrating of two non-homologous fragments. The incidence of non-homology can bias similarity, frequency measures and character based measures (Swofford et al. 1997). As the number and percentage of homoplasious characters increases, the likelihood of errors in the resultant phenetic and phylogenetic trees increases as well.

The percentage of non-homologous alleles resulting from indels in flanking regions, for example, can be tested for non-homology by sequencing. Kutil and Williams (2001) found persistent microsatellites with highly conserved flanking regions, a conserved repeat motif composition and variable unit numbers in pine cpDNA, and suggested sequencing orthologous microsatellites will be essential for making correct inferences about phylogenetic relationships. The polymorphic alleles caused by the presence of indels within the flanking regions were scored as identical when examined by denaturing polyacrylamide gel electrophoresis. The comparison of the number of sequence- and size-based alleles could help to find out homoplasies. Mogg et al. (2002) found approximately 10% of maize SSRs showing allele-size homoplasy.

There are different measurements to test homoplasy, such as successive approximation (Farris 1969), compatibility analysis (Meacham and Estabrook 1985) and the optimization method of Goloboff (1993). Ochieng (unpublished data) developed a method to test homoplasy. This method involves sequential character exclusion to create polytomies, followed by reinstatement in a stepwise manner and observation of changes in tree topology. For

soft polytomies that occurred farther in the past, homoplasy may overwrite a phylogenetic signal, such that each gene tree effectively becomes a hard polytomy with internal branch lengths that do not differ significantly from zero in a statistical framework.

Genetic distance measures for microsatellites

Classic genetic distances such as *Fst* (transformed for linearity with time) are based on the variance in allele frequencies (VAF) and are compatible with the infinite allele mutation model (IAM; Kimura and Crow 1964). Several genetic distances that make different assumptions on mutations have been developed for microsatellites. However, the appropriateness of each of these distance methods will vary from case to case, depending on the model of microsatellite evolution, mutation rates, effective population size, and time since divergence (Ochieng et al. 2007). The ideal distance measure will therefore depend on the characteristics of the SSRs and the phylogenetic question being addressed. One of these genetic metrics for microsatellites is *Rst* (Slatkin 1995), which uses the variance in repeat numbers (VRN) and is compatible with the stepwise mutation model (SMM; Kimura and Ohta 1978). Homoplasy is expected under SMM, which assumes loss or gain, with equal probability, of a single repeat unit through mutation, while the IAM expects no homoplasy.

Transferability of microsatellites among taxa

In comparison to a large amount of animal species, only a few SSRs are conserved in vascular plant species (Van Treuren et al. 1997; Steinkellner et al. 1997). In the animal field, successful microsatellite amplification has been possible even within very distantly related taxa (Rico et al. 1996), while in plants many phylogenetic studies utilizing microsatellites have been restricted to intra-specific relationships (Goldstein et al. 1999) or the use of the SSR flanking sequences in higher order phylogenies (Streelman et al. 1998; Zhu et al. 2000). The majority of phylogenies based on microsatellites derived from non-coding nuclear DNA were mostly performed within closely related crops and other cultivated species, e.g. comparisons between cultivated species and their wild relatives, or even within inbred lines. Jarne and Lagoda (1996) suggested that microsatellites are better used for groups separated by no more than a few thousand generations. However, since an increasing amount of sequence information from various plant genomes will arise in the future, including non-agricultural species, the problem of trans-specifity of SSR primers will be gradually reduced, and a priori knowledge of genome organization will improve finding trans-specific microsatellites in plants as well. Studies of transferability in *Actinidia* (Huang et al. 1998), *Prunus* (Dirlewanger et al. 2002), *Lolium* (Studer et al. 2006), and species of the A, B and D constituent genomes of *Triticum aestivum* (Sourdille et al. 2001) are promising, as they detected considerable proportions of cross-amplification between species. Nevertheless, the higher mutation rates within microsatellite regions compared to other non-coding nuclear markers, like ITS, will make it difficult to get good resolutions between more distantly related species.

What about AFLPs, RAPDs and ISSRs?

PCR-based techniques such as AFLPs, RAPDs or ISSRs have been employed in the plant kingdom mostly in population genetics. However, they have also been used for lower level phylogenetic reconstructions, where they cause the same problems as SSRs in phylogenetic studies, such as homology. Additionally these prevalent dominant markers raise other analytical issues not encountered with the co-dominant microsatellite markers. Vos et al. (1995) described the AFLP technique as a new fingerprinting method, which has many advantages, such as no prior genomic information needed and large number of polymorphism produced. However, it is a dominant marker with the problem of homoplasy. Additionally non-homology and non-independence of AFLP data will seriously misestimate similarity and distance, which need to be overcome with extensive testing. And other properties of AFLPs like scoring, bias introduced by dominance, reproducibility and the effect of polyploids will to some extent restrict its application in phylogeny.

Prospects on SSR application in plant phylogenetic analysis

The benefit of microsatellites for phylogenies is extremely small, mainly limited by homologous fragments, difficulties in finding the correct genetic distance measure and missing trans-specifity of primers between species. A bit more frequently used in phylogenetic analyses are cpDNA- and EST-derived microsatellites, which we did not introduce in this review of non-coding nuclear markers. However, microsatellites are an excellent tool to unravel patterns and processes within populations and therefore are essential in population genetics.

Tranposable elements (TE)

Mobile genetic elements that are able to change their position within the genome were first discovered by Barbara McClintock in maize almost 60 years ago (McClintock 1948). Since then, a considerable amount of such mobile

Table 2 Classification of transposable elements

Transposable elements		
Class I transposons	Long terminal repeat retrotransposons (LTR)	Ty1-copia
		Ty3-gypsy
	Non-long terminal repeat retrotransposons	Long interspersed elements (LINEs)
		Short interspersed elements (SINEs)
Class II transposons	Miniature inverted-repeat transposable elements (MITEs)	
	Others	

DNA elements has been detected and characterized. Currently, it seems quite clear that transposons are ubiquitous constituents of eukaryotic genomes. Thus, they have been found and studied throughout the plant kingdom (e.g. Flavell et al. 1992; Voytas et al. 1992; Bureau and Wessler 1994; Suoniemi et al. 1998; Kumar and Bennetzen 1999; Noma et al. 1999; Fedoroff 2000; Feschotte et al. 2002). Transposable elements have been classified into different types according to the mechanism of transposition (reviews in Grzebelus 2006; Wessler 2006; Ray 2007), namely retrotransposons (propagation via a RNA intermediate) and class II transposons (propagation via a DNA intermediate), which are further subdivided in several main groups (Table 2). There have also been attempts to achieve a phylogenetic classification of plant TE based on the sequence analysis of conserved regions (Hansen and Heslop-Harrison 2004).

Few attempts have been made to use TE as molecular markers for phylogenetic analyses at high taxonomic level in plants. On the contrary, a considerable amount of works have used TE for the study of different aspects of plant population genetics, including genome structure and evolution, genetic mapping and the assessment of genetic diversity and gene flow. These topics have been generally addressed for important crop species and their closest wild relatives (review in Grzebelus 2006; Ray 2007). Conversely, in animals, phylogenetic reconstructions based in mobile elements have been much more widely developed for a wide array of taxa, especially primates, while population-level analyses are still scarce outside humans (review in Ray 2007).

Transposable elements are increasingly being regarded as an under-utilized set of tools for researchers, which have great potential for investigating aspects of molecular ecology, such as genetic diversity, speciation, population and conservation genetics, and phylogeography (Ray 2007). Several molecular techniques have been designed to detect and assess DNA variation (DNA fingerprinting), which use PCR primers to amplify TE (Table 3; review in Schulman et al. 2004; Grzebelus 2006). These methods have been applied in several studies involving different plant species, most of them of agricultural interest [maize, rice, barley, rye, pea, sweet potato, cashew, banana, sugarcane and others (see Table 3 for references)].

Although there is a high number of techniques (Table 3), we will only address the most widely used ones here. Detailed information can be found in several specialized reviews (e.g. Kumar et al. 1997; Schulman et al. 2004; Grzebelus 2006; Ray 2007). The most popular transposon-based molecular marker system at present is termed S-SAP (Sequence-specific Amplification Polymorphism). It uses long terminal repeat (LTR) retrotransposon-specific primers in combination with AFLP primers. Numerous studies (Table 3) found S-SAP markers as efficient as or even more efficient than the well-known AFLP or RAPD techniques, as a result of being frequently more polymorphic and informative than those other markers. As a result, several authors have supported its usefulness in different applications of evolutionary biology (e.g. Ellis et al. 1998; Pearce et al. 2000; Vershinin et al. 2003; Schulman et al. 2004; Syed et al. 2005). However, although most fingerprinting techniques involving TE are applicable to a wide array of plant taxa, none of them has become as popular as RAPD, microsatellites and AFLPs. Nonetheless, the high level of polymorphism of these markers encourages their use for DNA fingerprinting purposes, but hinders their utility for phylogenetic reconstructions at the supra-specific level.

As for any transposon-based DNA profiling technique, the main disadvantage of the S-SAP approach is the need of sequence information to design transposon specific-primers. However, rapid transposon isolation methods based on PCR with adapter primers have been designed (Pearce et al. 1999). In addition, Wheelan et al. (2006) have recently designed an experimental method to identify all TE in a sample, called Transposon Insertion site Profiling chip (TIP-chip). Apart from its high variability, other advantages of the use of S-SAP markers include its even distribution across genetic maps, which contrasts with the behaviour of AFLP markers, often clustered in certain

Table 3 Main molecular marker techniques involving transposable elements applied to plant taxa

Molecular technique	Transposable element involved	Plant taxa studied	References
Alu-PCR	SINEs	*Musa*	Baurens et al. (1998)
		Saccharum	Alix et al. (1999)
Copia-SSR	LTR	*Hordeum*	Provan et al. (1999)
Hbr display or MITE-AFLP	MITEs	*Zea*	Casa et al. (2000, 2002, 2004) and Ju et al. (2004)
		Oryza	Park et al. (2003)
IMP	MITEs	*Hordeum*	Chang et al. (2001)
IRAP/ REMAP	LTR	*Aegilops*	Boyko et al. (2002)
		Hordeum	Leigh et al. (2003), Kalendar et al. (1999, 2000) and Brik et al. (2006)
		Malus	Antonius-Klemola et al. (2006)
		Musa	Chee et al. (2005)
LTR restriction site polymorphism	LTR	*Gossypium*	Hafez et al. (2006)
RBIP	LTR	*Pisum*	Flavell et al. (1998)
		Oryza	Vitte et al. (2004)
S-SAP	LTR	*Aegilops*	Nagy et al. (2006)
		Anacardium	Syed et al. (2005)
		Avena	Yu and Wise (2000)
		Cynara	Acquadro et al. (2006) and Lanteri et al. (2006)
		Elaeis, Cocos	Price et al. (2003)
		Hordeum	Leigh et al. (2003), Waugh et al. (1997), Rodriguez et al. (2006) and Soleimani et al. (2005, 2007)
		Ipomoea	Berenyi et al. (2002)
		Iris	Bouck et al. (2005) and Kentner et al. (2003)
		Malus	Venturi et al. (2006)
		Medicago	Porceddu et al. (2002)
		Pisum	Ellis et al. (1998), Pearce et al. (1999, 2000), Vershinin et al. (2003)
		Secale	Nagy and Lelley (2003)
		Triticum	Queen et al. (2004)
SINE insertion polymorphism	SINEs	*Brassica*	Tatout et al. (1999)
		Oryza	Cheng et al. (2002, 2003), Motohashi et al. (1997), Ohtsubo et al. (2004), and Xu et al. (2005, 2007)
SINE-SSAP	SINEs	*Brassica, Raphanus*	Prieto et al. (2005)
Transposon display	Class II transposon	*Daucus*	Grzebelus et al. (2007)
		Lotus	Holligan et al. (2006)
		Oryza	Kown et al. (2006)
		Petunia	Van den Broeck (1998)
Transposon signatures (CAPS)	LTR	*Zea*	Purugganan and Wessler (1995)
		Hibiscus	Lee et al. (2002b)

genomic areas. Transposable elements also show a remarkable preference to integrate into gene-rich regions, which facilitate the isolation of genes by transposon tagging (Gierl and Saedler 1992; Hanley et al. 2000). Finally, S-SAP markers can be used to monitor the transpositional activity of the element in an evolutionary timescale. There are some variants of the S-SAP technique (Table 3), depending of the specifity of the TE primer. Thus, AFLP

primers have also been used together with primers specific for class II transposons [e.g. van den Broeck et al. 1998 (*Petunia*); Holligan et al. 2006 (*Lotus*)], including the MITE superfamily, in a technique called Hbr display or MITE-AFLP [Casa et al. 2000 (*Zea*); Park et al. 2003 (*Oryza*)].

Other transposon-based PCR methods use LTR retrotransposon-specific primers in combination with microsatellite-specific primers. Different versions of this technique have been called copia-SSR (copia-Simple Sequence Repeat; Provan et al. 1999) and REMAP (Retrotransposon-Microsatellite Amplified Polymorphism; Kalendar et al. 1999). Specific primers have also been used to amplify the intervening sequence between LTR and MITEs elements, in the techniques termed IRAP (Inter-Retrotransposon Amplified Polymorphism; Kalendar et al. 1999) and IMPs (InterMITE polymorphisms; Chang et al. 2001), respectively. These markers were considerably more polymorphic than ISSR, and suitable for DNA fingerprinting (Kalendar et al. 1999, 2000; Manninen et al. 2000; Chang et al. 2001; Leigh et al. 2003; Antonius-Klemola et al. 2006).

In general, fingerprinting techniques involving TE generate multilocus banding patterns, which are often highly variable and generally behave as dominant markers [except for the co-dominant CAPS (Cleavage Amplified Polymorphic Sequences; Purugganan and Wessler 1995) and RBIP (Retrotransposon Based Insertion Polymorphism; Flavell et al. 1998)]. In addition, TE insertion based markers involve a more specific kind of genetic variability, i.e. transpositions events, than arbitrary markers systems such as RAPD or AFLPs, which detect polymorphism which may be of very different nature, from simple nucleotidic changes to extensive genomic rearrangements. Therefore, a more structured dataset could be expected to be generated from TE markers (Grzebelus et al. 2007). In animals, one of the most utilized groups of TE for population genetics and phylogenetic analyses has been retrotransposons, especially Short Interspersed Elements (SINEs). These elements seem to present low levels of homoplasy (Shedlock and Okada 2000; Okada et al. 2004; Shedlock et al. 2004; Ray et al. 2006; Deragon and Zhang 2006), and can be easily genotyped and analyzed, yielding very good resolution of evolutionary relationships and solving problematic phylogenies in several animal taxa (review in Ray 2007). On the other hand, in the plant kingdom, SINEs have been found in many families, but they have been less frequently developed as molecular markers. SINEs have been applied successfully to the study of the phylogenetic relationships between wild and cultivated *Brassica* and *Oryza* species, basing on the presence-absence of SINEs in certain loci (Table 3; review in Deragon and Zhang 2006). However, there are some processes that may potentially hinder phylogenetic inference based on SINEs, like lineage sorting, parallel insertion, precise excision and paralogous insertion (Hillis 1999; Ray et al. 2006).

As already mentioned, most studies dealing with phylogenetic reconstructions in plants based on TE have mainly tried to shed light on the origin, genome evolution and relationships between important crop species and their closest wild relatives (Table 3). It is important to note that these plants are usually submitted to strong artificial selection, so TE evolution may be different than in wild species living in natural conditions (Park et al. 2007). However, only a few wild plant groups have been studied in this regard, including some gymnosperms (Friesen et al. 2001; Stuart-Rogers and Flavell 2001) and Orobanchaceae (Park et al. 2007). The main purpose of these works was to analyse the diversity and evolution of different kind of retrotransposons (LTR; Table 2). Friesen et al. (2001) found no significant phylogenetic separation between very distant plant groups (angiosperms, gymnosperms, ferns), which they attributed to a very early radiation of retrotransposons during evolution, to horizontal transfer or to convergent evolution. In general, TE generate complex patterns due to a high level of polymorphism, which usually hinders phylogenetic inference. TE could also be useful for many other applications in evolutionary biology like hybridization and speciation research, as claimed by Kentner et al. (2003) and Bouck et al. (2005), who applied the technique S-SAP to produce genetic maps of two closely related wild species of *Iris* and used them successfully to study the genetic incompatibility and specific barriers between them.

In summary, although the use of TE as molecular markers is not yet as extended as that of other non-coding nuclear regions, it seems that they have the potential to provide a new source of nuclear DNA variation in plants, specially suitable for fingerprinting purposes or for exploring phylogenetic relationships between closely related species.

Nuclear intron sequences

Introns were first defined as sequences of non-coding DNA interspersed between coding regions (exons) of eukaryotic genes. Intron sequences are transcribed but they are removed before translation of the messenger RNA (mRNA) by a process known as splicing. Nowadays, introns have been found in all kingdoms, however, the extent and types of splicing can be very different between the major divisions. The splicing mechanisms can occur autonomously (self-splicing), with the help of enzymatic activities or with the so-called spliceosome. There are four

major classes of introns: self-splicing group I and group II introns, tRNA introns that splice by a mechanism different from that of spliceosomal introns, and spliceosomal introns. Group I introns are widespread and present in mRNA, tRNA and rRNA in mitochondria, chloroplasts and nuclear genomes of non-vertebrates, in bacteriophages and in eubacterial genomes. Group II introns are found in mitochondria and chloroplasts, but have also been observed in cyanobacteria and proteobacteria. tRNA and spliceosomal introns are confined to the nuclear genome of all eukaryotes (Rodríguez-Trelles et al. 2006).

Intron sequences can also be used as marker systems in phylogenetic studies, although they may require additional work compared to other markers, e.g. cloning the PCR product when alleles of different length (for suggestions on how to select nuclear marker sequences see Hughes et al. 2006; for an overview on theoretical concerns and practical issues see the review by Sang 2002; for primer selection see Strand et al. 1997). Nevertheless, there are several maker systems that seem to be rather popular and widely applicable since they were successfully used in a variety of plant families. The following summary is condensed in Table 4.

Antirrhinum majus FLORICAULA gene (FLO, Coen et al. 1990) or its *Arabidopsis thaliana* ortholog *LEAFY* (*LFY*, Weigel et al. 1992), encode a transcription factor that regulates floral meristem identity. *FLO/FLY* is usually single copy in angiosperms. However, in members of the Detarieae (Leguminosae) it was suggested that a second copy is present, which is probably a pseudogene (Archambault and Bruneau 2004). In non-angiosperms, like Isoetaceae, there are evidences for a second copy of this gene (Hoot et al. 2004).

FLO/LFY 2nd intron is the most variable intron within the gene. It was successfully applied for dating dispersal and radiation of the gymnosperms Gnetales (Won and Renner 2006), and determining the relationships within the Nellieae of the Rosaceae (Oh and Potter 2005). However, in some cases the *FLO/LFY* second intron turned out to be so variable that could not be aligned (Iochrominae, Solanaceae, de Witt Smith and Baum 2006). The first intron of *FLO/LFY* was found to be a variable marker in the orchid genus *Ophrys* (Schlüter et al. 2007).

Another marker system is *PISTILLATA* (*PI*) that encode a class B MADS-box transcription factor implicated in specifying petals and stamen in *Arabidopsis* (Krizek and Meyerowitz 1996). It was first introduced by Bailey and Doyle (1999) as a potential marker system. The *PI* locus in *Arabidopsis* has a length of 2.3 kb and includes several introns and exons of variable length. The first intron, which was subsequently used for phylogenetic analyses, is the largest one and has a length of about 1 kb. Lee et al. (2002a) sequenced the *PI* first intron for a phylogenetic study of *Lepidium*. They found 11.5% sequence divergence in the ingroup and most of the groups that they recovered were the same as in phylogenies based on ITS and cpDNA markers. However, the backbone species were different in the *PI* based tree. Bailey et al. (2002) used *PI* together with ITS and cpDNA markers to elucidate the phylogenetic relationships within the Halimolobine Brassicaceae, a group of North and Central American Brassicaceae genera. In their study, they found incongruence between cpDNA/*ITS* and *PI* based trees that did not fit with taxon phylogeny. However, the inclusion of the *PI* in a combined dataset seemed to stabilize the phylogeny in terms of bootstrap support. In addition to Brassicaceae, *PI* introns were applied in *Impatiens* (Balsaminaceae) for phylogenetic reconstruction (Janssens et al. 2007). Janssens et al. (2007) tested the phylogenetic utility of *PI* fourth and fifth introns of the two paralogues found in *Impatiens*. Topologies of the two obtained trees were found to be highly congruent with previously attained phylogenetic reconstructions based on chloroplast DNA data. Combination of the data obtained for *PI* fourth and fifth introns with chloroplast data resulted in a well-supported tree. This is consistent with the results from Bailey et al. (2002) where combination of both datasets also lead to a better bootstrap support.

Other markers that were infrequently but successfully used were *ncpGS* (nuclear plastid-expressed glutamine synthase, introns 7–10), applied in the Sinningieae (Gesneriaceae) (Perret et al. 2003), PEPC (phosphoenolpyruvat carboxylase, fourth intron), used in *Moringa* (Moringaceae) (Olson 2002) and *adh* (alcohol dehydrogenase), successfully applied in *Gossypium* giving a fully resolved tree (Small et al. 1998). GAP-DH gave also well-resolved and well-supported trees that were nearly free of homoplasy and showed no long branch attraction in the Olisbeoideae (Gesneriaceae) and phylogenetic relationships between *Manihot esculenta* and relatives (Euphorbiaceae) (Olsen and Schaal 1999).

Work with nuclear single copy gene introns is not as straight-forward as the work with the traditionally used ITS. However, problems associated with the use of of the above mentioned markers were easy to overcome. In some cases cloning was necessary, but the general result of the studies was satisfying. Nevertheless, this would represent a wrong picture of the applicability of single copy gene introns. Some markers have indeed caused big problems, like low mutation rates, high homoplasy levels and the presence of paralogues (Hughes et al. 2006). *MS* and *PRK* were used in the Areceae (Arecaceae) (Lewis and Doyle 2002). *MS* was shown to be of low utility and *PRK* was only applicable at higher levels and it was present in paralogues (Lewis et al. 2002). *sopdh* (sorbitol-6-phosphate-dehydrogenase) was tried in *Prunus* (Rosaceae), but it was found to have as many parsimony informative sites than

Table 4 Overview of intron sequences used in various phylogenetic reconstructions (for details refer to text)

Marker	Which intron was used?	Where applied	Average length (bp)	Cloning necessary?	Informative characters	Comments
FLO/LFY	2nd; 1st intron in Orchidaceae	Isoetaceae (Hoot et al. 2004) Rosaceae (Oh and Potter 2005) Solanaceae (De Witt Smith and Baum 2006) Platanaceae (Feng et al. 2005) Gnetales (Won and Renner 2006) Orchidaceae (Schlüter et al. 2007)	800–1,000; 3,000 for first intron	For some	Up to 187	Limited amount of homoplasy; in *Ophrys* (Orchidaceae) many indels
PISTILLATA	1st; also 4th and 5th	Brassicaceae (Bailey et al. 2002; Lee et al. 2002a)	600–1,000	For some	517 (of 1,400 aligned characters)	Backbone species different than for ITS
GAP-DH	Introns 1 to 4; exons included in analysis	Euphorbiaceae (Olsen and Schaal 1999)	1,000	Yes	64	Well resolved, well supported, nearly free of homoplasy
s6pdh	Introns 2 to 5	Rosaceae (Bortiri et al. 2002)	1,200–1,300	For some	390	More informative characters in s6pdh than in ITS; in general topology of ITS tree
MS	2nd	Arecaceae (Lewis and Doyle 2002)		For some		Low resolution
PRK	3rd, 4th	Arecaceae (Lewis and Doyle 2002)		For some		More suitable for higher relationships
adh	*adhC;* introns between exons 2 and 9	*Gossypium* (Small et al. 1998)	1600	Yes	25	Tree fully resolved
CHS		Elaeagnaceae (Bartish et al. 2006)		For some		No clear association of CHS lineages with geography
CYCLOIDEA		Gesneriaceae (Wang et al. 2004)	400–700	For some		High levels of homoplasy, paralogues!
ncpGS	7 to 10	Gesneriaceae (Perret et al. 2003) Oxalidaceae (Emshwiller and Doyle 1998)	360–400	For some		Higher polymorphism than cpDNA markers; good for generic and species level; variability similar to ITS
PEPC	4th	Moringaceae (Olson 2002)	500		39	Less informative characters in PEPC than in ITS
PgiC	In total 20 introns; here introns between exons 11 (15) to 21 used	Compositae (Ford et al. 2006)	20 introns ranging from <100 to 680		580 in the introns in the region from exon 11 to exon 21	Higher resolution than ITS
NIA or *NR*	3rd intron	Goodeniaceae (Howarth and Baum 2002)	1,200–1,500	Yes	33	Up to 5.4-fold more variable than ITS

ITS, and it was not possible to apply a molecular clock as different evolutionary rates occurred in the dataset (Bortiri et al. 2002). *CHS* (chalcone synthase) was used in a recent study on *Hippophae rhamnoides* to investigate its quaternary history (Bartish et al. 2006). The authors found no clear association of *CHS* lineages with geography, which they suspect might be due to the different transmission of *CHS* genes compared to cpDNA. *CYCLOIDEA* was tried as a marker system in *Titanotrichum* (Gesneriaceae) and was found to have eight times higher mutation rate than the ITS, but also very high levels of homoplasy and paralogues that caused additional problems (Wang et al. 2004).

The studies presented here give insight into benefits, but also problems that come up with the use of nuclear gene introns and their use in phylogenetic studies. Single copy gene markers may cause severe problems when paralogues have to be identified or when it turns out that most of the homoplasy is present in the dataset. Sometimes these markers are not more variable than standard ones like the ITS. Another problem is the availability of primers. There are no standard primers for single copy gene introns available as opposed to the nuclear standard marker ITS or the chloroplast marker trnL and trnF. However, there are also advantages in the use of single copy gene introns. Some of them have indeed more informative characters and are not subjected to concerted evolution like the ITS. Also, they are biparentally transmitted unlike the maternally (in most cases in angiosperms) inherited cpDNA and they enable us to identify hybrids and their parents.

In the light of the increasing number of molecular phylogenetic studies and the failure of traditionally used marker systems due to characteristics such as concerted evolution for the ITS or only maternal inheritance for chloroplast DNA (in Angiosperms in most cases) raises the need for new marker systems. Single copy gene introns offer to us such a system. We still have to deal with the difficulties that arise with their use such as additional work in the laboratory (cloning, more PCR optimization), difficulties in alignments and differences in usability of a marker system between different plant taxa. However, it is worth the effort in order to increase the quality of phylogenetic reconstructions and our knowledge on intron evolution in the nuclear genome.

Promoter sequences

One of the most important classes of non-coding DNA are the promoter sequences located directly upstream of the coding genes. They enable the organism to react to changes in environmental conditions, mirrored in the cell's physiological status, by modulating the rate of transcriptional initiation. Typically, promoters consist of a set of core elements, which are necessary to facilitate transcription by providing binding sites for the RNA polymerase and essential transcription factors, and a variable number of regulatory elements (Smale and Kadonaga 2003). Those elements are commonly referred to as proximal and distal promoter regions. Regulatory elements consist of sequence motifs that work as binding sites for proteins that modulate transcription initiation (transcription factors). Those transcription factor-binding sites are usually arranged in clusters of 6–15, thus exhibiting modular organisation (Fickett and Wasserman 2000; Wray et al. 2003).

The promoters of higher plants are, like all eukaryotic promoters, extremely difficult to characterize since they are highly diverse (Latchman 1998; Courey 2001; Stone and Wray 2001). One of the few common traits of eukaryotic promoters is the presence of a Pribnow- or TATA-box within the first 50 basepairs of many, but not all, promoters. This site binds a TATA binding protein assisting the RNA polymerase in forming the transcriptional complex. Regulatory elements, on the other hand, can be found several kilobasepairs upstream of the transcription start, making it hard to determine the range of sequence that still has to be considered part of a given promoter. Determining the presence of transcription factor binding sites also is not a straightforward procedure, since their sequences are highly diverse (Qiu 2003; Rombauts et al. 2003; Pavesi et al. 2004; Wasserman and Sandelin 2004; Tompa et al. 2005). Modulation of transcriptional rates might well include processes where sequences are suboptimal for binding of transcription factors. Examination of several well-characterized eukaryotic promoters leads to the estimation that approximately 10–50 actual binding sites comprise about 10–20% of the total sequence range (Wray et al. 2003; Zhang and Gerstein 2003). The number and spacing of the binding sites are highly variable between species (Buchanan et al. 1997). The position and the orientation of a binding site are influenced among other factors by DNA bending, cooperation between different transcription factors and steric hindrance.

It seems obvious that any mutations in the promoter regions of any given genome can have dramatic effects on an organism, and it has been proposed that changes of promoter sequences have been of much higher importance for evolution than changes in coding sequences. Nevertheless, the fact that promoters can span several kilobasepairs, in which the rate of mutation is prone to high variation due to differences in evolutionary constraints between transcription factor binding sites and other parts of the sequence (Wasserman et al. 2000), often makes them difficult marker systems for phylogenetic analyses. As expected, details about the evolution of promoters and transcription factor binding sites are not very well known. It is reported that binding sites for one transcription factor in one species can

differ by 20–30% without losing functionality (Collado-Vides et al. 1991). The amount of variability of those sequences also seems to be rather independent of the functional importance of a given site for transcription of the respective gene. Instead variability seems more dependent of phylogenetic lineage (Dermitzakis and Clark 2002). In addition, promoter evolution is marked by frequent rearrangements of transcription factor binding sites, including inversions, making it difficult to align promoter sequences by standard procedures. These rearrangements have been reported to increase in number with increasing evolutionary distance in vertebrates (Chuzhanova et al. 2000). Although not much data about their occurrence is available, it stands to reason that this would be the case in plants as well.

A phenomenon particularly abundant in plants is gene duplication, which may implement new gene functions or expression patterns. Often the function of a given gene is repartitioned in the course of evolution into several copies that have different expression patterns, implying variation in their promoter regions (Ohno 1970; Prince and Pickett 2002). This plays an especially influential role when transcriptional regulators of developmental genes are affected, suggesting a key role in evolution and making these mechanisms a probable source of diversity in higher plants (Doebley and Lukens 1998; Tautz 2000; Kellogg 2004). Thus, it is a prerequisite for the use of promoter sequences in phylogenetic studies to compare the right set of sequences to determine their orthologous and paralogous relationships (Theissen 2002).

Any phylogenetic approach to promoter sequences has to take into account that no full-length colinearity between sequences can be assumed, that evolutionary constraints applied to different regions of the analyzed sequences will be very different and that the overall level of variability will be relatively high (Holland et al. 2001). Alignments of sequences including such cryptically conserved regions as transcription factor binding sites will in most cases have to be done by starting with a local algorithm considering both strands, and subsequent manual aligning of the interjacent regions as far as possible. In any case, the availability of reasonably well alignable sequences will be smaller than it is the case with other marker systems, and the selection of a set of sequences meaningful for a given problem will pose more of a challenge. Frequently used global alignment programs like CLUSTALW (Thompson et al. 1994) will not prove helpful under these conditions (Notredame 2002; Morrison 2006). Software packages that are better suited to provide starting points for these tasks include MACAW (Schuler et al. 1991; Lawrence et al. 1993) and DIALIGN (Morgenstern et al. 2006). MACAW employs a Gibbs Sampling Strategy to optimize motif-finding and involves the user in finally selecting statistically pre-evaluated conserved patterns. DIALIGN is capable of doing anchored alignments and relying on the user's choice in deciding which patterns to use. Another software package to mention in this context is, of course, BLAST (Altschul et al. 1990). But in comparison to the more conventional global alignments of completely colinear coding sequences frequently used in phylogenetics, there is no ideal system available at the moment to automate the finding of regions of local similarity in long sequences (Schuler et al. 1991). Problems complicating this task include the large size of the searchspace involved and the necessity of relying on basic parameters that are difficult to define beforehand, like the fragment length, the allowable fraction of substituted characters and the minimal fraction of sequences in which the pattern in question occurs.

The method most frequently applied for the determination of transcription factor binding sites within promoter regions is phylogenetic footprinting (Wasserman et al. 2000; Bulyk 2003; Weitzman 2003; Zhang and Gerstein 2003). This method is based on the assumption that regions conserved between orthologous regulatory sequences are of functional importance, leading to higher evolutionary constraints than in the surrounding sequences. Keeping in mind how variable those binding sites can be, it becomes obvious that only a certain window of divergence is available for such analyses (De Bodt et al. 2006). On the one hand the sequences in question have to be sufficiently divergent to allow distinguishing between functionally important and non-functional parts, but on the other hand conservation will become difficult to detect much more rapidly than in coding sequences. The method has been successfully applied to promoter sequences of animals and yeasts (Wasserman et al. 2000; Bulyk 2003; Weitzman 2003; Zhang and Gerstein 2003), and to several plant sequences such as grasses (Guo and Moose 2003; Inada et al. 2003) and dicots, for example, from the Brassicaceae, Rubiaceae and Cucurbitaceae (Koch et al. 2001; Hong et al. 2003; Ayre et al. 2003; Manen 2000). In Brassicaceae it was also shown that the phylogeny inferred from *Chs* and *Apetala3* promoter sequences closely mimicks the accepted evolutionary relationships obtained from the analysis of more conventional marker systems (Koch et al. 2001). It's rather unlikely that this approach will be applicable to analyses of more distantly related eudicot plant species (De Bodt et al. 2006). There are several software tools available that are capable of helping, to a certain extent, in the process of finding conserved regions within promoter sequences (e.g. MEME, PhyloGibbs, PhyME and others, Das and Dai 2007; Gertz et al. 2006). Although they are mainly intended for the purpose of finding transcription factor binding sites, they also might prove useful as a means of generating starting points for alignments that are built from sequence regions that would provide adequate phylogenetic signals.

Although there are phylogenetic applications where promoters can prove useful, they are not to be counted among the more widely and easily applicable marker systems. Sequence region and taxa must be chosen carefully, and in most cases a higher effort will be necessary when creating alignments. In this respect, phylogenetic analyses could benefit from efforts to allow routine identification of transcription factor binding sites by means of phylogenetic footprinting. This also would help locate regions in promoter sequences that are most likely to provide usable phylogenetic signals. Given choices of adequate regions, the amount of variability usually encountered in promoter sequences makes them promising targets for the field of phylogeny, especially in light of their evolutionary significance.

References

Albach DC, Chase MW (2004) Incongruence in Veroniceae (Plantaginaceae): evidence from two plastid and a nuclear ribosomal DNA region. Molec Phylogenet Evol 32:183–197

Acquadro A, Portis E, Moglia A, Magurno F, Lanteri S (2006) Retrotransposon-based S-SAP as a platform for the analysis of genetic variation and linkage in globe artichoke. Genome 49:1149–1159

Aldrich PR, Michler CH, Sun W, Romero-Severson J (2002) Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). Mol Ecol Notes 2:472–474

Alix K, Paulet F, Glaszmann JC, ĎHont A (1999) Inter-Alu-like species-specific sequences in the *Saccharum* complex. Theor Appl Genet 99:239–243

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Molec Biol 215:403–410

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molec Phylogenet Evol 29:417–434

Álvarez AE, van de Weil CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theor Appl Genet 103:1283–1292

Andreasen K, Baldwin BG (2003) Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (*Sidalcea*, Malvaceae). Molec Phylogenet Evol 29:563–581

Antonius-Klemola K, Kalendar R, Schulman AH (2006) TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. Theor Appl Genet 112:999–1008

Apprill AM, Gates RD (2007) Recognizing diversity in coral symbiotic dinoflagellate communities. Molec Ecol 16:1127–1134

Archambault A, Bruneau A (2004) Phylogenetic Utility of the *LEAFY/FLORICAULA* gene in the Caesalpinoideae (Leguminosae): gene duplication and novel insertion. Syst Bot 29:609–626

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn R (eds) Evolution of genes and proteins. Sinauer, Sunderland, pp. 38–61

Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001) Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome 44:50–62

Ayre BG, Blair JE, Turgeon R (2003) Functional and phylogenetic analyses of a conserved regulatory program in the phloem of minor veins. Pl Physiol 133:1229–1239

Bailey D, Doyle J (1999) Potential phylogenetic utility of the low-copy nuclear gene pistillata in Dicotyledonous plants: comparison to nrDNA ITS and trnL intron in *Sphaerocardamum* and other Brassicaceae. Molec Phylogenet Evol 13:20–30

Bailey D, Price RA, Doyle JJ (2002) Systematics of the Halimolobine Brassicaceae: evidence from three loci and morphology. Syst Bot 27:318–332

Bailey D, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molec Phylogenet Evol 29:435–455

Bailey D, Koch MA, Mayer M, Mummenhoff K, O'Kane Jr S, Warwick S, Windham M, Al-Shehbaz I (2006) Toward a global phylogeny of the Brassicaceae. Molec Biol Evol 23:2142–2160

Baldwin BG, Markos S (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: congruence of ETS and ITS trees of *Calycadenia*. Molec Phylogenet Evol 10:449–463

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. Ann Missouri Bot Gard 82:247–277

Barkman TJ, Simpson BB (2002) Hybrid origin and parentage of *Dendrochilum acuiferum* (Orchidaceae) inferred in a phylogenetic context using nuclear and plastid DNA sequence data. Syst Bot 27:209–220

Bartish IV, Kadereit JW, Comes HP (2006) Late Quaternary history of *Hippophaë rhamnoides* L. (Elaeagnaceae) inferred from chalcone synthase intron (Chsi) sequences and chloroplast DNA variation. Molec Ecol 15:4065–4083

Baurens FC, Noyer JL, Lanaud C, Lagoda PJL (1998) Inter-Alu PCR like genomic profiling in banana. Euphytica 99:84–91

Bayly MJ, Ladiges PA (2007) Divergent paralogues of ribosomal DNA in eucalypts (Myrtaceae). Molec Phylogenet Evol 44:346–356

Becerra JX (2003) Evolution of Mexican *Bursera* (Burseraceae) inferred from ITS, ETS, and 5S nuclear ribosomal DNA sequences. Molec Phylogenet Evol 26:300–309

Bena G, Jubier MF, Olivieri I, Lejeune B (1998) Ribosomal external and internal transcribed spacers: combined used in the phylogenetic analysis of *Medicago* (Leguminosae). J Molec Evol 46:299–306

Berenyi M, Gichucki ST, Schmidt J, Burg K (2002) Ty1-copia retrotransposons-based S-SAP for genetic analysis of sweetpotato. Theor Appl Genet 105:862–869

Bhatia S, Singh Negi M, Lakshmikumaran M (1996) Structural analysis of the rDNA intergenic spacer of *Brassica nigra*: evolutionary divergence of the spacers of the three diploid *Brassica* species. J Molec Evol 43:460–468

Bhattacharya D, Bera AK, Bera BC, Maity A, Das SK (2007) Genotypic characterisation of Indian cattle, buffalo and sheep isolates of *Echinococcus granulosus*. Vet Par 143:371–374

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W (2003) Noncoding plastidic *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. J Evol Biol 16:558–576

Bortiri E, Oh S-H, Gao F-Y, Potter D (2002) The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). Amer J Bot 89:1697–1708

Bouck A, Peeler R, Arnold ML, Wessler SR (2005) Genetic mapping of species boundaries in Lousiana irises using IRRE retrotransposon display markers. Genetics 171:1289–1303

Boyko E, Kalendar R, Korzun V, Fellers J, Korol A, Schulman AH, Gill BS (2002) A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating retrotransposons and defence-related genes: insights into cereal chromosome structure and fucntion. Pl Molec Biol 48:767–790

Brik AF, Kalendar RN, Stratula OR, Sivolap YM (2006) IRAP and REMAP analyses of barley varieties of Odessa breeding. Cytol Genet 40(3):24–33

Bruford MW, Wyne RK (1993) Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939–943

Buchanan KL, Smith EA, Dou S, Corcoran LM, Webb CF (1997) Family-specific differences in transcription efficiency of Ig heavy chain promoters. J Immunol 159:1247–1254

Buckler ESI, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145:821–832

Bulyk ML (2003) Computational prediction of transcription-factor binding site locations. Genome Biol 5:201

Bureau TE, Wessler SR (1994) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Pl Cell 6:907–916

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) The MITE family Heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Casa AM, Mitchell SE, Smith OS, Register JC III, Wessler SR, Kresovich S (2002) Evaluation of Hbr (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. Theor Appl Genet 104:104–110

Casa AM, Nagel A, Wessler SR (2004) MITE display. Methods Mol Biol 260:175–188

Chan R, Baldwin BG, Ornduff R (2002) Cryptic goldfields: a molecular phylogenetic reinvestigation of *Lasthenia californica* sensu lato and close relatives (Compositae: Heliantheae sensu lato). Amer J Bot 89:1103–1112

Chang RY, ÓDonoughue LS, Bureau TE (2001) Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. Theor Appl Genet 102:773–781

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgesensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M (2007) A proposal for a standardized protocol to barcode all land plants. Taxon 56:295–299

Chee HT, Siang HT, Chai LH, Faridah QZ, Othman YR, Heslop-Harrison JS, Kalendar R, Schulman AH (2005) Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. J Pl Biol 48(1):96–105

Chen X, Cho YG, McCouch SR (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Molec Genet Genom 268:331–343

Cheng C, Tsuchimoto S, Ohtsubo H, Ohtsubo E (2002) Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. Genes Genet Syst 77:323–334

Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice based on the interspersion pattern of SINEs. Molec Biol Evol 20:67–75

Chuzhanova NA, Krawczak M, Nemytikova LA, Gusev VD, Cooper DN (2000) Promoter shuffling has occurred during the evolution of the vertebrate growth hormone gene. Gene 254:9–18

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R (1990) Floricaula: a homeotic gene required for flower development in *Antirrhinum majus*. Cell 63:1311–1322

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 19:370–375

Collado-Vides J, Magasanik B, Gralla JD (1991) Control site location and transcriptional regulation in *Escherichia coli*. Microbiol Rev 55:371–394

Courey AJ (2001) Cooperativity in transcriptional control. Curr Biol 11:R250–252

Das MK, Dai HK (2007) A survey of DNA motif finding algorithms. BMC Bioinformatics 8(Suppl 7):S21

De Bodt S, Theissen G, Van de Peer Y (2006) Promoter analysis of MADS-box genes in Eudicots through phylogenetic footprinting. Molec Biol Evol 23:1293–1303

Deragon JM, Zhang X (2006) Short Interspersed Elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. Syst Biol 55(6):949–956

Dermitzakis ET, Clark AG (2002) Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Molec Biol Evol 19:1114–1121

Desquesnes M, Davila AMR (2002) Applications of PCR-based tools for detection and identification of animal trypanosomes: a review and perspectives. Vet Par 109:213–231

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

De Witt Smith S, Baum DA (2006) Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). Amer J Bot 93:1140–1153

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach (*Prunus persica* (L.) Batsch) and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet 105:127–138

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Pl Cell 10:1075–1082

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Quart Rev Biol 70:297–320

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Molec Gen Genet 260:9–19

Emshwiller E, Doyle JJ (1998) Origins of domestication and polyploidy in oca (*Oxalis tuberosa*: Oxalidaceae): nrDNA ITS data. Amer J Bot 85:975–985

Farris JS (1969) A successive approximations approach to character weighting. Syst Zool 18:374–385

Fedoroff N (2000) Transposons and genome evolution in plants. Proc Natl Acad Sci USA 97:7002–7007

Fehrer J, Gemeinholzer B, Chrtek J Jr, Bräutigam S (2007) Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). Molec Phylogenet Evol 42:347–361

Feliner GN, Aguilar JF, Rosselló JA (2001) Can extensive reticulation and concerted evolution result in a cladistically structured molecular data set? Cladistics 17:301–312

Feng Y, Oh S-H, Manos PS (2005) Phylogeny and historical biogeography of the genus *Platanus* as inferred from nuclear and chloroplast DNA. Syst Bot 30:786–799

Fernández M, Polcano C, Ruiz ML, Pérez de la Vega M (2000) A comparative study of the structure of the rDNA intergenic spacer of *Lens culinaris* Medik., and other legume species. Genome 43:597–603

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meet genomics. Nat Rev Genet 3:329–341

Fickett JW, Wasserman WW (2000) Discovery and modeling of transcriptional regulatory regions. Curr Opin Biotechnol 11:19–24

Fisher PJ, Gardner RC, Richardson TE (1996). Single locus microsatellites isolated using 5́ anchored PCR. Nucleic Acids Res 24:4369–4371

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Flavell AJ, Knox MR, Pearce SR, Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Pl J 16:643–650

Ford BA, Iranpour M, Naczi RFC, Starr JR, Jerome C (2006) Phylogeny of *Carex* subg. *Vignea* (Cyperaceae) based on non-coding nrDNA sequence data. Syst Bot 31:70–82

Friesen N, Brandes A, Heslp-Harrison JS (2001) Diversity, origin and distribution of retrotransposons (gypsy and copia) in conifers. Molec Biol Evol 18(7):1176–1188

Fuertes Aguilar J, Rosselló JA, Nieto Feliner G (1999) Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). Molec Ecol 8 (8):1341–1346

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6:211–217

Gertz J, Fay JC, Cohen BA (2006) Phylogeny based discovery of regulatory elements. BMC Bioinformatics 7:266

Gierl A, Saedler H (1992) Plant-transposable elements and gene tagging. Pl Molec Biol 19:39–49

Glenn TC, Schable NA (2005). Isolating microsatellite DNA loci. Methods Enzymol 395:202–222

Goertzen LR, Francisco-Ortega J, Santos-Guerra A, Mower JP, Linder CR, Jansen RK (2002) Molecular systematics of the *Asteriscus* alliance (Asteraceae: Inulae) II: combined nuclear and chloroplast data. Syst Bot 27:815–823

Goldstein DB, Pollock DD (1997) Launching Microsatellites: a review of mutation processes and methods of phylogenetic inference. Heredity 88:335–342

Goldstein DB, Roemer G, Smith D, Reich DE, Bergman A, Wayne R (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151:797–801

Goloboff PA (1993) Estimating character weights during tree search. Cladistics 9:83–91

Grzebelus D (2006) Transposon insertion polymorphism as a new source of molecular markers. J Fruit Ornamental Plant Res 14(Suppl 1):21–29

Grzebelus D, Jagosz B, Simon PW (2007) The DcMaster Transposon Display maps polymorphic insertion sites in the carrot (*Daucus carota* L.) genome. Gene 390:67–74

Guo H, Moose SP (2003) Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. Pl Cell 15:1143–1158

Hadjiolova KV, Georgiev OI, Nosikov VV, Hadjiolov AA (1984) Localization and structure of endonuclease cleavage sites involved in the processing of the rat 32S precursor to ribosomal RNA. Biochem J 220:105–116

Hadjiolova KV, Normann A, Cavaillé J, Soupène E, Mazan S, Hadjiolov AA, Bachellerie JP (1994) Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. Molec Cell Biol 14:4044–4056

Hafez E, Ghany AG, Zaki E (2006) LTR-retrotransposon-based molecular makers in cultivated Egyptian cottons *Gossypium barbadense* L. Afr J Biotechnol 5:1200–1204

Hanley S, Edwards D, Stevenson D, Haines S, Hegarty M, Schuch W, Edwards KJ (2000) Identification of transposon-tagged genes by the random sequencing of mutator-tagged DNA fragments from *Zea mays*. Pl J 22:557–566

Hansen CN, Heslop-Harrison JS (2004) Sequences and phylogenies of plant retroviruses, viruses and transposable elements. Adv Bot Res 41:165–193

Harpke D, Peterson A (2006) Non-concerted ITS evolution in *Mammillaria* (Cactaceae). Molec Phylogenet Evol 41:579–593

Hayden MJ, Sharp PJ (2001) Sequence-tagged microsatellite profiling (STMP): a rapid technique for developing SSR markers. Nucleic Acids Res 29:e43

Hayden MJ, Good G, Sharp PJ (2002) Sequence-tagged microsatellite profiling (STMP): improved isolation of DNA sequence flanking target SSRs. Nucleic Acids Res 30:e129

Hershkovitz MA, Lewis LA (1996) Deep-level diagnostic value of the rDNA-ITS region. Mol Biol Evol 13:1276–1295

Hillis D (1999) SINEs of the perfect character. Proc Natl Acad Sci 96:9979–9981

Hillis DM, Dixon MT (1991) Ribosomal DNA—molecular evolution and phylogenetic inference. Quart Rev Biol 66:410–453

Holligan D, Zhang X, Jiang N, Pritham EJ, Wessler SR (2006) The transposable element landscape of the model legume *Lotus japonicus*. Genetics 174:2215–2228

Hong RL, Hamaguchi L, Busch MA, Weigel D (2003) Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. Pl Cell 15:1296–1309

Holland JB, Holland SJ, Sharopova N, Rhyne DC (2001) Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. Genome 44:1065–1076

Hoot SB, Napier NS, Taylor WC (2004) Revealing unknown or extinct lineages within *Isoetes* (Isoetaceae) using DNA sequences from hybrids. Amer J Bot 91:899–904

Howarth DG, Baum DA (2002) Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. Molec Phylogenet Evol 23:525–528

Huang WG, Cipriani G, Morgante M, Testolin R (1998) Microsatellite DNA in *Actinidia chinensis*: isolation, characterisation, and homology in related species. Theor Appl Genet 97:1269–1278

Hughes CE, Eastwood RJ, Bailey CD (2006) From famine to feast? Selecting nuclear DNA sequence loci for plant species level phylogeny reconstruction. Philos Trans Roy Soc Lond B Biol Sci 361:211–225

Inada DC, Bashir A, Lee C, Thomas BC, Ko C, Goff SA, Freeling M (2003) Conserved noncoding sequences in the grasses. Gen Res 13:2030–2041

Islam MS, Lian CL, Kameyama N, Wu B, Hogetsu T (2004) Development of microsatellite markers in *Rhizophora stylosa* using a dual-suppression-polymerase chain reaction technique. Molec Ecol Notes 4:110–112

Islam MS, Lian CL, Kameyama N, Wu B, Hogetsu T (2006) Development and characterization of ten new microsatellite markers in a mangrove tree species *Bruguiera gymnorrhiza* (L.). Molec Ecol Notes 6:30–32

Janssens S, Geuten K, Viaene T, Yuan Y-M, Song Y, Smets E (2007) Phylogenetic utility of the AP3/DEF K-domain and its molecular evolution in impatiens (Balsamineae). Molec Phylogenet Evol 43:225–239

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 10:424–429

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Jousselin E, Rasplus J-Y, Kjellberg F (2003) Convergence and coevolution in a mutualism: Evidence from a molecular phylogeny of *Ficus*. Evolution 57:1255–1269

Ju KL, Jong YP, Se HC, Jin HK, Jae KC, Min H-K, Park C-H, Kim N-S (2004) Genetic mapping of maize with the intermated Mo17 × KW7 population using MITE-AFLP and SSR markers. Kor J Genet 26:63–72

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711

Kalendar R, Tanskanen J, Immonenn S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by Bare-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607

Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proc Natl Acad Sci USA 91:88–92

Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6:36

Kellogg EA (2004) Evolution of developmental traits. Curr Opin Pl Biol 7:92–98

Kentner EK, Arnold ML, Wessler SR (2003) Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana *Iris* species and their use as molecular markers. Genetics 164:685–697

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci 75:2868–2872

Koch MA, Al-Shehbaz IA, Mummenhoff K (2003a) Molecular systematics, evolution and population biology in the Mustard family, Brassicaceae: a review of a decade of studies. Ann Missouri Bot Gard 90:151–171

Koch MA, Dobeš C, Mitchell-Olds T (2003b) Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molec Biol Evol 20:338–350

Koch MA, Weisshaar B, Kroymann J, Haubold B, Mitchell-Olds T (2001) Comparative genomics and regulatory evolution: conservation and function of the Chs and Apetala3 promoters. Molec Biol Evol 18:1882–1891

Kölliker R, Enkerli J, Widmer F (2006) Characterization of novel microsatellite loci for red clover (*Trifolium pratense* L.) from enriched genomic libraries. Molec Ecol Notes 6:50–53

Kown SJ, Ju KL, Hong SW, Park YJ, McNally KL, Kim NS (2006) Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by Rim2/Hipa CACTA transposon display. Gen Genet Syst 81:93–101

Krizek BA, Meyerowitz EM (1996) The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development 122:11–22

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annual Rev Genet 33:479–532

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100:205–217

Kumpatla SP, Manley MK, Horne EC, Gupta M, Thompson SA (2004) An improved enrichment procedure to develop multiple repeat classes of cotton microsatellite markers. Pl Molec Biol Rep 22:85–86

Kutil BL, Williams CG (2001) Triplet-repeat microsatellites shared among hard and soft pines. J Heredity 92:327–332

Lalev AI, Nazar RN (1999) Structural equivalence in the transcribed spacer of pre-rRNA transcripts in *Schizosaccharomzces pombe*. Nucleic Acids Res 27:3071–3078

Lanteri S, Acquadro A, Comino C, Mauro R, Mauromicale G, Portis E (2006) A first linkage map of global artichoke (*Cynara cardunculus* var. *scolymus* L.) based on AFLP, S-SAP, M-AFLP and microsatellite markers. Theor Appl Genet 112:1532–1542

Latchman DS (1998) Transcription factors: an overview. Academic Press, San Diego

Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC (1993) Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science 262:208–214

Lee J-Y, Mummenhoff K, Bowman JL (2002a) Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L (Brassicaceae). Proc Natl Acad Sci USA 99:16835–16840

Lee SJ, Jeung JU, Cho SK, Um BY, Chung WI, Bae JM, Shin JS (2002b) Diversity and varietal classification of *Hibiscus syriacus* L. With the heterogeneity within retrotransposon-like elements. Mol Cell 13:362–368

Leigh F, Kalendar R, Lea V, Lee D, Donni P, Schulman AH (2003). Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Gen Genet 269:464–474

Lewis CE, Doyle JJ (2002) A phylogenetic analysis of tribe Areceae (Arecaceae) using two low-copy nuclear genes. Pl Syst Evol 236:1–17

Li D, Zhang X (2002) Physical localization of the 18S-5·8S-26S rDNA and sequence analysis of ITS Regions in *Thinopyrum ponticum* (Poaceae: Triticeae): implications for concerted evolution. Ann Bot 90:445–452

Li M, Zheng X, Zhu Y, Wang X, Liang S, Li L, Wu X (2006) Development and characterization of SSR markers in lychee (*Litchi chinensis*). Molec Ecol Notes 6:1205–1207

Lian C, Hogetsu T (2002) Development of microsatellite markers in black locust (*Robinia pseudoacacia*) using a dual-supression-PCR technique. Molec Ecol Notes 2:211–213

Lian C, Zhou Z, Hogetsu T (2001) A simple method for developing microsatellite markers using amplified fragments of inter simple sequence repeat (ISSR). J Pl Res 114:381–385

Linder CR, Rieseberg LH (2004) Reconstructing patterns of reticulate evolution in plants. Amer J Bot 91:1700–1708

Linder CR, Goertzen LR, Heuvel BV, Francisco-Ortega J, Jansen RK (2000) The complete external transcribed spacer of 18S–26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. Molec Phylogenet Evol 14:285–303

Liu JS, Schardl CL (1994) A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Pl Molec Biol 26:775–778

Mai JC, Coleman AW (1997) The Internal Transcribed Spacer 2 exhibits a common secondary structure in green algae and flowering plants. J Molec Evol 44:258–271

Manen JF (2000) Relaxation of evolutionary constraints in promoters of the plastid gene *atp*B in a particular Rubiaceae lineage. Pl Syst Evol 224:235–241

Manninen O, Kalendar R, Robinson J, Schulman AH (2000) Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol Gen Genet 264:325–334

McClintock B (1948) Mutable loci in maize. Carnegie Inst Wash Yearb 47:155–169

McIntyre CL, Clarke BC, Appels R (1988) DNA sequence analyses of the ribosomal spacer regions in the Triticeae. Pl Syst Evol 160:91–104

Meacham CA, Estabrook GF (1985) Compatibility methods in systematics. Annual Rev Ecol Syst 16:431–446

Mishima M, Ohmido N, Fukui K, Yahara T (2002) Trends in site number change of rDNA loci during polyploid evolution in *Sanguisorba* (Rosaceae). Chromosoma 110:550–558

Mogg R, Batley J, Hanley S, Edwards D, ÓSullivan H, Edwards KJ (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor Appl Genet 105:532–543

Morgenstern B, Prohaska SJ, Pöhler D, Stadler PF (2006) Multiple sequence alignment with user-defined anchor points. Alg Mol Biol 1:1748–7188

Morrison DA (2006) L. A. S. JOHNSON REVIEW No. 8. Multiple sequence alignment for phylogenetic purposes. Austral Syst Bot 19:479–539

Motohashi R, Mochizuki K, Ohtsubo H, Ohtsubo E (1997) Structures and distribution of pSINE1 members in rice genomes. Theor Appl Genet 95:359–368

Muir G, Fleming CC, Schlötterer C (2001) Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L Molec Biol Evol 18:112–119

Murakami A (2001) Structural differences in the intergenic spacer of 18S–26S rDNA and molecular phylogeny using partial external transcribed spacer sequence in hop, *Humulus lupulus*. Breed Sci 51:163–170

Nagy ED, Lelley T (2003) Genetic and physical mapping of sequence-specific amplified polymorphic (SSAP) markers on the 1RS chromosome arm of rye in a wheat background. Theor Appl Genet 107:1271–1277

Nagy ED, Molnár I, Schneider A, Kovács G, Molnár-Láng M (2006) Characterization of chromosome-specific S-SAP markers and their use in studying genetic diversity in *Aegilops* species. Genome 49:289–296

Nagylaki T, Petes TD (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. Genetics 100:315–337

Nickrent DL, Patrick JA (1998) The nuclear ribosomal DNA intergenic spacers of wild and cultivated soybean have low variation and cryptic subrepeats. Genome 41:183–192

Nieto Feliner G, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. Molec Phylogenet Evol 44:911–919

Nieto Feliner G, Larena BG, Aguilar JF (2004) Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). Ann Bot 93:189–200

Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol Gen Genet 261:71–79

Notredame C (2002) Recent progress in multiple sequence alignment: a survey. Pharmacogenomics 3:131–144

Noyes RD (2006) Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae). Amer J Bot 93:470–479

Ochieng JW, Muigai AWT, Ede GN (2007) Phylogenetics in plant biotechnology: principles, obstacles and opportunities for the resource poor. Afr J Biotechnol 66:639–649

Oh S-H, Potter D (2005) Molecular phylogenetic systematics and biogeography of tribe Neillieae (Rosaceae) using DNA sequences of cpDNA, rDNA, and LEAFY. Amer J Bot 92:179–192

Ohno S (1970) Evolution by gene duplication. Springer, Heidelberg

Ohta T, Dover GA (1983) Population genetics of multigene families that are dispersed into 2 or more chromosomes. Proc Natl Acad Sci USA 80:4079–4083

Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. Genetics 108:501–521

Ohtsubo H, Cheng C, Ohsawa I, Tsuchimoto S, Ohtsubo E (2004) Rice retroposon p-SINE1 and origin of cultivated rice. Breed Sci 54:1–11

Okada N, Shedlock AM, Nikaido M (2004) Retroposon mapping in molecular systematics. In: Millar WJ, Capy P (eds) Mobile genetic elements: protocols and genomic applications, Humana Press, Totowa, pp 189–226

Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N, Kato M (2005) Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Molec Biol Evol 22:285–296

Olsen KM, Schaal BA (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proc Natl Acad Sci USA 96:5586–5591

Olson ME (2002) Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). Syst Bot 27:55–73

Park KC, Kim NH, Cho YS, Kang KH, Lee JK, Kim NS (2003) Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. Theor Appl Genet 197:203–209

Park JM, Schneeweiss GM, Weiss-Schneeweiss H (2007) Diversity and evolution of Ty1-copia and Ty3-gypsy retroelements in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). Gene 387:75–86

Pavesi G, Mauri G, Pesole G (2004) In silico representation and discovery of transcription factor binding sites. Brief Bioinf 5:217–236

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis THN, Flavell A.J (1999) Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. Pl J 19:711–717

Pearce SR, Knox M, Ellis THN, Flavell AJ, Kumar A (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet 263:898–907

Perret M, Chautems A, Spichiger R, Kite G, Savolainen V (2003) Systematics and evolution of tribe Sinningieae (Gesneriaceae): evidence from phylogenetic analyses of six plastid DNA regions and nuclear ncpGS. Amer J Bot 90:445–460

Perry KL, Palukaitis P (1990) Transcription of tomato ribosomal DNA and the organization of the intergenic spacer. Mol Gen Genet 221:102–112

Plovanich AE, Panero JL (2004) A phylogeny of the ITS and ETS for *Montanoa* (Asteraceae: Heliantheae). Molec Phylogenet Evol 31:815–821

Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F (2002) Development of S-SAP markers base on a LTR-like sequence from *Medicago sativa* L. Mol Genet Genom 267:107–114

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3:827–837

Price Z, Schulman AH, Mayes S (2003) Development of new marker methods—an example from oil palm. Pl Genet Res 1:103–113

Prieto JL, Pouilly N, Jenczewski E, Deragon JM, Chevre AM (2005) Development of crop-specific transposable element (SINE) markers for studying gene flow from oilseed rape to wild radish. Theor Appl Genet 111:446–455

Provan J, Thomas WTB, Forster BP, Powell W (1999) Copia-SSR: a simple marker technique which can be used on total genomic DNA. Genome 42:363–366

Purugganan MD, Wessler SR (1995) Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. Molec Ecol 4:265–269

Qiu P (2003) Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. Biochem Biophys Res Commun 309:495–501

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Genet Genom 271:91–97

Ray DA, Xing J, Salem AH, Batzer MA (2006) SINEs of a *nearly* perfect character. Syst Biol 55:928–935

Ray DA (2007) SINEs of progress: mobile element applications to molecular ecology. Molec Ecol 16:19–33

Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. Proc Biol Sci 263:549–557

Rodriguez M, ÓSullivan D, Donini P, Papa R, Chiapparino E, Leigh F, Atiene G (2006) Integration of retrotransposons-based markers in a linkage map of barley. Molec Breed 17:173–184

Rodríguez-Trelles F, Tarrío R, Ayala FJ (2006) Origins and evolution of spliceosomal introns. Annual Rev Genet 40:47–76

Rogers SO, Bendich AJ (1987) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Pl Molec Biol 9:509–520

Rombauts S, Florquin K, Lescot M, Marchal K, Rouze P, Van de Peer Y (2003) Computational approaches to identify promoters and cis-regulatory elements in plant genomes. Pl Physiol 132:1162–1176

Rossetto M, McNally J, Henry RJ (2002). Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. Theor Appl Genet 104:61–66

Rosselló J, Lázaro A, Cosín R, Molins A (2007) A phylogeographic split in *Buxus balearica* (Buxaceae) as evidenced by nuclear ribosomal markers: when ITS paralogues are welcome. J Molec Evol 64:143–157

Ruggiero MV, Procaccini G (2004) The rDNA ITS region in the lessepsian marine angiosperm *Halophila stipulacea* (Forssk.) Aschers (Hydrocharitaceae): intragenomic variability and putative pseudogenic sequences. J Molec Evol 58:115–121

Sanderson MJ, Doyle JJ (1992) Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst Biol 41:4–17

Sang T (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit Rev Biochem Molec Biol 37(3):121–147

Sang T, Crawford DJ, Stuessy TF (1995) Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. Proc Natl Acad Sci USA 92:6813–6817

Schlötterer C, Hauser MT, Von Haesler A, Tautz D (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Molec Biol Evol 11:513–522

Schlüter PM, Kohl G, Stuessy TF, Paulus HF (2007) A screen of low-copy nuclear genes reveals the *LFY* gene as phylogenetically informative in closely related species of orchids (*Ophrys*). Taxon 56:493–504

Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. Proteins 9:180–190

Schulman AH, Flavell AJ, Ellis TH (2004) The application of LTR retrotranspons as molecular markers in plants. Methods Molec Biol 260:145–173

Schultz J, Müller T, Achtziger M, Seibel PN, Dandekar T, Wolf M (2006) The internal transcribed spacer 2 database—a web server for (not only) low level phylogenetic analyses. Nucleic Acids Res 34:W704–W707

Shedlock AM, Takahashi K, Okada N (2004) SINEs of speciation: tracking lineages with retroposons. Trends Ecol Evol 19:545–553

Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. Bioessays 22:148–160

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annual Rev Biochem 72:449–479

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. Amer J Bot 85:1301–1315

Soleimani VD, Baum BR, Johnson DA (2005) Genetic diversity among barley cultivars assessed by sequence-specific amplification polymorphism. Theor Appl Genet 110:1290–1300

Soleimani VD, Baum BR, Johnson DA (2007) Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data. Genet Resour Crop Evol 54:83–97

Sourdille P, Tavaud M, Charmet G, Bernard M (2001) Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes. Theor Appl Genet 103:346–352

Springer MS, Hollar LJ, Burk A (1995) Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals. Molec Biol Evol 12:1138–1150

Stappen JV, Marant S, Volckaert G (2003) Molecular characterization and phylogenetic utility of the rDNA external transcribed spacer region in *Stylosanthes* (Fabaceae). Theor Appl Genet 207:291–298

Starr JR, Harris SA, Simpson DA (2003) Potential of the 5′ and 3′ ends of the intergenic spacer (IGS) of rDNA in the Cyperaceae: new sequences for lower-level phylogenies in sedges with an example from *Uncinia* Pers. Int J Pl Sci 164:213–227

Steinkellner H, Lexer C, Turetschek E, Glössl J (1997) Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. Molec Ecol 6:1189–1194

Stephan W (1989) Tandem-repetitive non-coding DNA: forms and forces. Molec Biol Evol 6:198–212

Stone JR, Wray GA (2001) Rapid evolution of cis-regulatory sequences via local point mutations. Molec Biol Evol 18:1764–1770

Strand AE, Leebens-Mack J, Milligan BG (1997) Nuclear DNA-based markers for plant evolutionary biology. Molec Ecol 6:113–118

Streelman JT, Zardoya R, Meyer A, Karl SA (1998) Multilocus phylogeny of chichlid fishes (Pisces: Perciformes): evolutionary comparison of microsatellite and single-copy nuclear loci. Molec Biol Evol 15:798–808

Stuart-Rogers C, Flavell AJ (2001) The evolution of Ty1-copia group retrotransposons in gymnosperms. Molec Biol Evol 18(2):155–163

Studer B, Widmer F, Enkerli J, Kölliker R (2006) Development of novel microsatellite markers for the grassland species *Lolium multiflorum*, *Lolium perenne* and *Festuca pratensis*. Molec Ecol 6:1108–1110

Suoniemi A, Tanskanen J, Schulman AH (1998) Gypsy-like retrotransposons are widespread in the plant kingdom. Pl J 13:699–705

Syed NH, Sureshundar S, Wilkinson MJ, Bhau BS, Cavalcanti JJV, Flavell AJ (2005) Ty1-copia retrotransposon-based S-SAP marker development in cashew (*Anacardum occidentale* L.). Theor Appl Genet 110:1195–1202

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1997) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. 2nd edn. Sinauer Assoc. Inc., Massachusetts, pp 407–514

Takaiwa F, Kikuchi S, Oono K (1990) The complete nucleotide sequence of the intergenic spacer between 25S and 17S rDNAs in the rice. Pl Molec Biol 15:933–935

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389–399

Tang SX, Knapp SJ (2003) Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003

Tatout C, Warwick S, Lenoir A, Deragon JM (1999) SINE insertions as clade markers for wild crucifer species. Molec Biol Evol 16:1614–1621

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Tautz D (2000) Evolution of transcriptional regulation. Curr Opin Genet Dev 10:575–579

Theissen G (2002) Secret life of genes. Nature 415:741

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

Tompa M, Li N, Bailey TL, Church GM, de Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Regnier M, Simonis N, Sinha S, Thijs G, van Heiden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z (2005) Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 23:137–144

Tucci GF, Simeone MC, Gregori C, Maggini F (1994) Intergenic spacers of rRNA genes in three species of the Cynareae (Asteraceae). Pl Syst Evol 190:187–193

Van den Broeck D, Maes T, Saber M, Zetho J, de Keukeleire P, ĎHauw M, van Montagu M, Gerats T (1998). Transposon display identifies individual transposable elements in high copy number lines. Pl J 13:121–129

Van Glabeke S, Coart E, Honnay O, Roldán-Ruiz I (2007) Isolation and characterization of polymorphic microsatellite markers in *Anthyllis vulneraria*. Molec Ecol Notes 7:477–479

Van Treuren R, Kuittinen H, Kärkkäinen K, Baena-Gonzalez E, Savolainen O (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Molec Biol Evol 14:220–229

Vargas P, McAllister HA, Morton C, Jury SL, Wilkinson MJ (1999) Polyploid speciation in *Hedera* (Araliaceae): phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. Pl Syst Evol 219:165–179

Venema J, Tollervey D (1999) Ribosome synthesis in *Saccharomyces cerevisiae*. Annual Rev Genet 33:261–311

Venturi S, Dondini L, Donini L, Sansavini S (2006). Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. Theor Appl Genet 112:440–444

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis TH (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. Molec Biol Evol 20:2067–2075

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Gen Genomics 272:504–511

Volkov RA, Komarova NY, Panchuk II, Hemleben DV (2003) Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. *Petota* (genus *Solanum*). Molec Phylogenet Evol 29:187–202

Volkov RA, Komarova NY, Hemleben V (2007) Ribosomal DNA in plant hybrids: inheritance, rearrangements, expression. Syst Biodiv 5:261–276

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Voytas DF, Cummings MP, Konieczcy A, Ausubel FM, Rodermel SR (1992) Copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

Walsh JB (1987) Persistence of tandem arrays: Implications for satellite and simple sequence DNAs. Genetics 115:553–567

Walton C, Somboon P, O'Loughlin SM, Zhang S, Harbach RE, Linton YM, Chen B, Nolan K, Duong S, Fong MY, Vythilingum I, Mohammed ZD, Trung HD, Butlin RK (2007) Genetic diversity and molecular identification of mosquito species in the *Anopheles maculatus* group using the ITS2 region of rDNA. Infect Genet Evol 7:93–102

Wang CN, Möller M, Cronk QCB (2004) Phylogenetic position of *Titanotrichum oldhamii* (Gesneriaceae) inferred from four different gene regions. Syst Bot 29:407–418

Warwick SI, Sauder CA, Al-Shehbaz IA, Jacquemoud F (2006) Phylogenetic relationships in the Brassicaceae tribes of Anchoniae, Chorisporeae, Euclidieae, and Hesperideae based on nuclear ribosomal ITS DNA sequences. Ann Missouri Bot Gard 94:56–78

Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5:276–287

Wasserman WW, Palumbo M, Thompson W, Fickett JW, Lawrence C.E (2000) Human-mouse genome comparisons to locate regulatory sites. Nat Genet 26:225–228

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements and the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM (1992) LEAFY controls floral meristem identity in *Arabidopsis*. Cell 29:843–859

Weitzman JB (2003) Tracking evolution's footprints in the genome. J Bioinf 2:9

Wendel JF, Schnabel A, Seelanan T (1995a) An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Molec Phylogenet Evol 4:298–313

Wendel JF, Schnabel A, Seelanan T (1995b) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92:280–284

Wessler SR (2006) Eukaryotic transposable elements: teaching old genomes new tricks. In: Caporale L (ed) The implicit genome, Oxford University Press, Oxford, pp 138–165

Wheelan SJ, Scheifele LS, Martínez-Murillo F, Irizarry RA, Boeke JD (2006) Transposon insertion site profiling chip (TIP-chip). Proc Natl Acad Sci 103:17632–17637

Wheeler WC, Honeycutt RL (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Molec Biol Evol 5:90–96

Wichman SR, Wright SD, Cameron EK, Keeling DJ, Gardner RC (2002) Elevated genetic heterogeneity and Pleistocene climatic instability: inferences from nrDNA in New Zealand *Coprosma* (Rubiaceae). J Biogeogr 29:943–954

Widmer A, Baltisberger M (1999) Molecular evidence for allopolyploid speciation and a single origin of the narrow endemic *Draba ladina* (Brassicaceae). Amer J Bot 86:1282–1289

Wissemann V (2000) Molekulargenetische und morphologisch-anatomische Untersuchungen zur Evolution und Genomzusammensetzung von Wildrosen der Sektion *Caninae* (DC.), Ser Bot Jahrb Syst 122:347–429

Wolf M, Achtziger M, Schultz J, Dandekar T, Muller T (2005) Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. RNA 11:1616–1623

Won H, Renner SS (2006) Dating dispersal and radiation in the gymnosperm *Gnetum* (Gnetales)—clock calibration when outgroup relationships are uncertain. Syst Biol 55:610–622

Woods K, Hilu KW, Borsch T, Wiersema JH (2005) Pattern of variation and systematics of *Nymphaea odorata*: II. Sequence information from ITS and *trn*L-*trn*F. Syst Bot 30(3):481–493

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) The evolution of transcriptional regulation in eukaryotes. Molec Biol Evol 20:1377–1419

Wu KS, Jones R, Danneberger L, Scolnik PA (1994) Detection of microsatellite polymorphisms without cloning. Nucleic Acids Res 22:3257–3258

Xu JH, Kurata N, Akimoto M, Ohtsubo H, Ohtsubo E (2005) Identification and characterization of Australian wild rice strains of *Oryza meridionalis* and *Oryza rufipogon* by SINE insertion polymorphism. Genes Genet Syst 80:129–134

Yan XF, Lian CL, Hogetsu T (2006) Development of microsatellite markers in ginkgo (*Ginkgo biloba* L.). Molec Ecol Notes 6:301–302

Xu JH, Cheng C, Tsuchimoto S, Ohtsubo H, Ohtsubo E (2007) Phylogenetic analysis of *Oryza rupifogon* strains and their relations to Oryza sative strains by insertion polymorphism of rice SINEs. Genes Genet Syst 82(3):217–229

Yang YW, Lai KN, Tai PY, Ma DP, Li WH (1999) Molecular phylogenetic studies of *Brassica*, *Rorippa*, *Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S–25S rDNA. Molec Phylogenet Evol 13:455–462

Yu GX, Wise RP (2000) An anchored AFLP and retrotransposon-based map of diploid *Avena*. Genome 43:736–749

Zhang Z, Gerstein M (2003) Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements. J Bioinf 2:11

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Molec Evol 50:324–338

Zwettler D, Vieira CP, Schlötterer C (2002) Polymorphic microsatellites in *Antirrhinum* (Scrophulariaceae), a genus with low levels of nuclear sequence variability. J Heredity 93:217–221