Molecular Ecology (1995) 4, 265–269

SHORT COMMUNICATION

Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA

M. D. PURUGGANAN* and S. R. WESSLER*†
*Departments of *Botany and †Genetics, University of Georgia, Athens, GA 30602 USA*

Abstract

Transposable elements are mobile sequences found in nuclear genomes and can potentially serve as molecular markers in various phylogenetic and population genetic investigations. A PCR-based method that utilizes restriction site variation of element copies within a genome is developed. These patterns of site variation, referred to as transposon signatures, are useful in differentiating between closely related groups. Signature data using the *magellan* retrotransposon, for example, is useful in examining relationships within the genus *Zea* and *Tripsacum*. This method allows transposable elements, or even other multiple-copy nuclear DNA sequences, to be generally utilized as molecular markers in discriminating between other closely related species and subspecies.

Keywords: retrotransposon, *Zea*, *Tripsacum*, restriction site polymorphism, DNA fingerprinting

Received 22 July 1994; revision accepted 11 November 1994

Introduction

Transposable elements are a class of genetic loci that are capable of movement within nuclear genomes (Berg & Howe 1989). These mobile sequences have been isolated from a large number of prokaryotic and eukaryotic species (Berg & Howe 1989; Voytas *et al.* 1992; Robertson 1993). Molecular geneticists have identified two major classes of tranposable elements that differ from each other both in their structure and mechanism of transposition. Class I elements include the retrotransposons, which are evolutionarily related to retroviruses and replicate via an RNA intermediate. The yeast *Ty* and *Drosophila melanogaster copia* and *gypsy* elements are typical class I retrotransposons. Class II transposons, on the other hand, move via a DNA intermediate and include the maize *Ac/Ds* and *Drosophila P* and *mariner* element families. Together, members of these two element classes make up a significant portion of eukaryotic genomes (Bingham & Zachar 1989); in *Drosophila melanogaster*, for example, transposon sequences may account for 10% of the genome. In maize, retrotransposon insertions are apparently responsible for approximately one-third of spontaneous mutations (Varagona *et al.* 1992).

Correspondence: Michael Purugganan, Department of Biology, University of California at San Diego, La Jolla, CA 92093, USA. Tel. +1 619 534 2514. Fax +1 619 534 7108. E-mail: mikep@jeeves.ucsd.edu

Transposable elements possess unique properties that make them suitable molecular markers for studying relationships between closely related species and populations. At least one class of transposable elements, the class I retrotransposons, appear to evolve at significantly higher rates than conventional nuclear loci (Holland *et al.* 1982). Moreover, various methods are currently available to routinely isolate element sequences from different species which can then be utilized in genetic studies (Voytas *et al.* 1992; Flavell *et al.* 1992; Robertson 1993). However, the presence of multiple, often dissimilar, copies of a transposon within a species genome makes the use of these mobile genes as molecular markers problematic. Nevertheless, transposable elements have been used in a limited fashion to reveal phylogenetic relationships between closely related species (Daniels *et al.* 1983). The distribution patterns and nucleotide sequence of several elements have also been used to infer instances of interspecific hybridization between plant taxa (Purugganan & Wessler 1994; Talbert *et al.* 1990) and ecological association between *Drosophila* species and parasitic mites (Houck *et al.* 1991). Insertion sequence hybridization patterns, referred to as IS fingerprints and handprints, have also been used to type *Rhizobium* strains (Kosier *et al.* 1993).

In this paper, a rapid method for generating restriction site variation data from PCR-amplified transposable element sequences is presented. Transposable element PCR-

amplified DNAs are isolated with primers that can simultaneously recognize different copies of a particular element within an organism's genome. The PCR products are cut with four- and five-cutter restriction enzymes, and the restriction sites within the amplified copies from each species are scored as being present, absent or polymorphic. The pattern of restriction sites that are either fixed or polymorphic between transposable element copies residing within a genome are referred to as the transposon signature; these patterns can readily differentiate closely related plant species and subspecies. In principle, generating these signature patterns utilize the same principles as PCR-RFLP analysis of single-copy nuclear DNA (scnDNA) (Karl & Avise 1993; Jarmann *et al.* 1993).

The *magellan* element, a *Ty3/gypsy*-like retrotransposon first isolated from maize (*Zea mays* ssp. *mays*), was used in this study (Purugganan & Wessler 1994). Transposon signature data using this element was utilized to analyse relationships within the genus *Zea*. This genus comprises maize as well as the Mexican and Guatemalan teosintes, which are native to Mesoamerica (Doebley *et al.* 1987). Based on several morphological and ecological characteristics, systematists have divided the genus into two sections, mays and luxuriantes (Doebley 1990). *Zea mays* is the only member of the section mays. This species, however, includes several genetically distinct subspecies, including Z. *mays* ssp. *mays*, and the Mexican teosintes Z. *mays* ssp. *parviglumis* and Z. *mays* ssp. *huehuetenangensis*. The section luxuriantes includes the Guatemalan annual teosinte *Z. luxurians* and the Mexican perennial teosinte *Zea diploperennis*. For over a century, the relationships within the genus *Zea* were the subject of considerable controversy, although the intrageneric phylogeny has been largely resolved with recent work using both allozyme and chloroplast DNA analysis (Doebley *et al.* 1987).

Restriction site polymorphism analysis of a limited region of the *magellan* retrotransposon appears sufficient to discriminate between even closely related *Zea mays* subspecies. This analysis reveals that transposon signatures are suitable molecular markers for phylogenetic and population genetic investigations when few nuclear markers are available. This method can easily be extended to other tranposable elements found in other species, and may also be generally applicable in designing molecular markers from multiple-copy nuclear DNA sequences (mcnDNA).

Materials and methods

DNA isolation and PCR analysis

Seed for different Mexican and Guatemalan teosintes were a gift from J. Doebley (University of Minnesota). *Tripsacum andersonii* genomic DNA were obtained from Shawn White (University of Georgia). *Zea* species and *Tripsacum andersonii* genomic DNA were isolated from leaf tissue using standard protocols (Dellaporta *et al.* 1983).

PCR-amplified *magellan* DNA was generated from *Zea mays* ssp. *mays*, Z. *mays* ssp. *parviglumis*, Z. *mays* ssp. *huehuetenangensis*, Z. *luxurians* and Z. *diploperennis*. *Magellan* sequences were also amplified from *Tripsacum andersonii* (see Results and Discussion). *Magellan* PCR primers were designed to amplify an ~ 600-bp sequence at the 5′ end of the retrotransposon (see Fig. 1A). The 5′ primer (MLTR1) (5′-TGTCAGGAGACTGACGCAGC-3′) recognizes the beginning of a 341-bp-long terminal repeat (LTR) sequence found at the element's ends. The 3′ primer (MINT1) (5′-GGGTCGTTGCGGTCTACTGC-3′) recognizes an internal sequence approximately 260 base pairs downstream of the element's 5′ LTR. PCR reactions contained 200 ng of genomic DNA and 70 ng of primers were added to standard reaction buffer [16.6 mM NH_4SO_4, 67 mM Tris (pH 8.8), 6.7 mM $MgCl_2$, 6.7 μM EDTA] with 0.2 mM each of dNTP and 2.5 units of Taq polymerase (Perkin Elmer-Cetus). The reactions were cycled 35 times at 95 °C for 1 min, 55 °C for 2 min and 72 °C for 2 min, followed by a 72 °C incubation for 10 min.

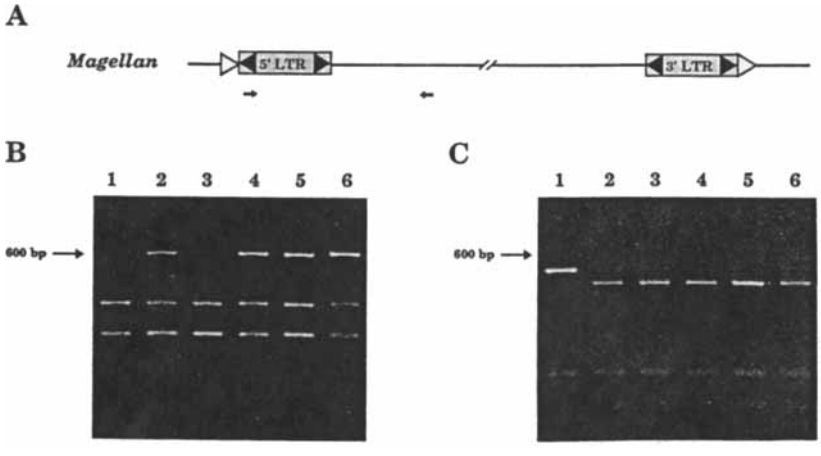

Fig. 1 (A) Schematic diagram of the *magellan* retrotransposon. The position of the long terminal repeats (LTRs) are shown in the figure. The open triangles depict the direct repeat sequences generated upon insertion of an element. The relative positions of the PCR primers used in this study are indicated by the arrows. The results of PCR-RFLP analysis of *magellan* PCR-amplified DNA using (B) *Hin*cII and (C) *Alu*I are shown. The species/subspecies PCR-amplified DNA shown in the gels are as follows: lane 1, *Zea mays* ssp. *mays*; lane 2, *Zea mays* ssp. *parviglumis*; lane 3, *Zea mays* ssp. *huehuetangensis*; lane 4, *Zea diploperennis*; lane 5, *Zea luxurians*; lane 6, *Tripsacum andersonii*.

Restriction site variation analysis

For each species, PCR-amplified DNA generated in several independent reactions were pooled. The PCR-amplified DNA was precipitated using ethanol and the DNA resuspended to give a final DNA concentration of 0.5–1.0 μg/μL. One-half to one microgram of DNA was cut separately with each of the following restriction enzymes: *Rsa*I, *Mbo*I, *Hinc*II, *Hae*III, *Cfo*I, *Alu*I, *Hin*fI, *Eco*RII, *Hpa*II and *Bgl*I. The cut DNA was run in a 3% NuSieve/1% regular agarose gel (FMC Bioproducts) and visualized using ethidium bromide staining.

Restriction maps were constructed for the PCR-amplified *magellan* terminal sequence from each taxa. PAUP was used to infer the relationships between taxa (Swofford 1993). Sites were scored as being present, absent or polymorphic. Restriction site characters were treated as ordered, with polymorphic sites intermediate between site gain or loss.

Results and Discussion

PCR amplification generates an expected 600-bp *magellan* sequence from all members of the genus *Zea* examined. This 600-bp region contains the magellan 5′ long terminal repeat (LTR) as well as about 260 bp of internal *magellan* sequence. The PCR-amplified band should contain copies from different *magellan* family members residing within a genome. Based on Southern blot analysis, *magellan* is a low-copy number retrotransposon that is present in 4–10 copies within the genomes of members of the genus *Zea* (M. D. Purugganan and S. R. Wessler, unpublished data).

Ten different restriction enzymes (nine four-cutter and one five cutter enzyme) were used to analyse the PCR-amplified DNA from the different *Zea* species and subspecies. Seven restriction enzymes recognized at least one site in the entire analysis (see Figs 1B–C and 2). Within this 600-bp region, a total of 16 sites were recognized. This restriction site polymorphism analysis surveyed 11% of the bases in the PCR-amplified *magellan* sequence.

Within a species genome, several restriction sites may be present within some *magellan* copies while absent in other copies. These intragenomic restriction site polymorphisms that appear in the analysis result from the amplification of divergent copies of an element family within a genome (see Fig. 1B–C). The level of molecular diversity between *magellan* copies within a genome are substantial

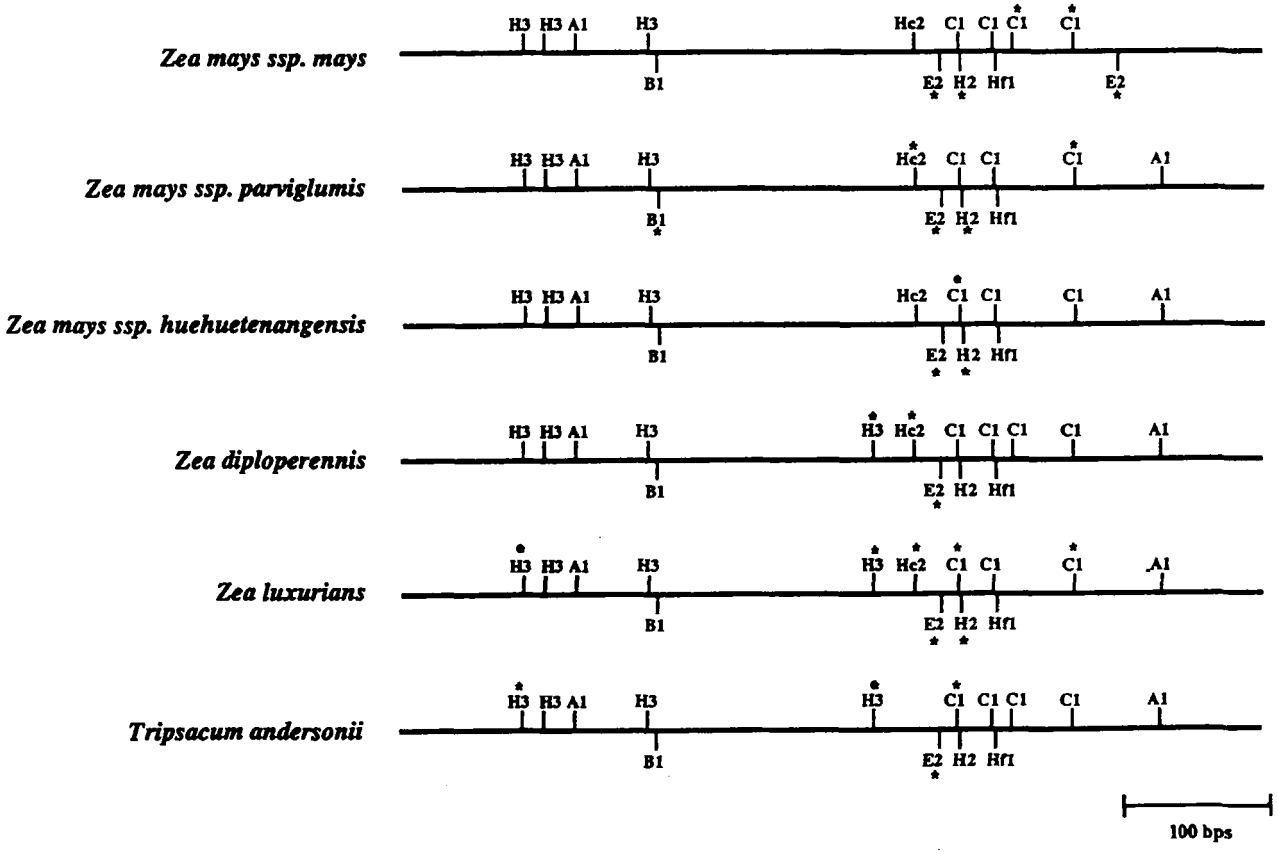

Fig. 2 Restriction maps of PCR-amplified *magellan* DNA from different *Zea* and *Tripsacum* species/subspecies. The stars indicate restriction sites that are polymorphic between magellan copies within a species/subspecies genome. The sites depicted are: A1, *Alu*I; B1, *Bgl*I C1, *Cfo*I; E2, *Eco*RII; H2, *Hpa*II; H3, *Hae*III; Hc2, *Hinc*II; Hf1, *Hin*fI.

in all species tested; the proportion of intragenomic polymorphic sites within species genomes range from 0.20–0.50. These high levels of intragenomic polymorphisms are confirmed by sequence analysis of *magellan* copies within *Zea* genomes (Purugganan & Wessler 1994), and are due to the presence of different *magellan* subfamilies within a genome, and the introduction of new element copies from other species via hybridization. The high retrotransposon intercopy sequence divergence is also promoted by the reverse transcriptase-mediated retrotransposition process, which is error-prone and results in mutation rates as much as 10^3 times the rate for conventional genes (Holland *et al.* 1982).

The restriction site analysis allows us to construct multiple-copy restriction maps (transposon signatures) for each species using *magellan* 5′ terminal sequences (see Fig. 2). These transposon signatures are species-specific, and even different subspecies within *Zea mays* display distinctive patterns of restriction site variation. These transposon patterns are capable of discriminating even between groups that appear to have diverged only in the past 10 000 years (*Z. mays* ssp. *mays* vs. ssp. *parviglumis* and ssp. *huehuetenangensis*).

These transposon signatures can accurately reconstruct the relationships between the different species (see Fig. 3). The phylogenetic tree, which was inferred using parsimony analysis, correctly places all *Zea mays* subspecies in one monophyletic group, separate from the two *Zea* species within section luxuriantes. The inferred tree is concordant with the phylogeny of the genus as inferred from chloroplast DNA restriction site polymorphism data (Doebley *et al.* 1987).

The *magellan* signature also establishes the relationship between *Tripsacum andersonii* and members of the genus *Zea*. *Tripsacum andersonii* is a perennial rhizomatous grass found in mesic woodlands in Central and South America (Larson & Doebley 1994). Based primarily on evidence from cytological studies and transposable element distribution patterns, this species is believed to be an intergeneric hybrid between an unknown *Zea* and

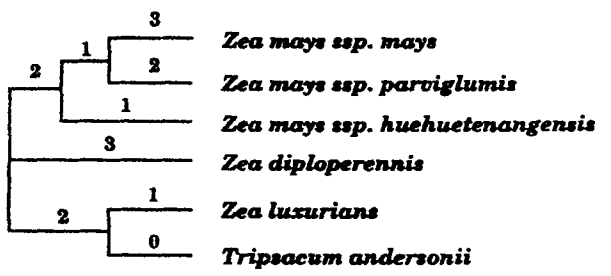

Fig. 3 Phylogeny of study species using transposon signature data from *magellan*. The tree is a 50% majority rule consensus of most-parsimonious trees inferred under maximum parsimony. Numbers above the branches indicate number of changes along the branch. The tree length is 15, with a consistency index of 0.73.

Tripsacum progenitor (Talbert *et al.* 1990). *Magellan* is absent in all *Tripsacum* species analysed except *T. andersonii*. Based on detailed analysis of *magellan* sequence data, it has been proposed that *Z. luxurians* may be the *Zea* progenitor of *T. andersonii* (Purugganan & Wessler 1994). This is substantiated by the *magellan* signature data, which shows a very close relationship between *Z. luxurians* and *T. andersonii*. Recently, chloroplast DNA data has provided additional evidence that confirms *Z. luxurians* as the maternal progenitor of *T. andersonii* (Larson & Doebley 1994).

This data demonstrates the utility of a PCR-based method to determine patterns of restriction site variation between transposable element copies within a genome. Substantial sequence diversity appears to occur between element copies within and between species, and allows these mobile sequences to be utilized as molecular markers. Transposable element proliferation via transposition, duplication and even interspecific transfer, provides a potentially large pool of useful polymorphic nuclear loci for phylogenetic studies. Since current methods allow investigators to readily acquire new element sequences from any species under study (Voytas *et al.* 1992; Robertson 1993), transposon signature analysis can be easily applied to the investigation of other species.

The transposon signatures provide a rapid assay of levels of element polymorphism within a species genome, and a method to discriminate between even closely related plant subspecies. The patterns of restriction site variation can be used to infer intrageneric phylogenies, and even intraspecific, relationships. This method is also more rapid than techniques based on genomic hybridization patterns, such as insertion sequence (IS) fingerprinting and handprinting (Kosier *et al.* 1993). Moreover, unlike fingerprinting, transposon signature analysis should remain useful even when the transposable elements utilized transpose frequently. Indeed, for transposon signatures to be useful in distinguishing closely related groups, the elements should be sufficiently active in establishing divergent family members that are population- or species-specific.

It must be noted that not all members of a transposable element family within a species genome may be amplified in the PCR reaction. Variation in primer site sequence may result in biasing the amplification towards particular family members. Using low annealing temperatures as well as pooling separate reactions may circumvent much of this problem. Moreover, the use of more accurate thermostable polymerases, such as *Pfu* and *Vent* polymerases, may reduce enzyme-induced variation in the amplified products.

The use of these polymorphic transposable element sequences may be particularly useful for plant molecular phylogenetic and population genetic studies. Among

plant systems, few polymorphic nuclear markers have been identified that are useful in differentiating closely related taxa (Soltis *et al.* 1992). This technique may also be extended to restriction site mapping of multiple copy nuclear DNA sequences (mcnDNA). Gene duplications via polyploidization and interspecific hybridization are prevalent among plant species, and many plant genes are present in multiple copies within nuclear genomes. The problem of isolating orthologous loci for analysis has generally made it difficult to use nuclear plant genes in evolutionary and ecological studies. This method, however, potentially allows investigators to use rapidly evolving small multigene families in the nuclear genome as molecular markers.

Acknowledgements

The authors wish to thank J. Doebley for seed and S. White for providing DNA samples, and J. F. McDonald, M. L. Arnold, E. Friar, B. C. McCaig and S. Karl for valuable discussions. This work was funded in part by grants from the US National Institutes of Health to S.R.W. and the Alfred P. Sloan Foundation to M.D.P.

References

Berg DE, Howe MM (1989) *Mobile DNA*. American Society of Microbiology, Washington, DC.

Bingham PM, Zachar Z (1989) In: *Mobile DNA* (eds Berg D, Howe MM). American Society of Microbiology, Washington, DC.

Daniels GK, Fox GM, Lowenstein D *et al.* (1983) Species-specific homogeneity of the primate *Alu* family of repeated DNA sequence. *Nucleic Acids Research*, **11**, 7579–7593.

Dellaporta S, Wood J, Hicks J (1983) A miniprep protocol for isolating plant DNA. *Plant Molecula Biology and Reproduction*, **1**, 19–21.

Doebley JF (1990) Molecular Systematics of *Zea*. *Maydica*, **35**, 143–150.

Doebley JF, Renfroe W, Blanton A (1987) Restriction site variation in the *Zea* chloroplast genome. *Genetics*, **117**, 139–147.

Flavell A, Dunbar E, Anderson R *et al.* (1992) *Ty1-copia* group retrotransposons are ubiquitous and heterogenous in higher plants. *Nucleic Acids Research*, **20**, 3639–3644.

Holland J, Spindler K, Horodsky F *et al.* (1982) Rapid evolution of RNA genomes. *Science*, **215**, 1577–1585.

Houck M, Clark, JB, Peterson KR, Kidwell MG (1991) Possible horizontal transfer of *Drosophila P* gene by the mite *Proctolaelaps regalis*. *Science*, **253**, 1125–1129.

Jamann S, Fernandez M, Normand P (1993) Typing method for nitrogen-fixing bacteria based on RFLP – Application to the characterization of *Frankia* strains. *Molecular Evolution*, **2**, 17–26.

Karl S, Avise J (1993) PCR-based assays of mendelian polymorphisms from anonymous single-copy nuclear DNA – Techniques and applications for population genetics. *Molecular Biol. Evolution*, **10**, 342–361.

Kosier B, Puhler A, Simon R (1993) Monitoring the diversity of *R. melliloti* field and microcosm isolates with a novel rapid genotyping method using insertion elements. *Molecular Ecology*, **2**, 35–46.

Larson SR, Doebley, JF (1994) Restriction site variation in the chloroplast genome of *Tripsacum* – phylogeny and rates of sequence evolution. *Systematic Botany*, **19**, 21–34.

Purugganan. MD, Wessler, SR (1994) Molecular evolution of *magellan*, a maize *Ty3/gypsy*-like retrotransposon. *Proceedings of the National Academy of Sciences of the USA*, **91**, 11674–11678.

Robertson H (1993) The *mariner* transposable element is widespread in insects. *Nature*, **362**, 242–245.

Soltis P. Soltis D, Doyle J (1992) *Molecular Systematics of Plants*. Chapman and Hall, New York.

Swofford D (1993) *PAUP. Phylogenetic Analysis using Parsimony*. Illinois Natural History Survey, Champaign.

Talbert LE, Doebley J, Larson S, Chandler V (1990) *Tripsacum andersonii* is a natural hybrid involving *Zea* and *Tripsacum* – Molecular evidence. *American Journal of Botany*, **77**, 722-726.

Varagona MJ, Purugganan MD, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons in the maize *waxy* gene. *Plant Cell*, **4**, 811–820.

Voytas DF, Cummings MP, Konieczny A *et al.* (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proceedings of the National Academy of Sciences of the USA*, **89**, 7124–7128.