TEMPORAL WAVES OF MINIATURE INVERTED-REPEAT TRANSPOSABLE ELEMENT AMPLIFICATION DURING RICE GENOME EVOLUTION

by

ALEXANDER NAGEL

(Under the Direction of Susan R. Wessler)

ABSTRACT

Miniature inverted-repeat transposable elements (MITEs) are the single most abundant type of transposable element in the rice genome with at least 100 families and almost 100,000 elements. They are nonautonomous derivatives of DNA transposons. Little is known about how they spread and why they have been so successful. Availability of the entire rice genome sequence has permitted a whole genome analysis of five *Tourist*-like MITE families (*Castaway, Explorer, Olo*, *Tou* and *Wanderer*) in order to gain insight into their evolution. MITE sequences were found to cluster into distinct subfamilies ranging in age from around 1 million years to around 22 million years. These results suggest that MITEs spread through the rice genome in successive rounds of amplification. As part of this analysis transposon display was applied to these MITE families and 225 markers were mapped onto the 12 rice chromosomes.

INDEX WORDS: Evolution, Georgia, Mapping, Marker, MITE, Transposon, Transposon Display, Rice, Thesis

TEMPORAL WAVES OF MINIATURE INVERTED-REPEAT TRANSPOSABLE ELEMENT AMPLIFICATION DURING RICE GENOME EVOLUTION

by

ALEXANDER NAGEL

Diploma, The University of Tuebingen, Germany, 1996

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements of the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2003

TEMPORAL WAVES OF MINIATURE INVERTED-REPEAT TRANSPOSABLE ELEMENT AMPLIFICATION DURING RICE GENOME EVOLUTION

by

ALEXANDER NAGEL

Approved:

Major Professor:	Susan Wessler
Committee:	Michael Scanlon John McDonald

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2003

ACKNOWLEDGMENTS

I would like to thank my advisor Susan Wessler, for her continuous support, guidance and encouragement throughout this work. I thank my committee members, John McDonald and Michael Scanlon for their advice, interest in this work and inspiration. I am indebted to past and present members of the Wessler lab for helpful discussions and suggestions. In particular, I want to thank Dawn Holligan for all the time she put into the mapping project and Cedric Feschotte, Mark Osterlund, Ning Jiang and Xiaoyu Zhang for offering invaluable help and comments. All of them were a pleasure to work with. Special thanks to the McDonald lab for technical support and to Carolyn Lawrence, Brian Perkins, David Brown and Greg Derda for their help with computer-related problems. Thanks to the greenhouse staff and to other members of the Botany and Genetics Departments, who provided information, material, advice or uplifting comments.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Transposable elements

Transposable elements (TEs) can be divided into two major classes on the basis of their mode of transposition. Class I elements (retrotransposons) transpose via reverse transcription of a RNA intermediate to double stranded cDNA, which is then integrated into the host genome *et al.* *et al.* . Most retrotransposons can be further divided into those that have long terminal repeats (LTRs) (*copia*, *gypsy*) and those which lack this feature but terminate at one end with a poly (A) tract [long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs)] . The focus of this thesis is on members of the other major class of TEs, class II or DNA elements.

DNA transposons move via a DNA intermediate facilitated by an element-encoded transposase. The ends of DNA transposons consist of terminal inverted repeats (TIRs), which, either alone or with subterminal sequences, are recognized by the transposase (Kunze *et al.*, 1997). Transposase initiates a staggered cut next to the TIR releasing the element from its flanking sequence and then catalyzing the insertion of the transposon at a new locus. This mechanism produces short sequence duplications [target site duplications (TSDs)] at the insertion site (Kunze *et al.*, 1997; Saedler & Nevers, 1985).

Transposons that encode functional transposase are referred to as autonomous elements (Feschotte *et al*., 2002). In contrast, nonautonomous TEs are usually defective autonomous elements that are still able to transpose in response to a transposase provided in trans. Examples of nonautonomous transposons include *Ds*, *dSpm* and miniature inverted-repeat transposable elements (MITEs) (Feschotte *et al*., 2002; Kunze *et al*., 1997).

Miniature inverted-repeat transposable elements (MITEs)

MITEs are a special class of nonautonomous DNA transposable element. They were first discovered when a 128 bp insertion into a mutant allele of the maize *waxy* gene was found to be a member of a highly repetitive class of sequences . MITEs are widely distributed in plants, where they have been identified in *Arabidopsis* and in several grass species *et al.* *et al.* and in animal genomes including *C. elegans* *et al.* , mosquito , fish *et al.* , *Xenopus* and human . Several studies show that MITEs, like other DNA transposons, are preferentially found close to genes in maize and rice *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* , suggesting that they might be important players in gene evolution. With their short length (<600 bp), MITEs do not possess any coding sequences, but are nonetheless able to attain copy numbers of hundreds to thousands of elements in a genome (Feschotte *et al*., 2002).

The presence of TIRs and TSDs of conserved size and sequence within a MITE family, suggest that MITEs may have originated as deletion derivatives of autonomous DNA transposable elements. The creation of TSDs is a hallmark of transposase activity during element insertion and TSD length is highly conserved among different DNA transposons in a superfamily *et al.* . A DNA transposon superfamily is defined as a group of sequences that share the same TSD, have similar TIRs, and similarities and/or specific signatures in their transpsosases *et al.* *et al.* *et al.* *et al.* .

MITEs can be classified into two major superfamilies based on shared similarities in their TIRs as well as their TSDs. *Tourist*-like MITEs insert preferentially into TAA nucleotides and the first seven bases of their TIRs are usually GGGCCTG/GGGCATC *et al.* . *Stowaway*-like MITEs insert preferentially into the dinucleotide TA and the first ten bases of their TIRs are usually CTCCCTCCGT *et al.* . The internal sequences of elements from different superfamilies vary widely in composition as well as length. MITE sequences that share over 85% similarity over their entire length have been arbitrarily grouped into families *et al.* . This study focuses on five *Tourist*-like MITE families previously identified in rice.

Shared sequence similarity in TIR sequences as well as TSDs suggested that elements from the *mariner*/Tc1 DNA transposon superfamily are the autonomous partners of *Stowaway*-like MITEs *et al.* . Originally identified in animals, *mariner*-like elements (MLEs) have also been shown to be widespread and abundant in plants

(Feschotte & Wessler, 2002). However, in rice, only a few cases were found where the similarities between the TIRS of *Stowaway*-like MITEs and the TIRs of MLEs extended into the subterminal regions. This means that the MLEs currently in the rice genome are not the direct progenitors of most of the *Stowaway*-like MITE families present in rice today.

Sequences of the MITE superfamily *Tourist* have been associated with the DNA transposon superfamily *PIF*/*Pong*, based in part on conserved TSDs (TTA/TAA) and TIR sequences *et al.* . Both *PIF* and *Pong* elements are widespread in dicot as well as monocot plant species and were also identified in animals *et al.* . A few clear-cut relationships, between *PIF*-like TEs and *Tourist*-like MITEs could be identified in rice. The highest sequence similarity was observed between the transposon *OsPIF6* and elements from the MITE family *Castaway* (~90%) *et al.* . This result indicates that *Castaway* elements arose as deletion derivatives from *OsPIF6*-like transposons.

A model for the origin and amplification of MITEs is presented in Figure 1.1 *et al.* . Transposition of an autonomous element is mediated by the element-encoded transposase (Figure 1.1, part 1). This may lead to an increase in copy number and/or the production of internally deleted derivatives (Figure 1.1, part 2). Host epigenetic mechanisms are hypothesized to repress transposition and prevent significant copy number increase. Autonomous and nonautonomous elements degrade over time by the accumulation of point mutations as well as insertions/deletions (Figure 1.1, part 3). New autonomous elements may be introduced into the genome from another line or species by horizontal transfer (the transfer of genetic material from one species to another that does not involve a cross) or genetic introgression (the integration of genetic material introduced via a cross between two lines). Alternatively, a previously silenced transposon might be reactivated when host defense mechanisms are relaxed. The expressed transposase might then be able to recognize distantly related MITEs (Figure 1.1, part 4), leading to their amplification (Figure 1.1, part 5).

MITEs in *Oryza*

There are about 100,000 MITEs in the rice genome that are organized into many distinct families (Bureau *et al*., 1996; Feschotte *et al*., 2003; Jiang *et al*., 2003; Mao *et al*., 2000; Song *et al*., 1998; Tarchini *et al*., 2000; Zhang *et al*., 2003). Whole genome analysis of both *O*. *sativa* ssp. indica and *O*. *sativa* ssp. japonica showed that around 40% of all TEs are MITEs, comprising about 26 Mbp or 6% of the genome *et al.* *et al.* . The availability of draft sequences of the two *O*. *sativa* subspecies, indica and japonica, enabled the characterization and classification of the two transposon superfamilies, MLEs and *PIF*/*Pong*, as well as the identification of the first active DNA transposon in rice (Feschotte *et al*., 2003; Jiang *et al*., 2003; Zhang *et al*., 2003).

Compared with the other cereal crops rice has the smallest genome (430 Mbp). The genomes of sorghum, maize, barley, and wheat are estimated to be 1000, 3000, 5000 and 16,000 Mbp, respectively *et al.* *et al.* . The difference in genome size is mainly due to differential numbers of LTR retrotranspons, mostly inserted in intergenic regions (Feschotte *et al*., 2002; SanMiguel *et al*., 1996), while the haploid gene content remains similar between these different crops. LTR retrotransposons constitute 15% of the rice genome, 50–80% of the maize and >70% of the barley genome (Jiang &

Wessler, 2001; Meyers *et al.*, 2001; SanMiguel & Bennetzen, 1998; Vicient *et al.*, 1999). The rice genome is regarded as the model for the grass genetic system due to its compact genome size.

Rice is one of the most important crop species in the world. Around 60% of the world's population depends on it for most of their daily calorie intake. The genus *Oryza* contains around 23 different species from Asia, Africa, Australia and Central and South America. Two species are cultivated for food consumption, while seven are tetraploid (Figure 1.2) (Ge *et al.*, 1999). Distinct genome types (A-K) have been identified cytologically. In some cases the parental genome types of the tetraploid species could be determined (e.g. CCDD for *O. grandiglumis*) *et al.* . The two cultivated rice species, *O. sativa* and *O. glaberrima* both belong to the cluster of A genome species.

The non-cultivated *Oryza* species are generally referred to as wild rice. They have great potential as a source of germplasm for the improvement of the cultivated species for traits such as disease and drought resistance, and yield. Aside from their agricultural significance, the wild species are found world-wide (Asia, Africa, Australia and Central and South America) in a range of ecological adaptations (dry soils, inundation, salt, sun, shade). The difference in their genome size (from 300 Mbp to 900 Mbp) and the existence of diploid and tetraploid species makes the entire genus suitable for studying the impact of ecological and evolutionary questions.

Figure 1.1: Model for the origin and amplification of MITEs *et al.* . See text for explanation.

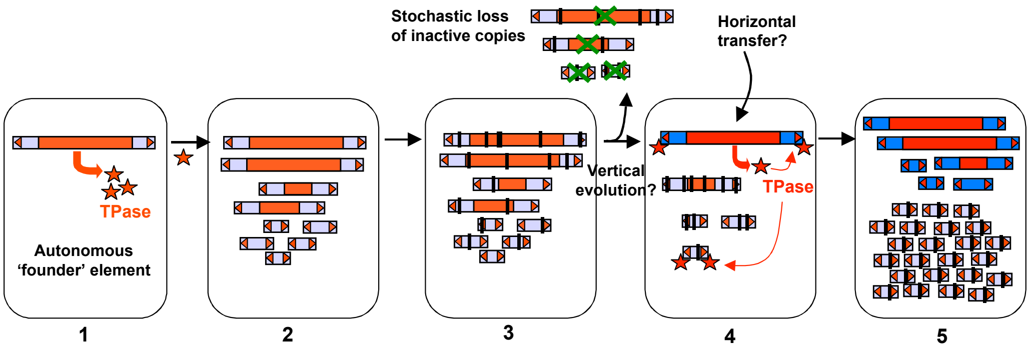
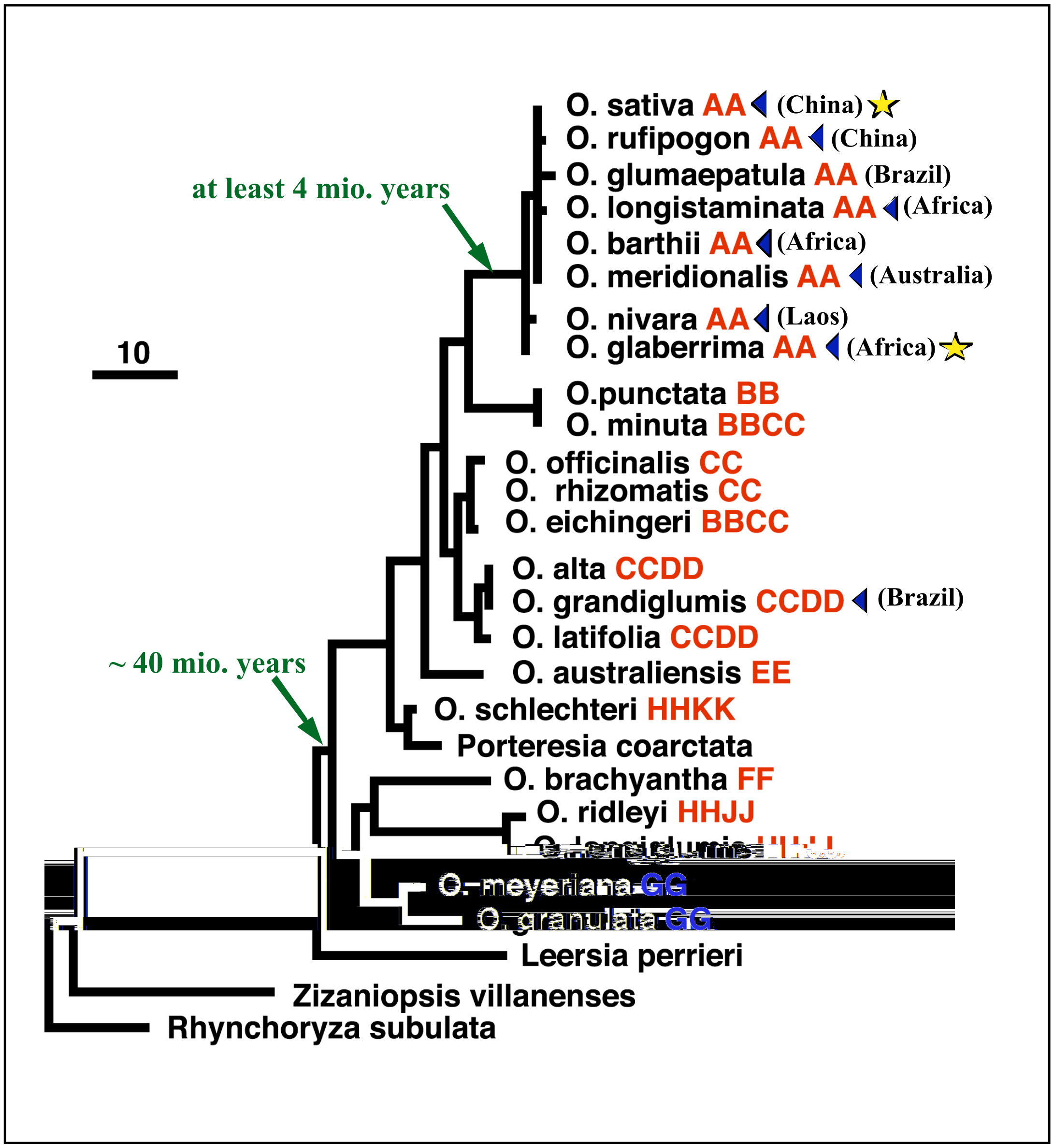

Stochastic loss
of inactive copies
Horizontal
transfer?
TPase
Autonomous
'founder' element
Vertical
evolution?
TPase
1
2
3
4
5

Figure 1.2: The most parsimonious tree generated from *mat*K gene sequences *et al.* . The genome type for each species is indicated in red uppercase letters after the species name. Species used in this study are indicated by blue arrows and their place of origin is given in brackets. *O. sativa* and *O. glaberrima* are the only two cultivated species (highlighted by yellow stars). Estimated divergence times for the genus Oryza as well as the cluster of AA genome species (Tao Sang personal communication).


at least 4 mio. years
~ 40 mio. years
10
O. sativa AA (China) ☆
O. rufipogon AA (China)
O. glumaepatula AA (Brazil)
O. longistaminata AA (Africa)
O. barthii AA (Africa)
O. meridionalis AA (Australia)
O. nivara AA (Laos)
O. glaberrima AA (Africa) ☆
O.punctata BB
O. minuta BBCC
O. officinalis CC
O. rhizomatis CC
O. eichingeri BBCC
O. alta CCDD
O. grandiglumis CCDD (Brazil)
O. latifolia CCDD
O. australiensis EE
O. schlechteri HHKK
Porteresia coarctata
O. brachyantha FF
O. ridleyi HHJJ
O. longiglumis HHJJ
O. meyeriana GG
O. granulata GG
Leersia perrieri
Zizaniopsis villanenses
Rhynchoryza subulata

CHAPTER 2

TEMPORAL WAVES OF MINIATURE INVERTED-REPEAT TRANSPOSABLE ELEMENT AMPLIFICATION DURING RICE GENOME EVOLUTION

BACKGROUND AND RATIONALE

The proximity of MITEs to genes was exploited in the development of MITEs as informative molecular markers *et al.* using a technique called Transposon Display (TD). TD is a modified AFLP technique *et al.* that produces PCR products terminating in a TE and an adjacent restriction site in flanking genomic DNA *et al.* . Previously, TD was applied to the *Heartbreaker (Hbr)* MITE family in maize and more than 200 MITE markers were mapped in a recombinant inbred line mapping population *et al.*

The development of genetic markers is essential for the exploitation of genetic diversity present in wild rice species as well as for breeding programs. Given the abundance of MITEs in rice and the previous successful MITE based marker development in maize, development of TD and its use for marker development in rice seems to be a natural extension.

In this study TD was developed for the rice *Tourist*-like MITE families *Castaway*, *Explorer*, *Olo*, *Tourist* and *Wanderer*. Subsequently, 225 MITE markers were mapped in a doubled-haploid mapping population derived from an *Oryza sativa* ssp. japonica x

Oryza sativa ssp. *indica* cross. Successful application of TD to other AA genome species suggests that these MITE primers can be used in mapping populations derived from interspecific crosses for which sequence information is very limited or none.

The MITE sequences in the rice genome show relationships of varying degrees to each other and also to autonomous DNA transposons. While several studies *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* were able to identify and characterize the major groupings of MITE sequences, data on the phylogenetic relationships of closely related MITE sequences is limited (Yang *et al.*, 2001).

In previous studies important insights into the evolution of transposable elements were made possible by phylogenetic analysis. In the case of *Alu* (SINE), it was found that the numerous *Alu* copies in humans (~500,000) have arisen from only a small number so called 'master genes', elements that were able to undergo extensive amplification (Deininger *et al.*, 1992). In yeast it was found that frequent inter-element recombination is an important factor in the evolution of Ty1/Ty2 retroelements (Jordan & McDonald, 1999), and in *Drosophila melanogaster*, the LTR retrotransposons families, identified in whole genome analysis, amplified after the split of this species from its closest relative ~2.3 million years ago (Bowen & McDonald, 2001).

Here, results from a phylogenetic of five MITE families provide insights into MITE relationships on a subfamily level as well as their relative age. This study revealed the presence of numerous and distinct subfamilies, which appear to have arisen in temporal waves of amplification over the past 20 million years.

MATERIAL AND METHODS

Terminology

The term *Tourist* was originally used to name the first MITEs identified in maize and rice , but is now also used as a name for one of the two superfamilies of MITEs *et al.* . In order to avoid confusion about terminology, the following distinction is made. In this study the original *Tourist* rice MITE sequences (Bureau & Wessler, 1994a) are referred to as *Tou* sequences while the term *Tourist* is used to relate to the superfamily and includes all the MITE families characterized in this study *et al.* .

Plant material and genomic extractions

DNA from several wild rice species as well as from mapping population lines was provided by Dr. Susan McCouch (Cornell). Plant DNA was extracted as described *et al.* . The lines with AA genomes were as follows: *Oryza barthii* (139A, 139C), *O. rufipogon* (1786B, 160A), *O. glaberrima* (142B, 132), *O. nivara* (157A, 151C), *O. longistaminata* (WLO2), *O. spontanea* (176A), *O. meridionalis* (1743A, 1760C), *O. sativa* ssp. indica (IR64), *O. sativa* ssp. japonica (Azucena, Nipponbare). The two CCDD genome accessions used were: O. grandiglumis (1754C, 1763A). The doubled-haploid (DH) mapping population used was derived via anther culture from a cross between *O. sativa* ssp. japonica (Azucena) and *O. sativa* ssp. indica (IR64) *et al.* *et al.* . Coix DNA was extracted as described .

Transposon display

Primers for transposon display (TD) analysis were designed to recognize conserved regions of the aligned sequences from the subfamilies *Explorer*, *Castaway*, *Olo*, *Tou* and *Wanderer*. The Prime program (GCG/Wisconsin) was used and the resulting primer sequences modified to recognize the consensus sequence. The position of element specific primers was ideally chosen to be within 50 bp from the end of the element, but in one case (*Castaway*) the primers were located 100 bp from the end of the element. In several cases degenerate primers were designed in order to maximize the number of elements amplified.

Restriction/ligation (R/L) reactions contained 50-500 ng of genomic DNA in a total volume of 40 μl of 1 x OPA (One-Phor-All) buffer, 50 ng/μl BSA, 5 mM DTT and 5 units of BfaI and incubated for 1-3 h at 37ºC. Subsequently, 10 μl containing 1 x OPA-plus buffer (10 mM Tris-Acetate, 10 mM Mg-Acetate, 50 mM K-Acetate at pH 7.5), 50 ng/μl BSA, 5 mM DTT, 50 pmol BfaI-adapter (5'-GACGATGAGTCCTGAG, 5'-TACTCAGGACTCAT), 1.2 mM ATP and 1 unit T4 DNA ligase were added and incubation continued for 3-15 h at 37ºC. Reactions were diluted 4 fold with $T_{01}E$ buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0).

Preselective amplification (PA) was performed in 30 μl containing 3 μl of the 4 fold diluted R/L reaction, PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 0.3 μM BfaI+0 primer (GACGATGAGTCCTGAGTAG), 0.3 μM element specific preselective primer (*Castaway*: 5'-TGYGTTTTTCCTGYGGTYCAATC, *Explorer*: 5'–AGCCTTTGCTTTTAGATCGC, *Olo*: 5'-ATTCGTGTCCGAAAACCCCTTCC, *Tou*: 5'-CGTCTCGCRAATTAGTCCAAGATTATGG, *Wanderer*: 5'-ATTAATCACGYGCTAATGGACCGYTC) and 2.5 units Ampli*Taq* DNA polymerase (Perkin-Elmer/ABI). PCR conditions were as follows: 1 x (2 min/72ºC), 1 x (3 min/94ºC), 24 x (30 sec/94ºC, 30 sec/58ºC, 60 sec/72ºC), 1 x (5 min/72ºC). Reactions were diluted 10 fold with $T_{01}E$ buffer.

Selective amplification (SA) was performed in 10 μl containing 2.5 μl of the 10 fold diluted PA, PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTP, 0.5 μM *Bfa*I+N primer (5'-GACGATGAGTCCTGAGTAG+N), 0.5 pmol ^{33}P-labeled element specific selective primer (*Castaway*: 5'-TGYGGTCCAATCAAACGRYCATTC, *Explorer*: 5'-CCTTTG CTTTTAGATCGCTAAGAACACG, *Olo*: 5'-ATCCGGTCAAACGTYYGATGTGAC ACCC, *Tou*: 5'-CATCCGATGTGATAGGGACTTAAAAG, *Wanderer*: 5'-TAATGGA CCGCTCYGTTTTCC) and 0.2 units Ampli*Taq* DNA polymerase. PCR conditions were as follows: 1 x (5 min/94ºC), 9 x (30 sec/94ºC, 30 sec/67-59ºC, 60 sec/ 72ºC), 27 x (30 sec/94ºC, 30 sec/58ºC, 60 sec/ 72ºC), 1 x (5 min/72ºC). A touchdown protocol was used where the initial annealing temperature was 67ºC and decreased by 1ºC with each cycle to 58ºC *et al.* .

Samples were prepared for gel electrophoresis by adding 10 μl of loading-denaturing buffer (0.025% xylene cyanol, 0.025% bromophenol blue, 10 mM EDTA, 98% deionized formamide), denatured for 5 min at 94ºC and 3 μl were loaded on a 6% denaturing (7.5 M urea) acrylamide-bisacrylamide (19:1) gel in 1 x TBE buffer (890 mM Tris, 890 boric acid, 20 mM EDTA, pH 8.3). After electrophoresis (2 h, 35 mA) the gel was transferred to filter paper, dried, and exposed to X-ray film.

To recover and reamplify individual bands a needle-scratching technique was used as described *et al.* . The PCR parameters were the same as in the preselective amplification. Fragments of the correct size were gel purified if necessary, cloned

(Invitrogen TA cloning kit) and sequenced by the Molecular Genetics Instrumentation Facility (University of Georgia).

Genetic mapping

MITEs from the *Tou*-like families *Castaway, Explorer, Olo*, *Tou* and *Wanderer* were mapped in the IR64 x Azucena doubled-haploid (DH) mapping population as described in *et al.* *et al.* . TD was done with genomic DNA from the two parents, IR64 and Azucena, and from 96 DH lines. Polymorphic bands were scored by hand. Mapmaker 2.0 *et al.* was used on an Apple computer to place the scored markers on chromosomes and to determine their relative position with respect to an existing set of SSR framework markers *et al.* *et al.* .

Sequence analysis

The following databases were used for sequence analysis: Genbank, BAC ends database *et al.* , *Oryza sativa* ssp. indica genomic database *et al.* , and Syngenta's *O. sativa* ssp. japonica genomic database *et al.* . The GCG/Wisconsin package was used to perform FASTA searches, construct alignments (PILEUP), derive majority consensus sequences and calculate Kimura-2-parameter (K) distances. Phylogenetic analysis was performed with PAUP* version 4.0b8 (http://paup.csit.fsu.edu/) using the neighbor-joining method with default parameters. The time (T) since divergence was estimated using the formula T = K / 2r . The average substitution rate (r) used was 6.5 x 10^{-9} substitutions per synonymous site per year for grasses *et al.* .

RESULTS

MITE families *Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer*

Five *Tourist*-like MITE families, *Castaway*, *Explorer*, *Tou* and *Wanderer* that were among the first MITEs identified in the genome of *Oryza sativa* ssp. japonica were chosen for analysis *et al.* . The MITE-family *Olo* was originally isolated from *Oryza longistaminata* . Computer searches of the BAC end database *et al.* , Genbank and the *Oryza sativa* ssp. indica genomic database *et al.* revealed the presence of *Olo* in *O. sativa* ssp. japonica and *O sativa* ssp. indica and its close relationship to the *Tou* family (similarity between *Olo* and *Tou* consensus sequences is 66%). These MITE families have some of the highest copy numbers [*Castaway* (2500), *Explorer* (4000), *Tou* (2200), *Olo* (6800) and *Wanderer* (10000)] of all MITE families in the rice genome (Jiang, personal communication). *Oryza sativa* ssp. japonica c.v. Nipponbare BAC sequences from Genbank were searched for copies of the five MITE families using FASTA and the resulting sequences were aligned using PILEUP.

Alignments for each family including their respective majority consensus sequence are shown in Figures 2.1. The first three bases of the TIRs for all these families are GGG/GGC/GCC, except for *Tou* (GGT). Due to the limited amount of sequence information in public databases when *Explorer* was first described it was not possible at the time to determine TIRs and TSDs *et al.* . With the increased amount of sequence information currently available we were able to identify the TIRs for *Explorer*.

In the case of *Wanderer* a variable region of 35 bp was observed in the 3' subterminal region (Figure 2.1).

Development of transposon display (TD) display

In order to develop MITE based markers for the families *Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer* in the doubled-haploid (DH) mapping population derived from a cross between *O. sativa* ssp. indica (IR64) and *O. sativa* ssp. japonica (Azucena), TD was developed for these families. To amplify fragments anchored in these MITE families, pairs of nested primers were designed to recognize conserved regions of a MITE and which are directed outward from the element (Figures 2.1).

Genomic DNA from the parents of the DH mapping population *O. sativa* ssp. indica (IR64) and *O. sativa* ssp. japonica (Azucena) was used as template in TD. Comparison of TD profiles revealed that different MITE families showed characteristic levels of band polymorphism (Figure 2.2). The highest level of polymorphism was observed for *Tou* (55%), with the remaining elements ranging between 29-39% (Figure 2.2 and Table 2.1). Selected bands on the TD gel were cut out, reamplified, cloned, and sequenced. In all cases the sequences were found to be anchored in the correct element (Figure 2.3).

Chromosomal distribution of *Tourist*-like MITEs

TD was used to determine the map position of the MITEs from the five families. Two hundred and twenty five polymorphic bands from the *Castaway* (57), *Explorer* (34), *Olo* (36), *Tou* (44), and *Wanderer* (54) families were mapped using the IR64 x Azucena DH mapping population and a previously constructed framework map consisting of 276

SSR and 137 RFLP markers *et al.* (Figure 2.4 and 2.5, Table 2.2). The MITE markers were distributed over all 12 chromosomes without any significant clustering in any region.

MITEs in other AA genome species

The genus *Oryza* contains 23 different species, seven of which are tetraploid *et al.* . Several distinct genome types (A-K) have been identified cytologically. The youngest clade, including the cultivated species *O. sativa* and *O. glaberrima*, have genome type AA. In many cases the parental genome types of the tetraploid species could be determined (e.g. CCDD for *O. grandiglumis*) *et al.* . To test whether the element specific primers designed by using *O. sativa* ssp. japonica c.v. Nipponbare sequences could amplify MITE anchored fragments from other rice species, genomic DNA from several AA genome species as well as the CCDD genome species *O. grandiglumis* was used in TD analysis. Primers from all MITE families described above successfully amplified bands from all of the *Oryza* accessions including the more distantly related CCDD genome (Figure 2.2).

MITE sequences cluster into distinct subfamilies

It was of interest to determine phylogenetic relationships among MITEs and obtain a picture of the history of MITE transposition in the rice genome. To this end, MITE sequences found in the initial and successive FASTA searches were aligned and phylogenetic trees (PAUP) constructed for the MITE families *Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer* (Figure 2.6). Pairwise Kimura-2-parameter distances between groups of sequences were averaged and used to estimate their divergence times (Figure 2.6). Alignments of consensus sequences are shown in Figure 2.7.

Due to their sequence similarity (66%) *Olo* (Tou-B) and *Tou* (Tou-A) sequences were included in the same phylogenetic analysis, which revealed the presence of another major cluster, C. The three major clusters A, B and C contain several distinct subclusters of different ages (supported by bootstrap values) (Figure 2.6 and 2.7). The youngest cluster A1a is ~1 million years old, while the oldest, C3 is ~14 million years old. As can be seen from the alignment of consensus sequences (Figure 2.7) primers designed for *Olo* (Tou-B) and *Tou* (Tou-A) were specific for their respective clusters. The Tou-A primers were able to amplify elements from Tou-A1 and Tou-A2 and Tou-A3 but not Tou-A4.

Castaway has several clusters (A-D) (supported by bootstrap values) that appear to have amplified between ~7-10 million years ago, whereas F sequences appear older, having amplified ~17 million years ago. The ends of the recently identified *PIF*-like full length transposon sequences called *OsPIF6* were noted to have high similarity with *Castaway* sequences (Zhang, 2003) and were included in the analysis (Figure 2.6 and 2.7). The ends of *OsPIF6* sequences (Figure 2.7) are most similar to D sequences but are much younger than any of the MITE clusters (~300,000 years).

For *Explorer*, two major clusters (A and B) were identified. Subclusters A1 and B2 are ~17-20 million years old, while A1a and B3 appear to be ~5-6 million years. The sequence relationships are less complex for *Explorer* than was revealed for *Tou*/*Olo* and *Castaway*. The primers for TD of *Explorer* sequences recognize primarily A1 sequences, but could amplify divergent sequences of the other clusters as well (Figure 2.7).

Clusters for *Wanderer* (aside from A1a, A1b and A1c) are not supported by bootstrap values. The age for the A1 and A2 subclusters was estimated to be 19 and 22 million years, respectively (Figure 2.6). The three subclusters A1a, A1b and A1c are supported by bootstrap values and their average age was estimated to be ~2 million years. In successive FASTA searches, sequences with significant similarity to A1 and A2 were found and labeled B and C (Figure 2.7). The primers designed to amplify *Wanderer* elements in TD recognize A sequences only (Figure 2.7).

The similarity of sequences within major clusters indicated by upper case letters (A, B, …) was more than 75%. Similarities between sequences of major clusters could be as low as 62% (Tou-C/Tou-A, Cas-A/Cas-F), 65% (Exp-A/Exp-B) and 66% (Tou-B/Tou-A, Wan-A/Wan-C).

DISCUSSION

Previously, TD was developed for the *Hbr (Heartbreaker)* MITE family in maize *et al.* , for *Mutator* in maize (Frey *et al.*, 1998), for *dTph1* in petunia *et al.* , *mPing* in rice *et al.* *et al.* *et al.* *et al.* *et al.* *et al.* and *Stowaway-os1* and *Tabito II* in rice (Takagi, 2003). In this work, sequences from public databases were used to develop primers to recognize conserved regions of the high copy number *Tourist*-like MITE families *Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer* for TD and marker development. Successful amplification of these MITE families suggest, that with enough draft sequence information, TD can be quickly and routinely applied to new MITE families. For all MITE families described in this study, the same protocol was used, including identical PCR conditions. The high success rate indicates that the most important factor in developing TD is the design and choice of the element specific primers.

Polymorphic bands on TD are either due to the presence or absence of a MITE at a locus in the two parents (insertion site polymorphism) or due to restriction site polymorphism for the enzyme used in the restriction/ligation reaction. The level of band polymorphism observed in this study ranged between 29-55% (Table 2.1). An estimate of 21.8% restriction site polymorphism had been determined previously for the strains used in my analysis *et al.* . Using this number and the band polymorphism values from Table 2.1, the level of insertion site polymorphism for each family can be calculated. Accordingly the estimated level of insertion site polymorphism for *Castaway* is 7.3%, for *Explorer* 13.5%, *Olo* 17.8%, *Tou* 33.4% and *Wanderer* 15.4%. Therefore in the cases of *Castaway*, *Explorer*, *Olo* and *Wanderer* most of the visible band polymorphism can be attributed to restriction site polymorphism, while for *Tou* most of the band polymorphism can be attributed to insertion site polymorphism. The primers for TD of *Tou* were designed to specifically recognize Tou-A1a sequences (Figure 2.6). Tou-A1a is the youngest (~1 million years) of all the clusters observed in this study (Figure 2.6). A higher level of insertion site polymorphism is to be expected for comparatively young MITE sequences.

The primers developed for TD in *O. sativa* successfully amplified fragments from several other rice species (Figure 2.2). Thus, TD can potentially be used for population genetic and phylogenetic studies in a wide range of *Oryza* species. TD of *Hbr* *(Heartbreaker)* was successfully used to determine the genetic relationships among maize inbred lines *et al.* . Besides intraspecific mapping populations between *O. sativa* ssp. japonica and *O. sativa* ssp. indica lines, interspecific mapping populations derived from crosses between *O. sativa* x *O. longistaminata* *et al.* , *O. sativa* x *O. glaberrima* *et al.* and *O. sativa* x *O. glumaepatula* *et al.* have been developed. The data presented in this study indicate that the primers designed for761 cm BT 50 0 0.996265 Tc q 0.24 0 0 -0.24 31 761 cm BT 50 0 0 -5

odcifiations

rgtions

mrkeors

ssp. japonica

rs

povsides

al. *et al.* *et al.* *et al.* *et al.* . Sequences within superfamilies that have similarities over their entire length are grouped into families. Previously a family was defined as a collection of MITE sequences sharing >85% similarity over their entire length *et al.* .

The results from this study show that sequences with >75% similarity over their entire length were found to group in distinct major clusters labeled with upper case letters (A, B, …) (Figure 2.6). Sequences of different major clusters share similarities as low as 62% (Tou-C/Tou-A, Cas-A/Cas-F) at which level their common ancestry is still apparent (Figure 2.7). Using these results a family of MITEs could be redefined as a collection of sequences sharing >60% sequence similarity. Alternatively, in keeping with the previously proposed definition *et al.* , sequences of major clusters (A, B, …) (>75% similarity) would constitute a family, while related major clusters (>60% similarity) a family group. This latter scheme is used for further discussion in which for example Tou-A, Tou-B and Tou-C each is addressed as a family. Subfamilies are defined as sequences forming distinct clusters within families as detected in the phylogenetic trees and are indicated by numbers (e.g. Tou-A4, Cas-A3).

The phylogenetic analysis reveals the existence of several related MITE families (Cas-A-F, Exp-A-B, Tou-A-C, Wan-A-C) each of which contains multiple subfamilies, whose ages range between 1-22 million years (Figure 2.6). The presence of distinct families and subfamilies of varying age indicates temporal waves of MITE amplification during the last 20 million years. In contrast, the age of the genus *Oryza* is estimated to be around 40 million years, while the age of the AA genome-species is at least 4 million (Tao Sang personal communication) (Figure 1.2). Comparison of these values suggests that the MITE subfamilies analyzed in this study were amplified after the divergence of the genus *Oryza* and, for most, before the split of the AA genome species from their common ancestor. Since age estimates are averaged over many elements, amplification of some families may have occurred prior to and after these times.

Overall the age of subfamilies is consistent with the fact that TD could amplify MITEs from all the AA genome species including a CCDD genome species (Figure 2.2) and also that most of the band polymorphism of TD with two AA genome species is due to restriction site polymorphism and not insertion site polymorphism (except for the youngest subfamily Tou-A1a). The identification of the active MITE family *mPing* (Jiang *et al*., 2003) indicates that the process that led to the amplification of MITEs is still ongoing after more than 20 million years.

Some MITEs are thought to have originated as deletion derivatives of autonomous DNA transposons (Feschotte *et al*., 2002). Elements of the *Stowaway*-like and *Tourist*-like MITE superfamilies could be related to plant *mariner*-like elements (MLEs) and *PIF*/*Pong*-like elements based on TIR sequence similarities and shared TSDs *et al.* *et al.* . However strong relationships between the ends of autonomous elements and full length MITE sequences are rare for both MITE superfamilies in rice *et al.* *et al.* . For the MITEs studied here only *Castaway* sequences showed high similarity over their whole length to the ends of an autonomous transposon called *OsPIF6* *et al.* . Sequences of *OsPIF6* are most closely related to Cas-D sequences (Figure 2.6 and 2.7). The recent divergence time of *OsPIF6* sequences (~300,000 years) indicates that the activity of these elements was not responsible for the amplification of *Castaway* subfamilies most of which are at least 5 million years old.

To explain the lack of clear-cut relationships between DNA transposons and most of the MITE families, it has been proposed that transposons are lost stochastically from the rice genome whereas MITEs persist in the genome due to their much higher copy numbers (Figure 1.1) *et al.* . Alternatively, MITEs could originate *de novo* when TIRs flanking unrelated sequences are recognized by a transposase, in which case no associated transposon sequences exist in the first place *et al.* .

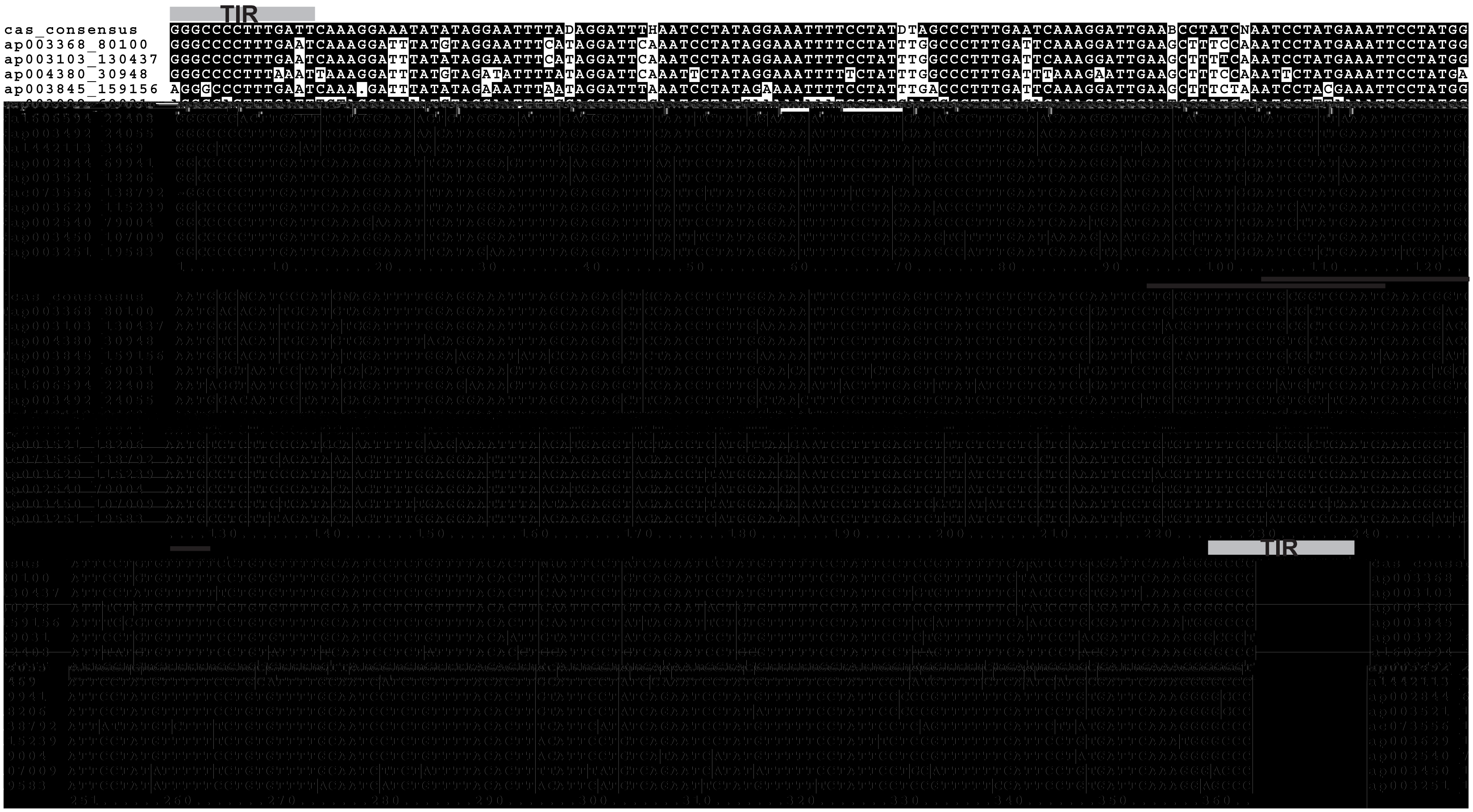
Even though candidate transposon sequences responsible for the amplification of the observed subfamilies could not be identified, their very existence indicates the presence of a transposase source several times during the last 20 million years. It has been suggested that cross-mobilization is a major mechanism for MITE amplification in the rice genome *et al.* . Cross-mobilization means that a transposase encoded by a related element recognizes and transposes a MITE sequence due to the fortuitously recognition of its TIRs.

Thus the following scenario could explain the patterns observed (Figure 2.6). An ancestral MITE sequence originates by internal deletion from an autonomous element that might have been responsible for the first wave of amplification. After loss of the transposon from the genome, temporal successive waves of MITE amplification would have been driven by related transposons active at different points in time.

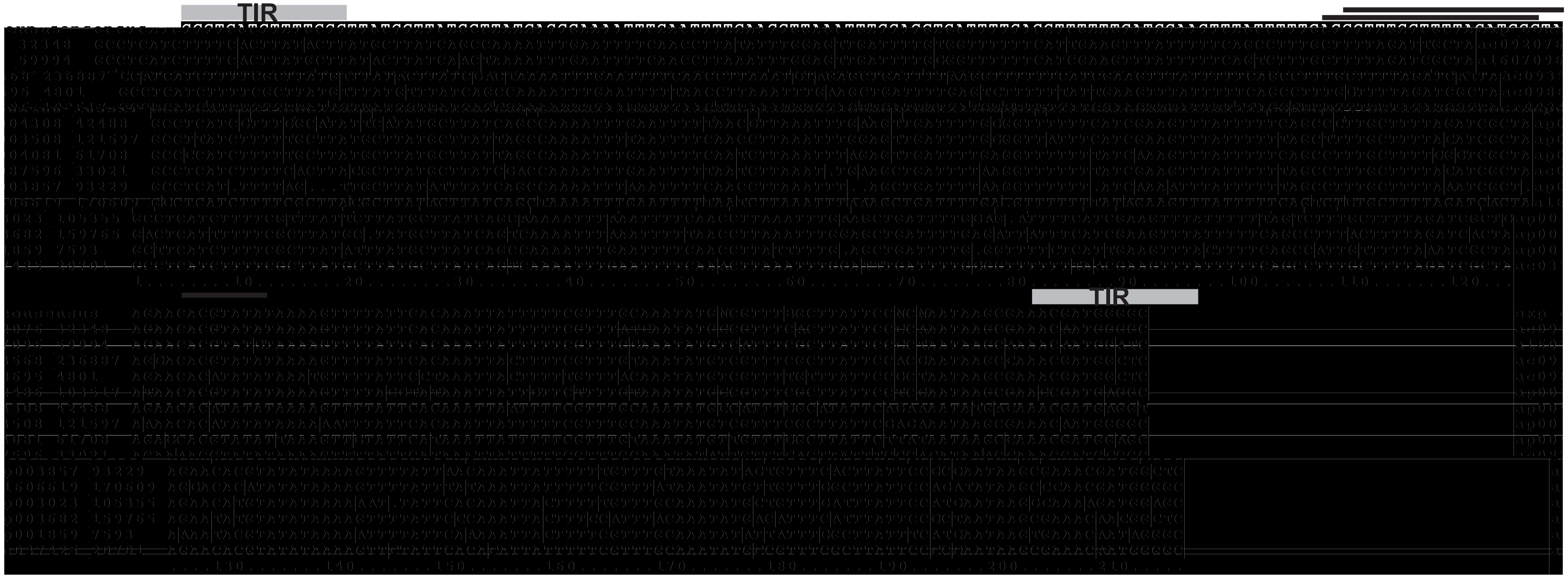
The recognition of TIRs seems to be very specific as suggested by the amplification of Tou-A1a from a single ancestral element out of many present Tou-A1 sequences that were not amplified (Figure 2.6). Similar cases observed are the recognition and amplification of the ancestral sequences of Wan-A1a,b,c while many candidate Wan-A1 sequences were not recognized and also Exp-A1a/Exp-A1 (Figure 2.6). The ancestral TIR sequences of Exp-A1a, Tou-A1a and Wan-A1a,b,c differ from those of Exp-A1, Tou-A1 and Wan-A1 by single base substitutions (Figure 2.7). These observations suggest that single base substitutions in the TIRs can determine whether or not a MITE sequence can be activated.

Figure 2.1: Alignment of *Tourist*-like MITE families *Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer*. The positions of the sequences recognized by the TD primers are indicated with black bars above the alignment. Grey boxes indicate TIRs and a bracket defines the variable region of *Wanderer*.

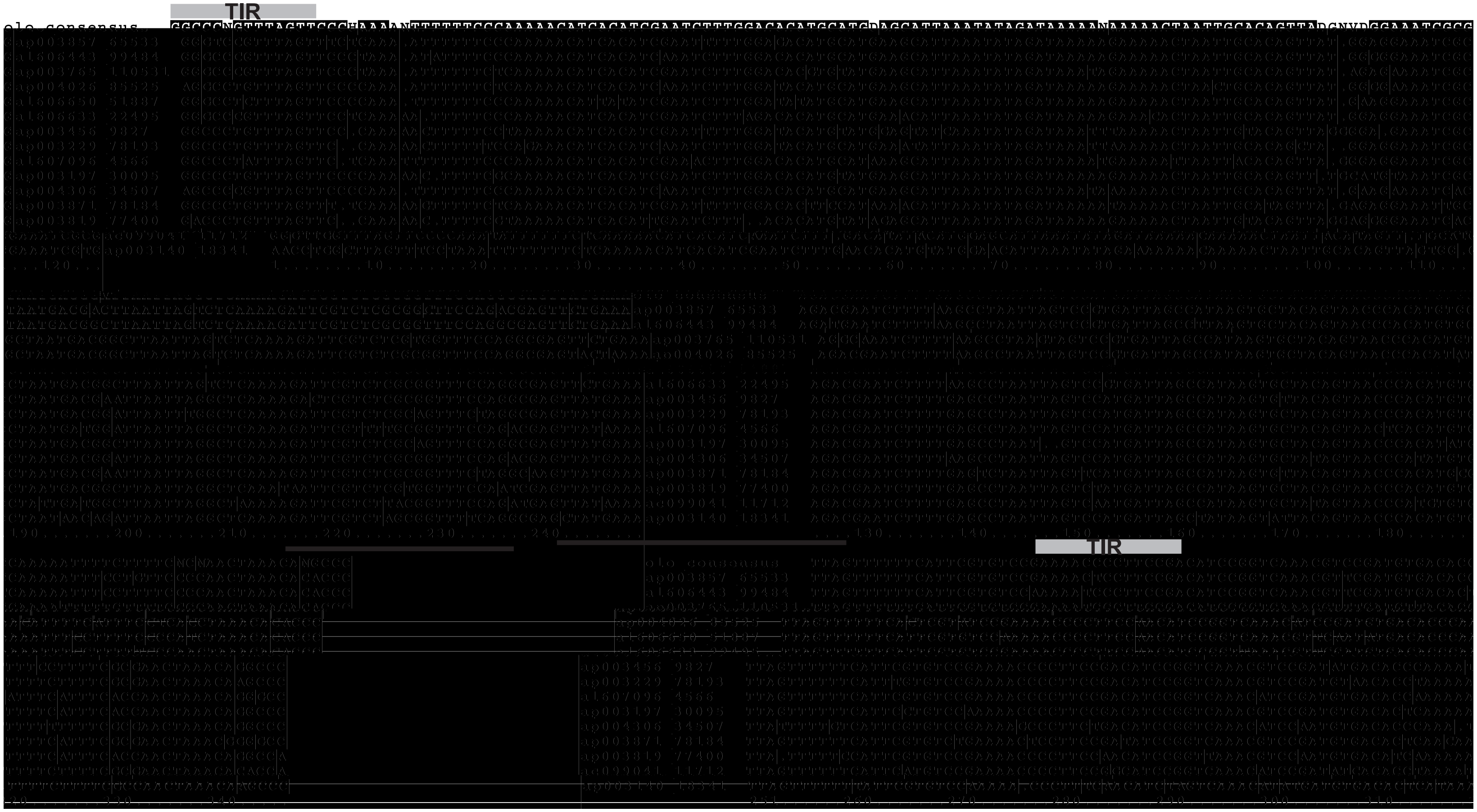
Castaway


TIR
cas_consensus GGGCCCCTTTGATTCAAAGGAAATATATAGGAATTTTADAGGATTTHAATCCTATAGGAAATTTTCCTATDTAGCCCTTTGAATCAAAGGATTGAABCCTATCNAATCCTATGAAATTCCTATGG
ap003368_80100 GGGCCCCTTTGAATCAAAGGATTTATGTAGGAATTTCATAGGATTCAAATCCTATAGGAAATTTTCCTATTTGGCCCTTTGATTCAAAGGATTGAAGCTTTCCAAATCCTATGAAATTCCTATGG
ap003103_130437 GGGCCCCTTTGAATCAAAGGATTTATATAGGAATTTCATAGGATTCAAATCCTATAGGAAATTTTCCTATTTGGCCCTTTGATTCAAAGGATTGAAGCTTTTCAAATCCTATGAAATTCCTATGG
ap004380_30948 GGGCCCCTTTAAATTAAAGGATTTATGTAGATATTTTATAGGATTCAAATTCTATAGGAAATTTTTCTATTTGGCCCTTTGATTTAAAGAATTGAAGCTTTCCAAATTCTATGAAATTCCTATGA
ap003845_159156 AGGGCCCTTTGAATCAAA.GATTTATATAGAAATTTAATAGGATTTAAATCCTATAGAAAATTTTCCTATTTGACCCTTTGATTCAAAGGATTGAAGCTTTCTAAATCCTACGAAATTCCTATGG
TIR

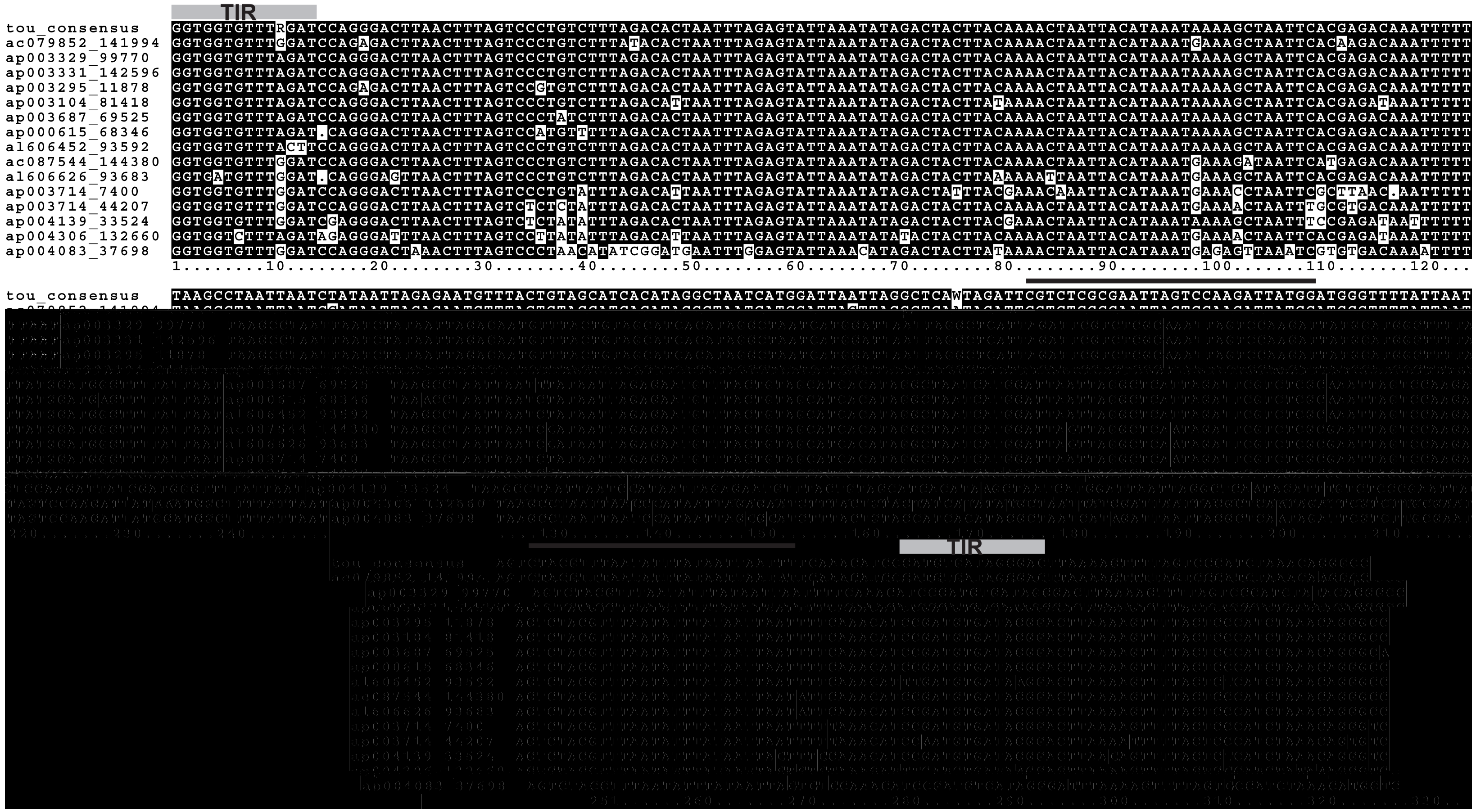
Explorer



Olo


TIR
olo consensus
GGCCCNGTTTAGTTCCCHAAAANTTTTTTCCCAAAAACATCACATCGAATCTTTGGACACATGCATGDAGCATTAAATATAGATAAAAANAAAAACTAATTGCACAGTTADGNVDGGAAATCGCG
TIR

Tou


TIR
tou_consensus
ac079852_141994
ap003329_99770
ap003331_142596
ap003295_11878
ap003104_81418
ap003687_69525
ap000615_68346
al606452_93592
ac087544_144380
al606626_93683
ap003714_7400
ap003714_44207
ap004139_33524
ap004306_132660
ap004083_37698
TIR

Wanderer

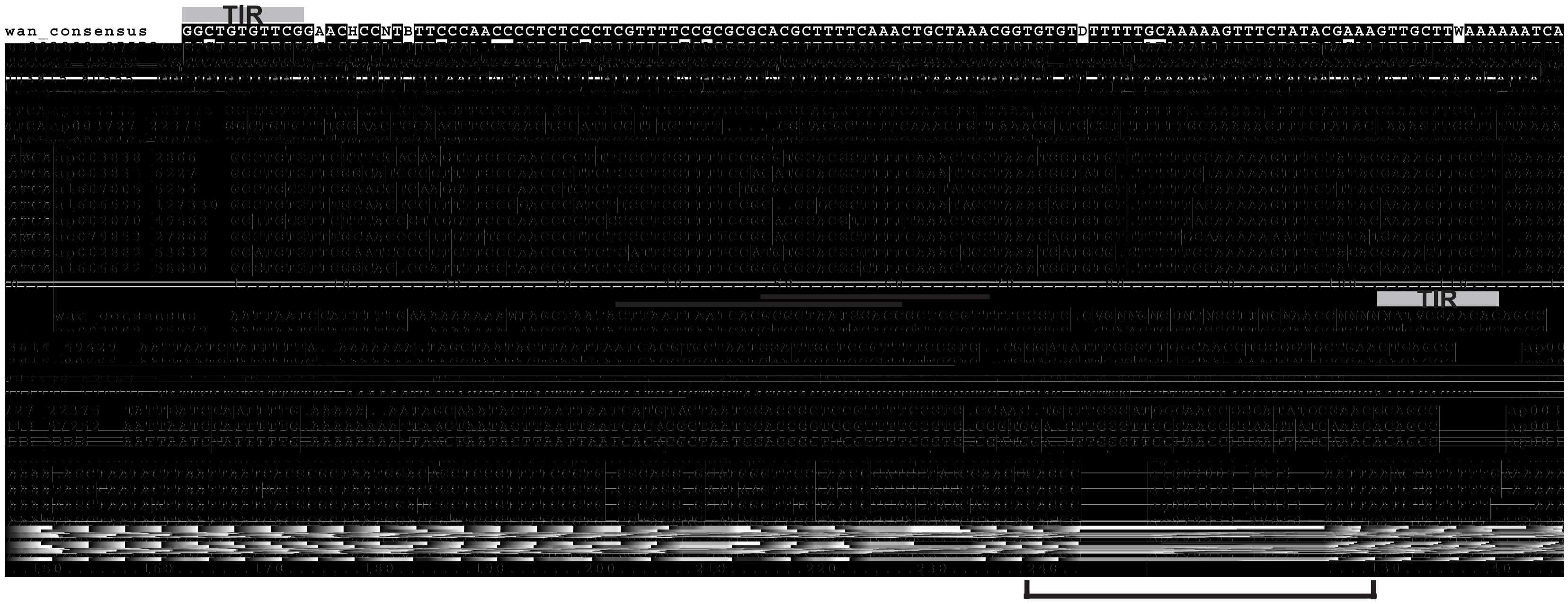
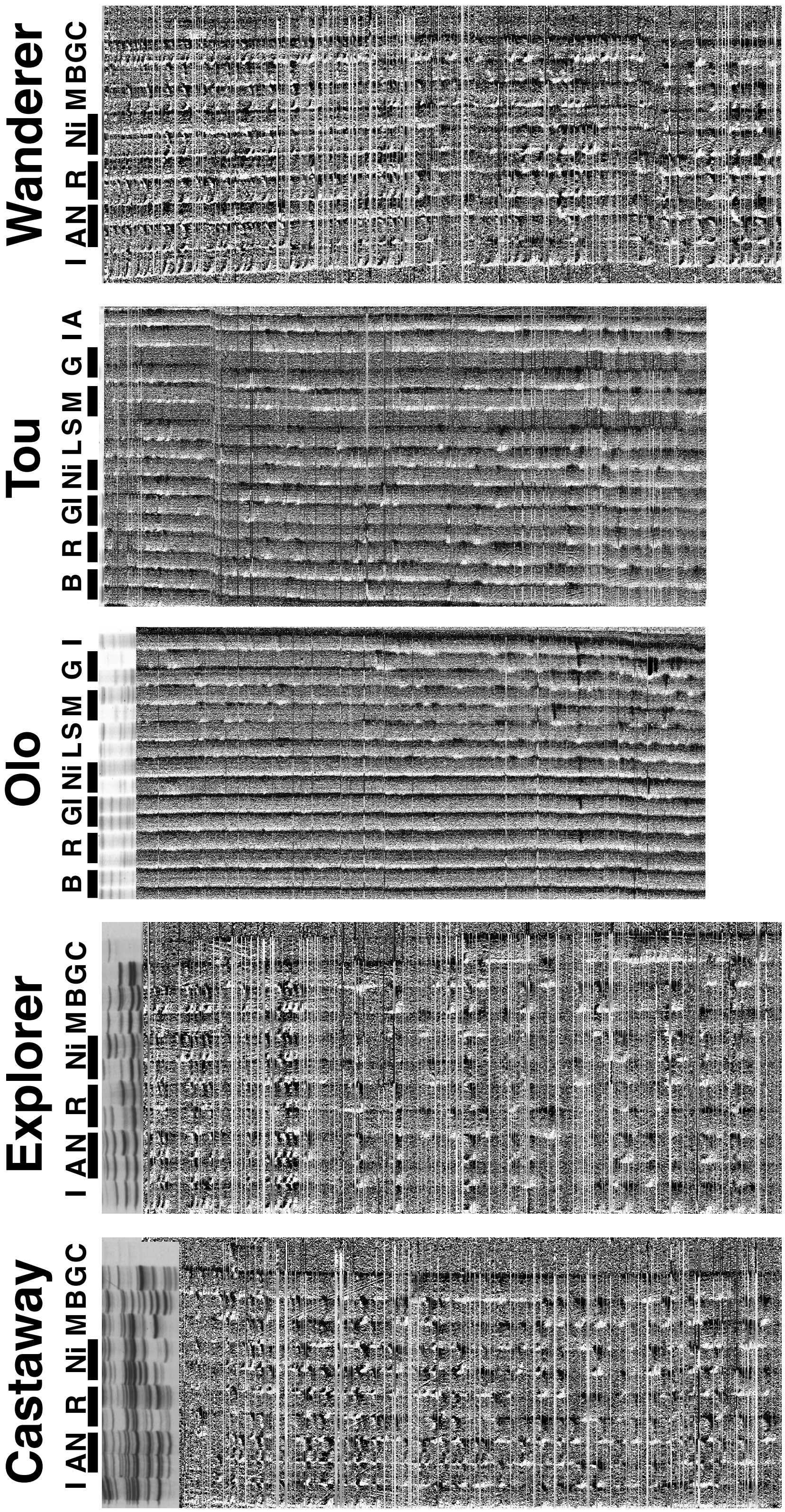
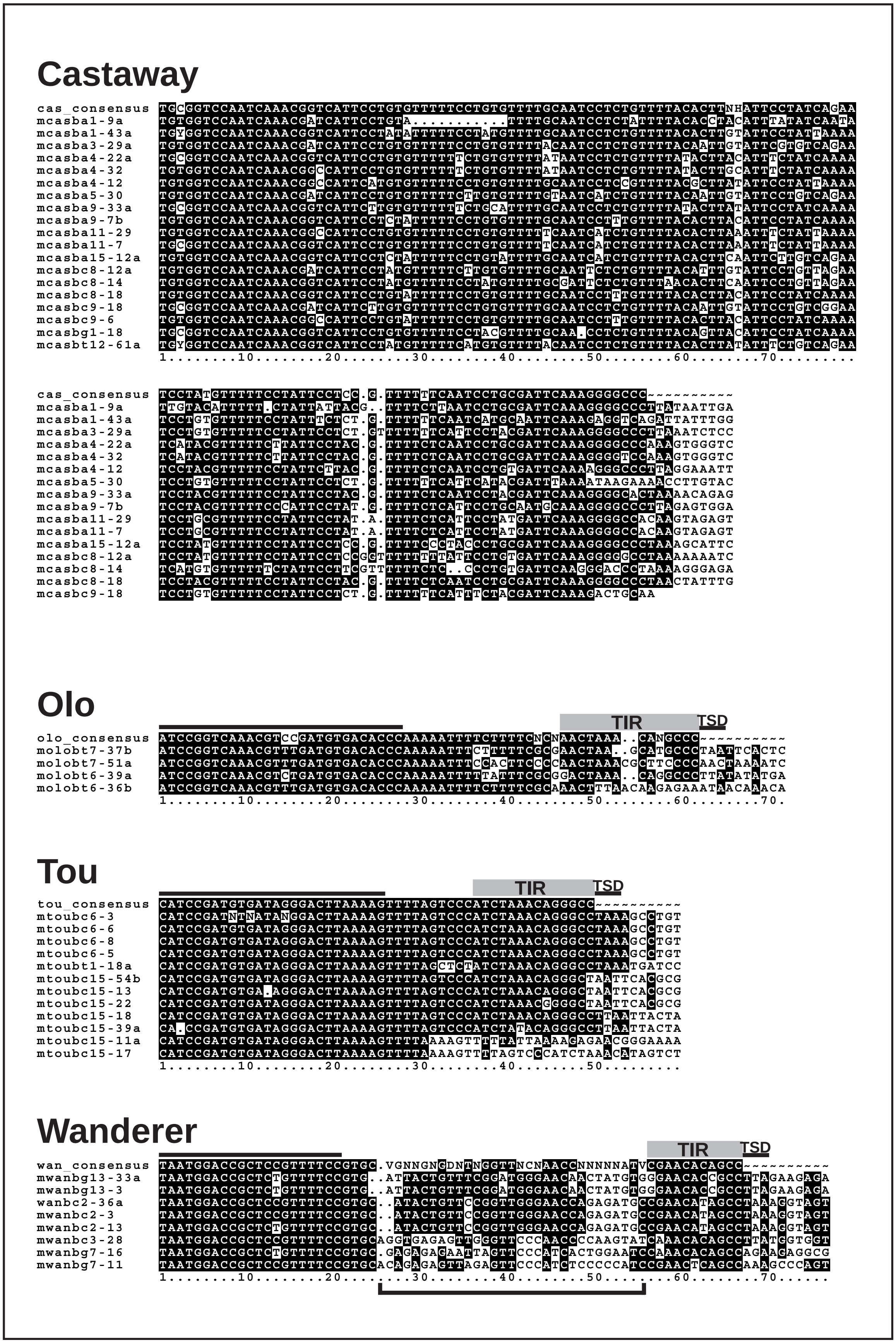

TIR
wan_consensus
GGCTGTGTTCGGAACHCCNTBTTCCCAACCCCTCTCCCTCGTTTTCCGCGCGCACGCTTTTCAAACTGCTAAACGGTGTGTDTTTTTGCAAAAAGTTTCTATACGAAAGTTGCTTWAAAAAATCA
TIR
wan_consensus

Figure 2.2: TD of different *Tourist*-like MITE families (*Castaway*, *Explorer*, *Olo*, *Tou* and *Wanderer*). Template DNA from the following species were used: I - *O. sativa* ssp. indica (IR64), A - *O. sativa* ssp. japonica (Azucena), N - *O. sativa* ssp. japonica, R - *O. rufipogon*, Ni - *O. nivara*, M - *O. meridionalis*, G - *O. grandiglumis*, C – Coix. The adapter specific primer in the selective amplification was *Bfa*I+A.



Castaway
I AN R Ni M B G C
Explorer
I AN R Ni M B G C
Olo
B R Gl Ni L S M G I
Tou
B R Gl Ni L S M G I A
Wanderer
I AN R Ni M B G C

Figure 2.3: Alignment of MITE anchored sequences isolated from TD gels. Bars above the alignment indicate primer positions. The variable region of *Wanderer* is highlighted by a bracket underneath the alignment.



Castaway

```
cas_consensus   TGCGGTCCAATCAAACGGTCATTCCTGTGTTTTTCCTGTGTTTTGCAATCCTCTGTTTTACACTTNHATTCCTATCAGAA
mcasba1-9a      TGTGGTCCAATCAAACGATCATTCCTGTA...........TTTTGCAATCCTCTATTTTACACCTACATTTATATCAATA
mcasba1-43a     TGYGGTCCAATCAAACGGTCATTCCTATATTTTTCCTATGTTTTGCAATCCTCTGTTTTACACTTGTATTCCTATTAAAA
mcasba3-29a     TGTGGTCCAATCAAACGATCATTCCTGTGTTTTTCCTGTGTTTTACAATCCTCTGTTTTACAATTGTATTCGTGTCAGAA
mcasba4-22a     TGCGGTCCAATCAAACGGTCATTCCTGTGTTTTTTCTGTGTTTTATAATCCTCTGTTTTATACTTACATTTCTATCAAAA
mcasba4-32      TGTGGTCCAATCAAACGGCCATTCCTGTGTTTTTTCTGTGTTTTATAATCCTCTGTTTTATACTTGCATTTCTATCAAAA
mcasba4-12      TGTGGTCCAATCAAACGGCCATTCATGTGTTTTTCCTGTGTTTTGCAATCCTCCGTTTTACGCTTATATTCCTATTAAAA
mcasba5-30      TGTGGTCCAATCAAACGATCATTCCTGTGTTTTTCTTGTGTTTTGTAATCATCTGTTTTACAATTGTATTCCTGTCAGAA
mcasba9-33a     TGCGGTCCAATCAAACGGTCATTCTTGTGTTTTTTCTGCATTTTGCAATCCTCTGTTTTATACTTATATTCCTATCAAAA
mcasba9-7b      TGTGGTCCAATCAAACGGTCATTCCTCTATTTTTCCTGTGTTTTGCAATCCTTTGTTTTACACTTACATTCCTATCAAAA
mcasba11-29     TGTGGTCCAATCAAACGGCCATTCCTGTGTTTTTCCTGTGTTTTTCAATCATCTGTTTTACACTTAAATTTCTATTAAAA
mcasba11-7      TGCGGTCCAATCAAACGGTCATTCCTGTGTTTTTCCTGTGTTTTTCAATCATCTGTTTTACACTTAAATTTCTATTAAAA
mcasba15-12a    TGTGGTCCAATCAAACGGTCATTCCTCTATTTTTCCTGTATTTTGCAATCATCTGTTTTACACTTCAATTCTTGTCAGAA
mcasbc8-12a     TGTGGTCCAATCAAACGATCATTCCTATGTTTTTCTTGTGTTTTGCAATTCTCTGTTTTACATTTGTATTCCTGTTAGAA
mcasbc8-14      TGTGGTCCAATCAAACGGTCATTCCTATGTTTTTCCTATGTTTTGCGATTCTCTGTTTAACACTTCAATTCCTGTTAGAA
mcasbc8-18      TGTGGTCCAATCAAACGGTCATTCCTGTATTTTTCCTGTGTTTTGCAATCCTTTGTTTTACACTTACATTCCTATCAAAA
mcasbc9-18      TGCGGTCCAATCAAACGATCATTCTTGTGTTTTTCCTGTGTTTTGCAATCCTCTGTTTTACAATTGTATTCCTGTCGGAA
mcasbc9-6       TGTGGTCCAATCAAACGGCCATTCCTGTATTTTTCCTGTGTTTTGCAATCCTTTGTTTTACACTTACATTCCTATCAAAA
mcasbg1-18      TGCGGTCCAATCAAACGGTCATTCCTGTGTTTTTCCTACGTTTTGCAA.CCTCTGTTTTACAGTTACATTCCTATCAAAA
mcasbt12-61a    TGYGGTCCAATCAAACGGTCATTCCTATGTTTTTCATGTGTTTTACAATCCTCTGTTTTACACTTATATTTCTGTCAGAA
                1........10........20........30........40........50........60........70.........
```

```
cas_consensus   TCCTATGTTTTTCCTATTCCTCC.G.TTTTTTCAATCCTGCGATTCAAAGGGGCCC~~~~~~~~~~
mcasba1-9a      TTGTACATTTTT.CTATTATTACG..TTTTCTTAATCCTGCGATTCAAAGGGGCCCTTATAATTGA
mcasba1-43a     TCCTGTGTTTTTCCTATTTCTCT.G.TTTTTTCAATCATGCAATTCAAAGAGGTCAGATTATTTGG
mcasba3-29a     TCCTGTGTTTTTCCTATTCCTCT.GTTTTTTTCATTCCTACGATTCAAAGGGGCCCTTAAATCTCC
mcasba4-22a     TCATACGTTTTTCTTATTCCTAC.G.TTTTCTCAATCCTGCGATTCAAAGGGGCCCAAAGTGGGTC
mcasba4-32      TCATACGTTTTTCTTATTCCTAC.G.TTTTCTCAATCCTGCGATTCAAAGGGGTCCAAAGTGGGTC
mcasba4-12      TCCTACGTTTTTCCTATTCTTAC.G.TTTTCTCAATCCTGTGATTCAAAAGGGCCCTTAGGAAATT
mcasba5-30      TCCTGTGTTTTTCCTATTCCTCT.G.TTTTTTCATTCATACGATTTAAAATAAGAAAACCTTGTAC
mcasba9-33a     TCCTACGTTTTTCCTATTCCTAC.G.TTTTCTCAATCCTACGATTCAAAGGGGCACTAAAACAGAG
mcasba9-7b      TCCTACGTTTTTCCCATTCCTAT.G.TTTTCTCATTCCTGCAATGCAAAGGGGCCCTTAGAGTGGA
mcasba11-29     TCCTGCGTTTTTCCTATTCCTAT.A.TTTTCTCATTCCTATGATTCAAAGGGGCCACAAGTAGAGT
mcasba11-7      TCCTGCGTTTTTCCTATTCCTAT.A.TTTTCTCATTCCTATGATTCAAAGGGGCCACAAGTAGAGT
mcasba15-12a    TCCTATGTTTTTCCTATTCCTCC.G.TTTTCCCTACCCTGCGATTCAAAGGGGCCCTAAAGCATTC
mcasbc8-12a     TCCTATGTTTTTCCTATTCCTCCGGTTTTTTTTATTCCTGTGATTCAAAGGGGGCCTAAAAAAATC
mcasbc8-14      TCATGTGTTTTTTCTATTCCTTCGTTTTTTCTC..CCCTGTGATTCAAGGGACCCTAAAAGGGAGA
mcasbc8-18      TCCTACGTTTTTCCTATTCCTAC.G.TTTTCTCAATCCTGCGATTCAAAGGGGCCCTAACTATTTG
mcasbc9-18      TCCTGTGTTTTTCCTATTCCTCT.G.TTTTTTCATTTCTACGATTCAAAGACTGCAA
```

Olo

```
olo_consensus   ATCCGGTCAAACGTCCGATGTGACACCCAAAAATTTTCTTTTCNCNAACTAAA..CANGCCC~~~~~~~~~~
molobt7-37b     ATCCGGTCAAACGTTTGATGTGACACCCAAAAATTTCTTTTTCGCGAACTAA..GCATGCCCTAATTCACTC
molobt7-51a     ATCCGGTCAAACGTTTGATGTGACACCCAAAAATTTCCACTTCCCCAACTAAACGCTTCCCCAACTAAAATC
molobt6-39a     ATCCGGTCAAACGTCTGATGTGACACCCAAAAATTTTTATTTCGCGGACTAAA..CAGGCCCTTATATATGA
molobt6-36b     ATCCGGTCAAACGTTTGATGTGACACCCAAAAATTTTCTTTTCGCAAACTTTAACAAGAGAAATAACAAACA
                1........10........20........30........40........50........60........70.
```

Tou

```
tou_consensus   CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCC~~~~~~~~~~
mtoubc6-3       CATCCGATNTNATANGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCCTAAAGCCTGT
mtoubc6-6       CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCCTAAAGCCTGT
mtoubc6-8       CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCCTAAAGCCTGT
mtoubc6-5       CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCCTAAAGCCTGT
mtoubt1-18a     CATCCGATGTGATAGGGACTTAAAAGTTTTAGCTCTATCTAAACAGGGCCTAAATGATCC
mtoubc15-54b    CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCTAATTCACGCG
mtoubc15-13     CATCCGATGTGA.AGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCTAATTCACGCG
mtoubc15-22     CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACGGGGCTAATTCACGCG
mtoubc15-18     CATCCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTAAACAGGGCCTTAATTACTA
mtoubc15-39a    CA.CCGATGTGATAGGGACTTAAAAGTTTTAGTCCCATCTATACAGGGCCTTAATTACTA
mtoubc15-11a    CATCCGATGTGATAGGGACTTAAAAGTTTTAAAAGTTTTATTAAAAGAGAACGGGAAAA
mtoubc15-17     CATCCGATGTGATAGGGACTTAAAAGTTTTAAAAGTTTTAGTCCCATCTAAACATAGTCT
                1........10........20........30........40........50.........
```

Wanderer

```
wan_consensus   TAATGGACCGCTCCGTTTTCCGTGC.VGNNGNGDNTNGGTTNCNAACCNNNNNATVCGAACACAGCC~~~~~~~~~~
mwanbg13-33a    TAATGGACCGCTCTGTTTTCCGTG..ATTACTGTTTCGGATGGGAACAACTATGTGGGAACACCGCCTTAGAAGAGA
mwanbg13-3      TAATGGACCGCTCTGTTTTCCGTG..ATTACTGTTTCGGATGGGAACAACTATGTGGGAACACCGCCTTAGAAGAGA
wanbc2-36a      TAATGGACCGCTCCGTTTTCCGTGC..ATACTGTTCCGGTTGGGAACCAGAGATGCCGAACATAGCCTAAAGGTAGT
mwanbc2-3       TAATGGACCGCTCCGTTTTCCGTGC..ATACTGTTCCGGTTGGGAACCAGAGATGCCGAACATAGCCTAAAGGTAGT
mwanbc2-13      TAATGGACCGCTCTGTTTTCCGTGC..ATACTGTTCCGGTTGGGAACCAGAGATGCCGAACATAGCCTAAAGGTAGT
mwanbc3-28      TAATGGACCGCTCCGTTTTCCGTGCAGGTGAGAGTTGGGTTCCCAACCCCAAGTATCAAACACAGCCTTATGGTGGT
mwanbg7-16      TAATGGACCGCTCTGTTTTCCGTGC.GAGAGAGAATTAGTTCCCATCACTGGAATCCAAACACAGCCAGAAGAGGCG
mwanbg7-11      TAATGGACCGCTCCGTTTTCCGTGCACAGAGAGTTAGAGTTCCCATCTCCCCCATCCGAACTCAGCCAAAGCCCAGT
                1........10........20........30........40........50........60........70......
```

Figure 2.4: Mapping of *Olo* elements. I and A label lanes with samples from the two parents of the mapping population, *O. sativa* ssp. indica (IR64) and *O. sativa* ssp. japonica (Azucena), respectively. DH lanes contain samples of 24 lines of the doubled-haploid (DH) mapping population. Bands which could be mapped are labeled with their corresponding marker name. The adapter specific primer in the selective amplification was *Bfa*I+C.

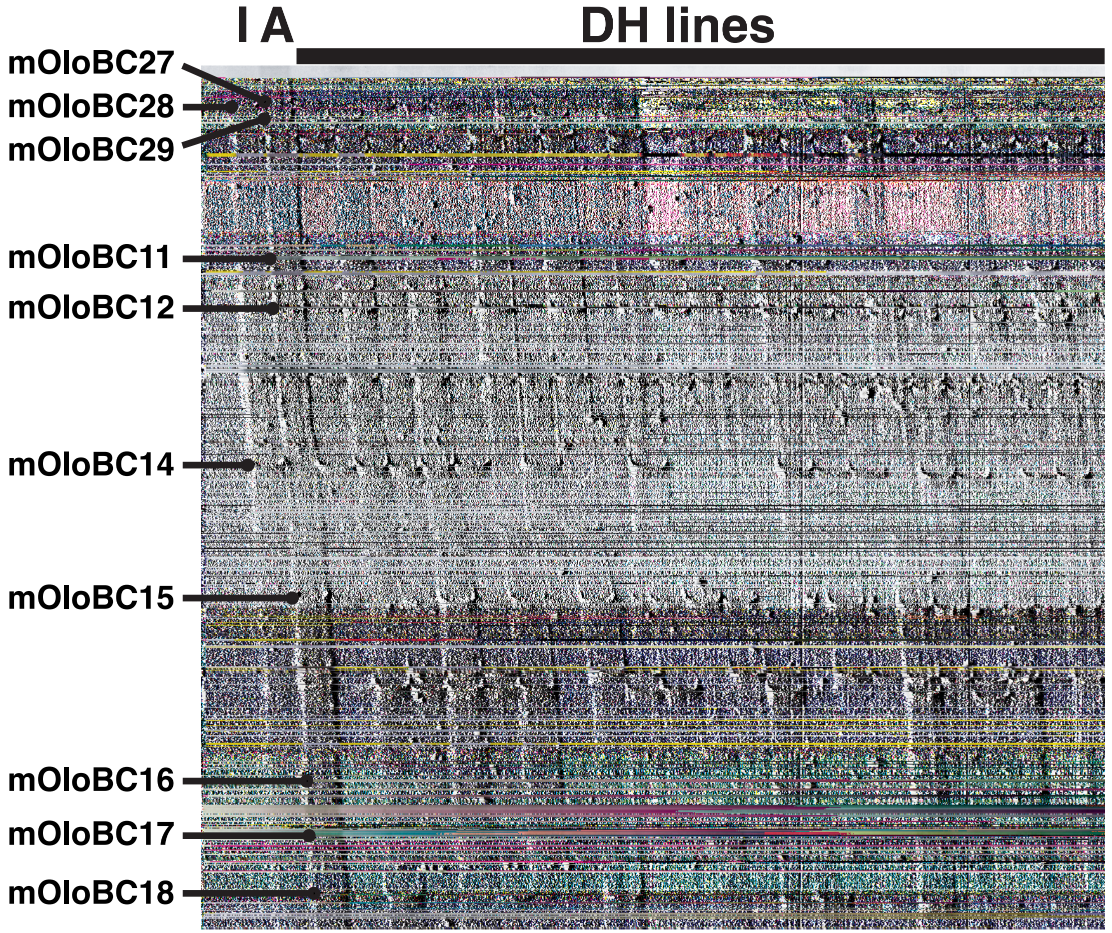
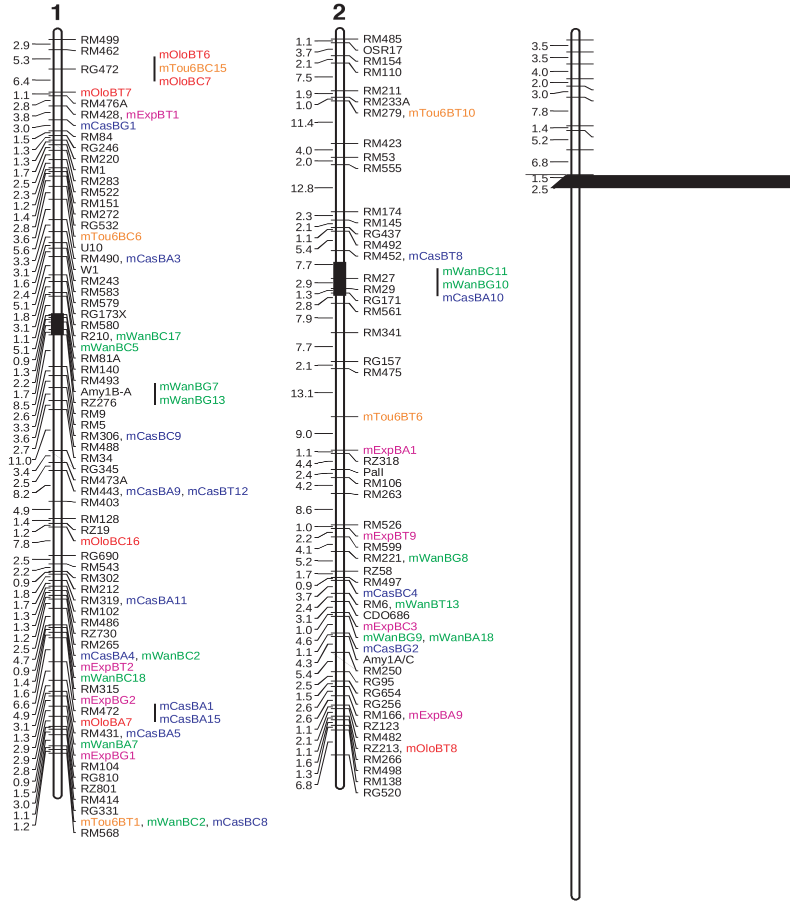

I A
DH lines
mOloBC27
mOloBC28
mOloBC29
mOloBC11
mOloBC12
mOloBC14
mOloBC15
mOloBC16
mOloBC17
mOloBC18

Figure 2.5: Genetic map of MITE markers. *Castaway*-markers are shown in blue, *Explorer*-markers purple, *Olo*-markers red, *Tou*-markers orange, and *Wanderer*-markers green. The position of centromeres is indicated with black boxes.


1
2.9 5.3 6.4 1.1 2.8 3.8 3.0 1.5 1.3 1.3 1.7 2.5 2.3 1.2 1.4 2.8 3.6 5.6 3.3 3.1 1.6 2.4 5.1 1.8 3.1 1.1 5.1 0.9 1.3 2.2 1.7 8.5 2.6 3.3 3.6 2.7 11.0 3.4 2.5 8.2 4.9 1.4 1.2 7.8 2.5 2.2 0.9 1.8 1.7 1.3 1.3 1.2 2.5 4.7 0.9 1.4 1.6 6.6 4.9 3.1 1.3 2.9 2.9 2.8 0.9 1.5 3.0 1.1 1.2
RM499
RM462
RG472
mOloBT6
mTou6BC15
mOloBC7
mOloBT7
RM476A
RM428, mExpBT1
mCasBG1
RM84
RG246
RM220
RM1
RM283
RM522
RM151
RM272
RG532
mTou6BC6
U10
RM490, mCasBA3
W1
RM243
RM583
RM579
RG173X
RM580
R210, mWanBC17
mWanBC5
RM81A
RM140
RM493
Amy1B-A
RZ276
mWanBG7
mWanBG13
RM9
RM5
RM306, mCasBC9
RM488
RM34
RG345
RM473A
RM443, mCasBA9, mCasBT12
RM403
RM128
RZ19
mOloBC16
RG690
RM543
RM302
RM212
RM319, mCasBA11
RM102
RM486
RZ730
RM265
mCasBA4, mWanBC2
mExpBT2
mWanBC18
RM315
mExpBG2
RM472
mOloBA7
mCasBA1
mCasBA15
RM431, mCasBA5
mWanBA7
mExpBG1
RM104
RG810
RZ801
RM414
RG331
mTou6BT1, mWanBC2, mCasBC8
RM568
2
1.1 3.7 2.1 7.5 1.9 1.0 11.4 4.0 2.0 12.8 2.3 2.1 1.1 5.4 7.7 2.9 1.3 2.8 7.9 7.7 2.1 13.1 9.0 1.1 4.4 2.4 4.2 8.6 1.0 2.2 4.1 5.2 1.7 0.9 3.7 2.4 3.1 1.0 4.6 1.1 4.3 5.4 2.5 1.5 2.6 2.6 1.1 2.1 1.1 1.6 1.3 6.8
RM485
OSR17
RM154
RM110
RM211
RM233A
RM279, mTou6BT10
RM423
RM53
RM555
RM174
RM145
RG437
RM492
RM452, mCasBT8
RM27
RM29
RG171
RM561
mWanBC11
mWanBG10
mCasBA10
RM341
RG157
RM475
mTou6BT6
mExpBA1
RZ318
Pall
RM106
RM263
RM526
mExpBT9
RM599
RM221, mWanBG8
RZ58
RM497
mCasBC4
RM6, mWanBT13
CDO686
mExpBC3
mWanBG9, mWanBA18
mCasBG2
Amy1A/C
RM250
RG95
RG654
RG256
RM166, mExpBA9
RZ123
RM482
RZ213, mOloBT8
RM266
RM498
RM138
RG520
3.5 3.5 4.0 2.0 3.0 7.8 1.4 5.2 6.8 1.5 2.5

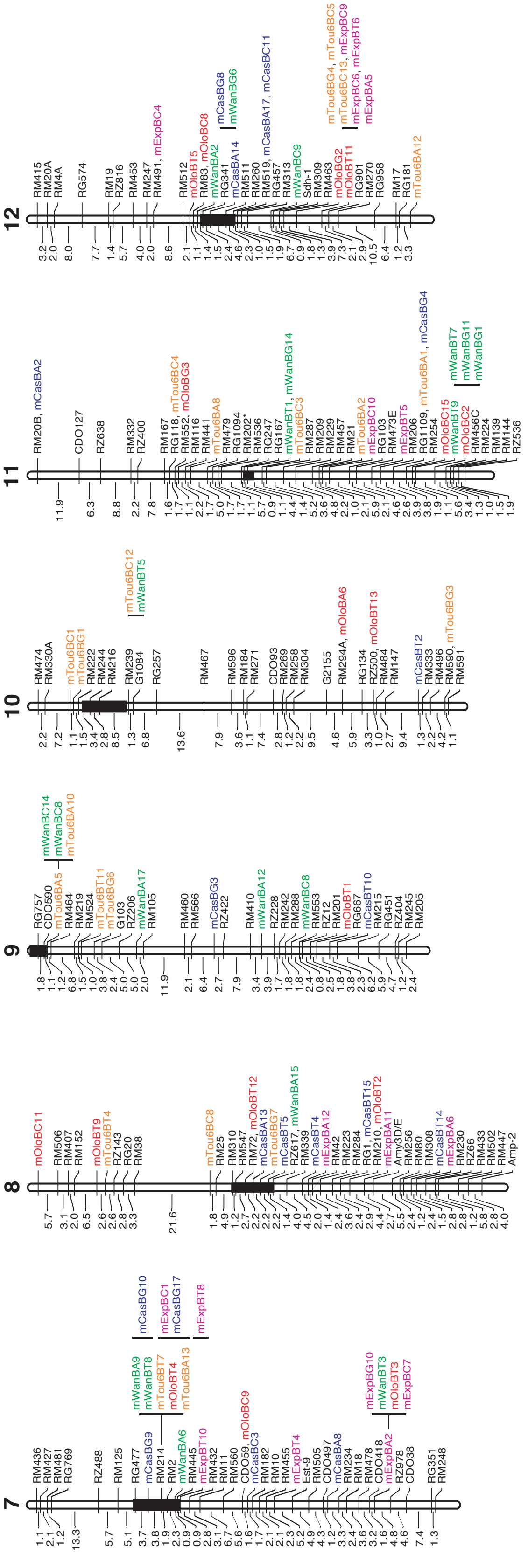
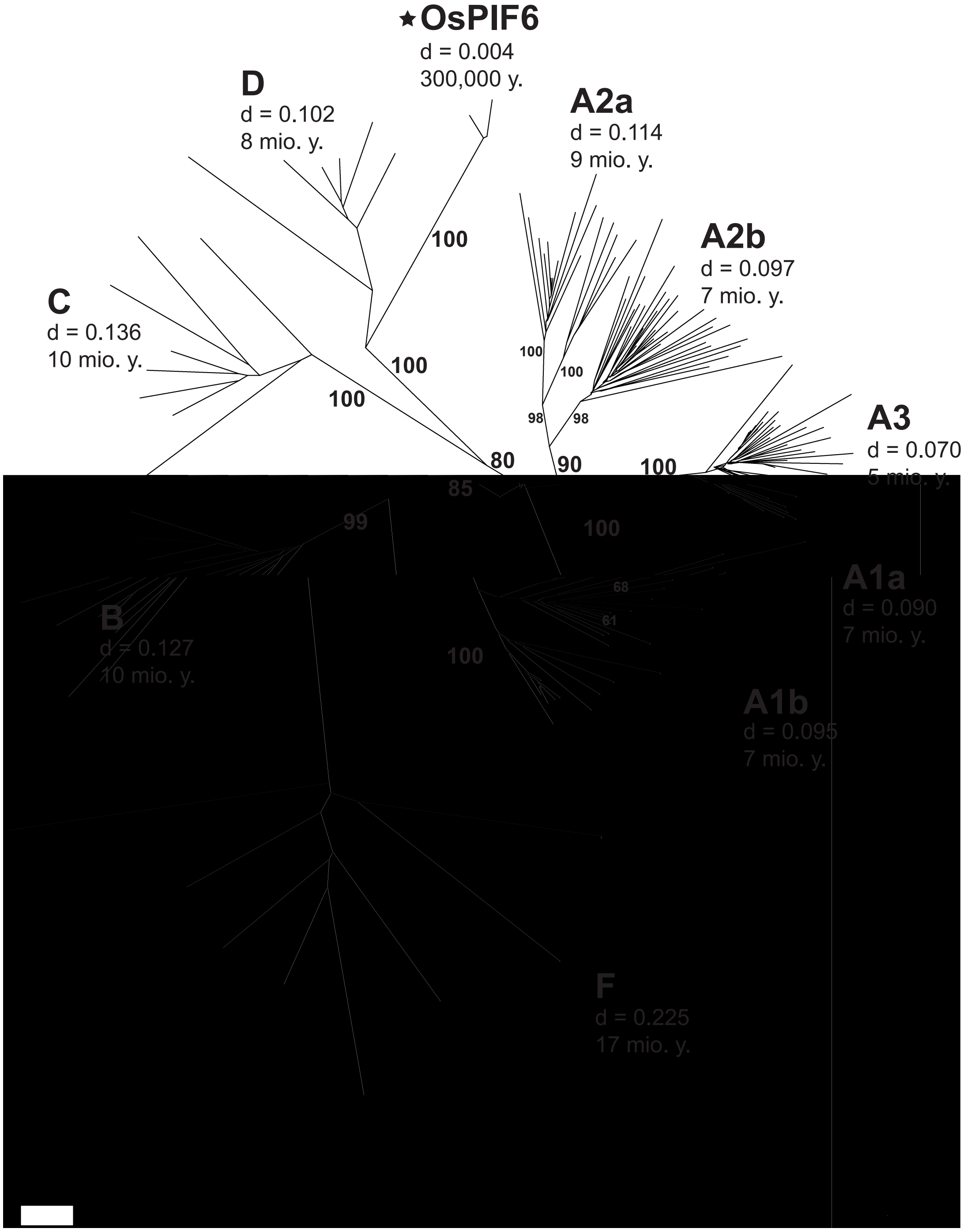
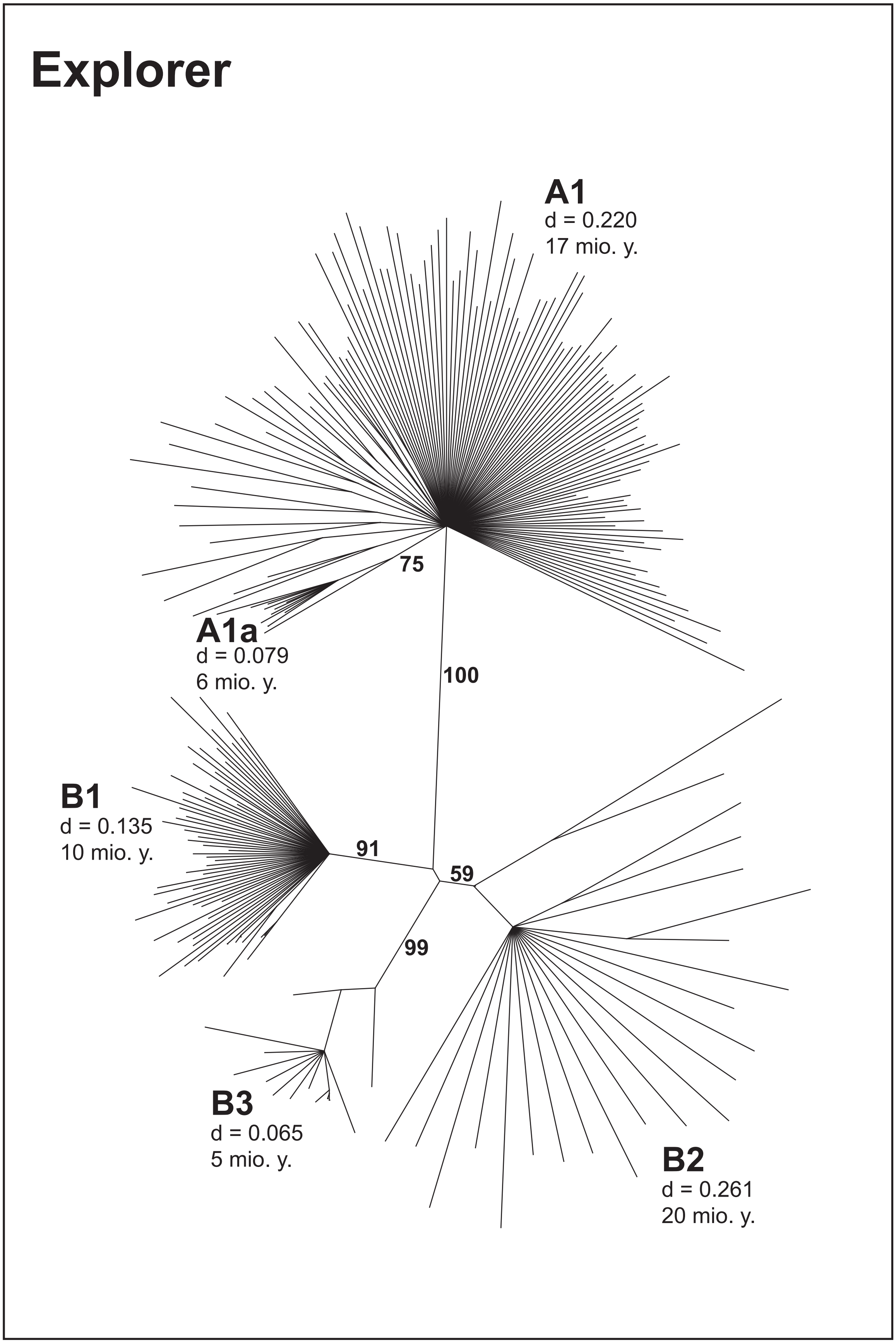

7
1.1
2.1
1.2
13.3
5.7
5.1
3.7
3.8
1.9
2.3
0.9
0.9
2.8
3.1
6.7
5.6
1.6
1.7
1.6
2.1
2.3
5.2
4.9
4.3
1.2
3.3
2.4
3.6
3.2
1.6
4.8
4.6
7.4
1.3
RM436
RM427
RM481
RG769
RZ488
RM125
RG477
mCasBG9
RM214
RM2
mWanBA6
RM445
mExpBT10
RM432
RM11
RM560
CDO59, mOloBC9
mCasBC3
RM182
RM10
RM455
mExpBT4
Est-9
RM505
CDO497
mCasBA8
RM234
RM18
RM478
CDO418
mExpBA2
RZ978
CDO38
RG351
RM248
mWanBA9
mWanBT8
mTou6BT7
mOloBT4
mTou6BA13
mCasBG10
mExpBC1
mCasBG17
mExpBT8
mExpBG10
mWanBT3
mOloBT3
mExpBC7
8
5.7
3.1
2.0
6.5
2.6
2.6
2.8
3.3
21.6
1.8
4.9
1.2
2.7
2.2
2.2
2.2
1.4
4.0
4.5
2.0
1.4
2.4
3.6
2.4
2.4
4.4
2.7
5.5
2.4
1.2
2.4
1.5
2.8
2.8
1.2
5.8
2.8
4.0
mOloBC11
RM506
RM407
RM152
mOloBT9
mTou6BT4
RZ143
RG20
RM38
mTou6BC8
RM25
RM310
RM547
RM72, mOloBT12
mCasBA13
mTou6BG7
mCasBT5
RZ617, mWanBA15
RM339
mCasBT4
mExpBA12
RM42
RM223
RM284
RG1, mCasBT15
RM210, mOloBT2
mExpBA11
Amy3D/E
RM256
RM80
RM308
mCasBT14
mExpBA6
RM230
RZ66
RM433
RM502
RM447
Amp-2
9
1.8
1.1
1.2
6.8
1.5
1.0
3.8
2.4
5.0
5.0
2.0
11.9
2.1
6.4
2.7
7.9
3.4
3.9
1.7
1.8
1.8
2.4
0.8
2.5
1.8
3.8
2.3
6.2
5.9
4.7
1.2
2.4
RG757
CDO590
mTou6BA5
RM464
RM219
RM524
mTou6BT11
mTou6BG6
G103
RZ206
mWanBA17
RM105
RM460
RM566
mCasBG3
RZ422
RM410
mWanBA12
RZ228
RM242
RM288
mWanBC8
RM553
RZ12
RM201
mOloBT1
RG667
mCasBT10
RM215
RG451
RZ404
RM245
RM205
mWanBC14
mWanBC8
mTou6BA10
10
2.2
7.2
1.1
1.5
3.4
2.8
8.5
1.3
6.8
13.6
7.9
3.6
1.1
7.4
2.8
1.2
2.2
9.5
4.6
5.9
3.3
1.0
2.7
9.4
1.3
2.2
4.2
1.1
RM474
RM330A
mTou6BC1
mTou6BG1
RM222
RM244
RM216
RM239
G1084
RG257
RM467
RM596
RM184
RM271
CDO93
RM269
RM228
RM304
G2155
RM294A, mOloBA6
RG134
RZ500, mOloBT13
RM484
RM147
mCasBT2
RM333
RM496
RM590, mTou6BG3
RM591
mTou6BC12
mWanBT5
11
11.9
6.3
8.8
2.2
7.8
1.6
1.7
1.1
2.2
1.7
5.0
1.7
1.1
5.7
0.9
1.1
4.1
1.4
5.2
3.6
4.8
2.2
1.0
2.1
5.9
2.1
4.6
2.6
3.9
3.8
1.9
1.1
5.6
3.4
1.3
1.0
1.5
1.9
RM20B, mCasBA2
CDO127
RZ638
RM332
RZ400
RM167
RG118, mTou6BC4
RM552, mOloBG3
RM116
RM441
mTou6BA8
RM479
RG1094
RM202*
RM536
RG247
RG167
mWanBT1, mWanBG14
mTou6BC3
RM287
RM209
RM229
RM457
RM21
mTou6BA2
mExpBC10
RG103
RM473E
mExpBT5
RM206
RG1109, mTou6BA1, mCasBG4
RM254
mOloBC15
mWanBT9
mOloBC2
RM456C
RM224
RM139
RM144
RZ536
mWanBT7
mWanBG11
mWanBG1
12
3.2
2.0
8.0
7.7
1.4
5.7
4.0
2.0
8.6
2.1
1.1
1.4
1.5
2.4
4.6
2.3
1.0
1.5
1.9
6.7
0.9
1.8
1.3
3.9
7.3
2.1
2.9
10.5
6.4
1.2
3.3
RM415
RM20A
RM4A
RG574
RM19
RZ816
RM453
RM247
RM491, mExpBC4
RM512
mOloBT5
RM83, mOloBC8
mWanBA2
RG341
mCasBA14
RM511
RM260
RM519, mCasBA17, mCasBC11
RG457
RM313
mWanBC9
Sdh-1
RM309
RM463
mOloBG2
mOloBT11
RG901
RM270
RG958
RM17
RG181
mTou6BA12
mCasBG8
mWanBG6
mTou6BG4, mTou6BC5
mTou6BC13, mExpBC9
mExpBC6, mExpBT6
mExpBA5

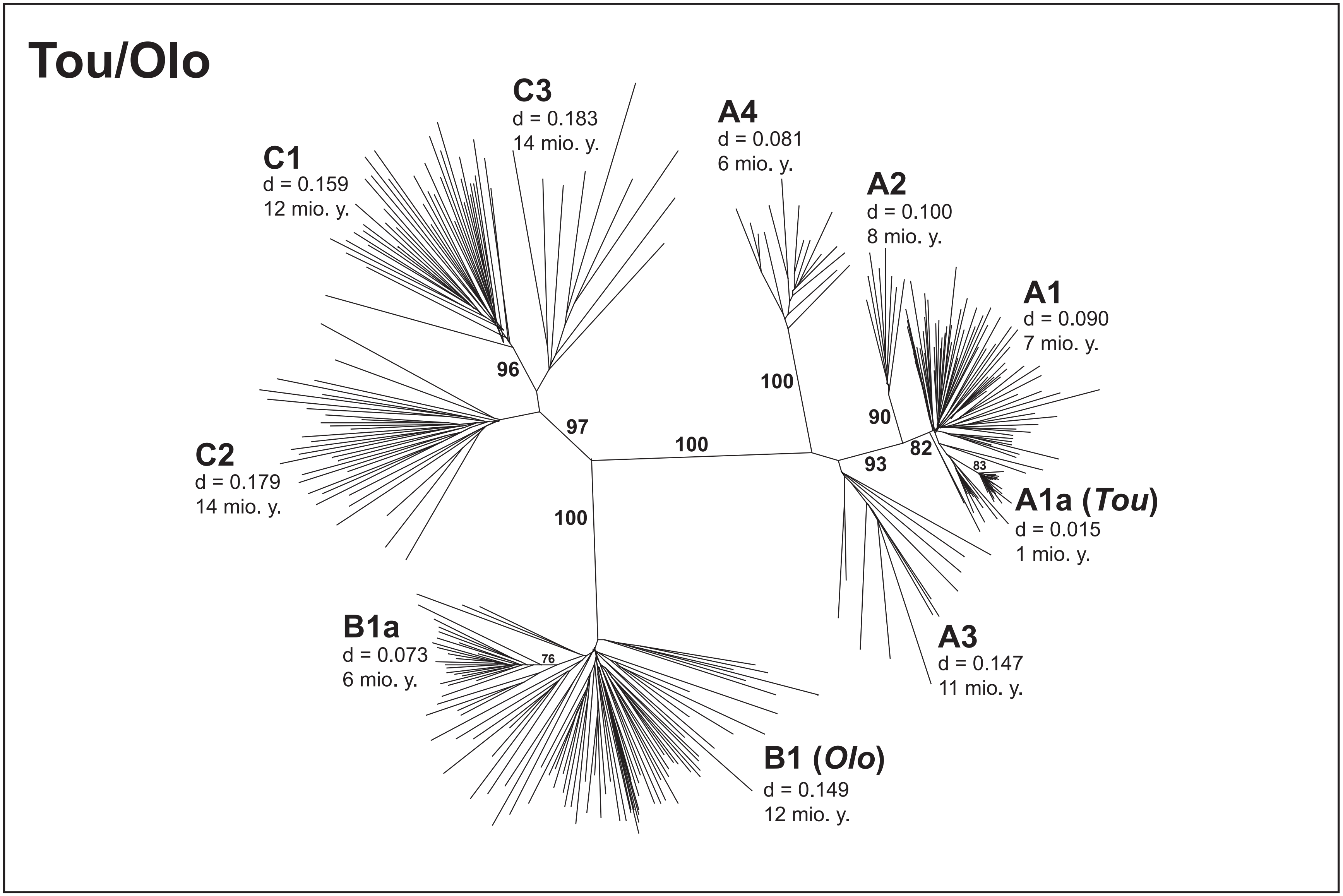
Figure 2.6: Phylogenetic relationships of *Castaway*, *Explorer*, *Tou*/*Olo* and *Wanderer* sequences. The MITEs *Olo* and *Tou*, for which TD primers have been developed, are represended by Tou-B and Tou-A, repectively. Bootstrap values were calculated from 1,000 replicates.

Castaway



[redacted]


Explorer
A1
d = 0.220
17 mio. y.
75
A1a
d = 0.079
6 mio. y.
100
B1
d = 0.135
10 mio. y.
91
59
99
B3
d = 0.065
5 mio. y.
B2
d = 0.261
20 mio. y.


Tou/Olo
C3
d = 0.183
14 mio. y.
C1
d = 0.159
12 mio. y.
A4
d = 0.081
6 mio. y.
A2
d = 0.100
8 mio. y.
A1
d = 0.090
7 mio. y.
96
97
100
100
90
82
93
83
C2
d = 0.179
14 mio. y.
100
A1a (Tou)
d = 0.015
1 mio. y.
B1a
d = 0.073
6 mio. y.
76
A3
d = 0.147
11 mio. y.
B1 (Olo)
d = 0.149
12 mio. y.

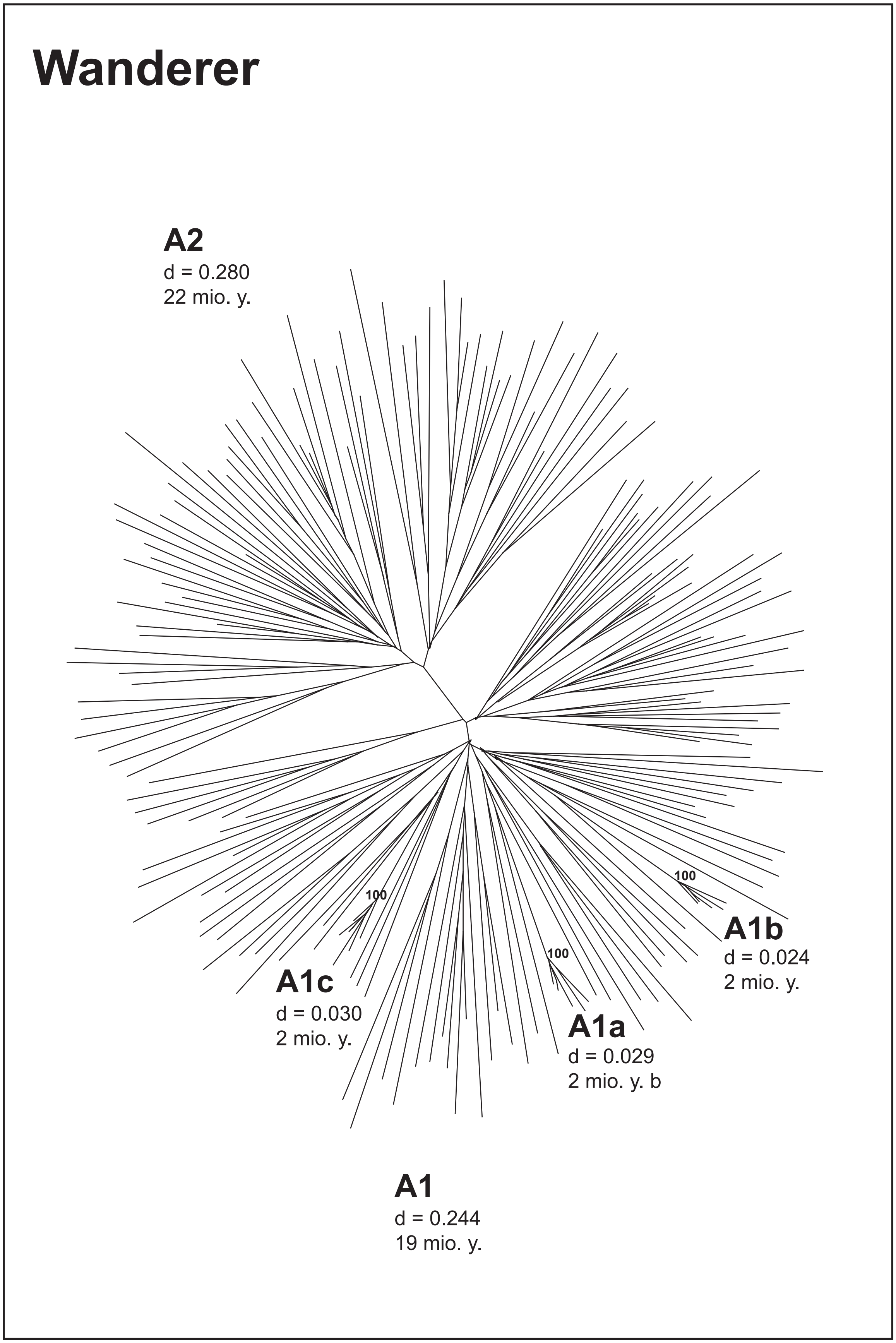
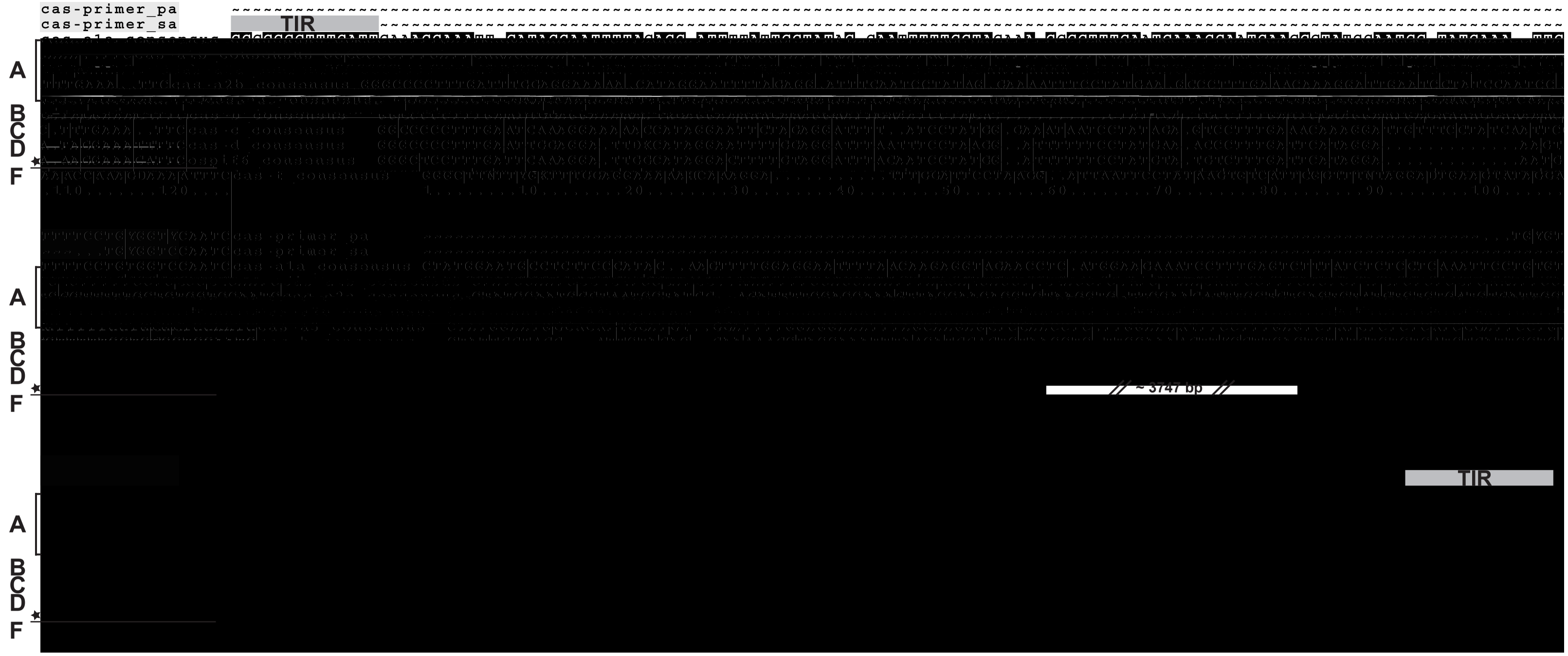
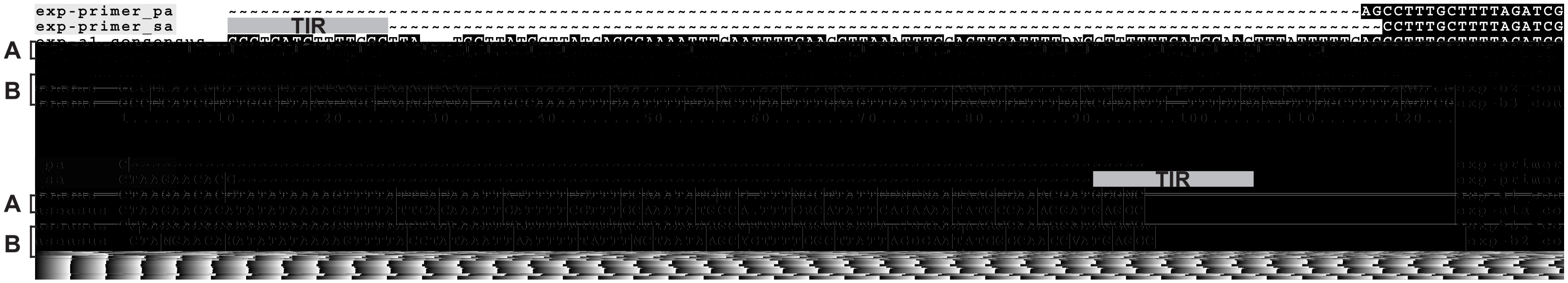

Wanderer
A2
d = 0.280
22 mio. y.
100
100
100
A1b
d = 0.024
2 mio. y.
A1c
d = 0.030
2 mio. y.
A1a
d = 0.029
2 mio. y. b
A1
d = 0.244
19 mio. y.

Figure 2.7: Alignments of subfamily consensus sequences for *Castaway*, *Explorer*, *Tou*/*Olo* and *Wanderer*. The first two sequences in each alignment are the primer sequences used in TD. Families are labeled with upper case letters and grouped by vertical brackets. For *Castaway* the *OsPIF6* consensus is underlined and highlighted by a star.

Castaway



Explorer



exp-primer_pa
exp-primer_sa
exp-a1 consensus
TIR
AGCCTTTGCTTTTAGATCG
CCTTTGCTTTTAGATCG
A
B
1
10
20
30
40
50
60
70
80
90
100
110
120
exp-b2_con
exp-b3_con
pa
sa
CTAAGAACACC
TIR
A
B

Tou/Olo


tou-a-primer_pa
tou-a-primer_sa
TIR
tou-a1_consensus
GGTGGTGTTTGGATCCAGGGACTTAA...CTTTAGT..CTCTG.........TATTTAGACACTAATTTAGAGTATTAAATATAGACTACTTACAAAACTAATTACATAAATGAAAGCTAATTT
tou-a1a_consensus
A
C
B
A
C
B
A
C
B
TIR

Wanderer

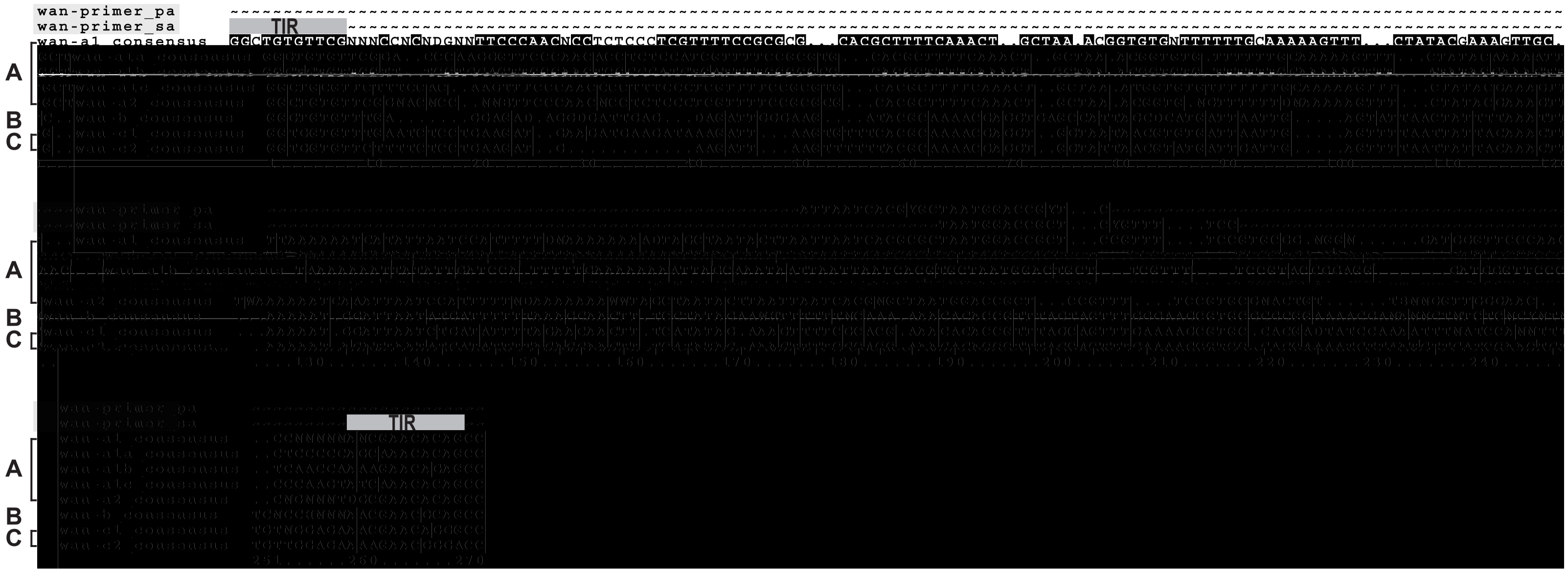

wan-primer_pa
wan-primer_sa
wan-a1 consensus
TIR
GGCTGTGTTCGNNNCCNCNDGNNTTCCCAACNCCTCTCCCTCGTTTTCCGCGCG...CACGCTTTTCAAACT..GCTAA.ACGGTGTGNTTTTTTGCAAAAAGTTT...CTATACGAAAGTTGC
A
B
C
A
B
C
A
B
C
TIR

Table 2.1: Level of band polymorphism observed on TD gels between the parents of the mapping population IR64 and Azucena.

	Number of bands				
	IR64	**Azucena**	**Total**	**Polymorphic**	**Polymorphism**
Castaway	208	235	443	129	29.1%
Explorer	197	177	374	132	35.3%
Olo	128	155	283	112	39.6%
Tou	111	112	223	123	55.2%
Wanderer	168	173	341	127	37.2%

Table 2.2: Number of MITE markers mapped in the doubled-haploid mapping population IR64 x Azucena.

Chromosome	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Sum**
Castaway	11	4	7	6	9	1	5	5	2	1	2	4	57
Explorer	4	4	3	3	3	0	7	3	0	0	2	5	34
Tou	3	2	2	7	4	4	2	3	4	4	5	4	44
Olo	5	1	4	5	2	2	3	4	1	2	3	4	36
Wanderer	8	6	7	9	2	2	4	1	5	1	6	3	54
Sum	31	17	23	30	20	9	21	16	12	8	18	20	225

CHAPTER 3

CONCLUSIONS

MITEs are the most abundant type of transposable element in rice. Their origin and mechanism of amplification are just beginning to be understood. It remains an open question why this kind of transposon is able to attain copy numbers up to several thousand for a single MITE family in the genomes of plants and animals. In order to address these questions, some basic information about the biology of MITEs was contributed in this thesis. Five *Tourist*-like MITE families were characterized in terms of their distribution, phylogeny, subfamily structure, time of divergence and sequence evolution. Transposon display (TD) was readily applied to different MITE families and used to genetically map 225 new MITE markers in rice, which were distributed among all 12 chromosomes in euchromatic as well as heterchromatic regions. It was shown that all MITE families studied here contain distinct subfamilies. The times of divergence of those subfamilies indicate that throughout a time period of around 22 million years ago to around 1 million years ago, MITEs spread through the genome in temporal waves of amplification. Successive rounds of amplification could also be concluded from the different levels of polymorphism observed on TD gels. The results indicate that features at the sequence level play a crucial role in whether or not MITE sequences are amplified.

REFERENCES

Bennetzen, J.L. 2000. The many hues of plant heterochromatin. *Genome Biology* 1: 107.1-107.4.

Bowen, N.J., and McDonald, J.F. 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* 11: 1527-40.

Brondani, C., Brondani, R.P.V., Rangel, P.H.N., and Ferreira, M.E. 2001. Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. *Hereditas* 134: 59-71.

Bureau, T.E., Ronald, P.C., and Wessler, S.R. 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc Natl Acad Sci U S A* 93: 8524-9.

Bureau, T.E., and Wessler, S.R. 1992. *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4: 1283-94.

Bureau, T.E., and Wessler, S.R. 1994a. Mobile inverted-repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. *Proc Natl Acad Sci U S A* 91: 1411-5.

Bureau, T.E., and Wessler, S.R. 1994b. *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* 6: 907-16.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S., and Wessler, S.R. 2000. Inaugural article: the MITE family *Heartbreaker* (*Hbr*): molecular markers in maize. *Proc Natl Acad Sci U S A* 97: 10083-9.

Casa, A.M., Mitchell, S.E., Smithe, O.S., Register III, J.C., Wessler, S.R., and 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.

Casacuberta, E., Casacuberta, J.M., Puigdomenech, P., and Monfort, A. 1998. Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: characterisation of the *Emigrant* family of elements. *Plant J* 16: 79-85.

Causse, M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P.C., Harrington, S.E., Second, G., McCouch, S.R., and Tanksley, S.D. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138: 1251-1274.

Deininger, P.L., Batzer, M.A., Hutchison, C.A., 3rd, and Edgell, M.H. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet* 8: 307-11.

Feschotte, C., Jiang, N., and Wessler, S.R. 2002. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3: 329-41.

Feschotte, C., and Mouches, C. 2000a. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol Biol Evol* 17: 730-7.

Feschotte, C., and Mouches, C. 2000b. Recent amplification of miniature inverted-repeat transposable elements in the vector mosquito *Culex pipiens*: characterization of the *Mimo* family. *Gene* 250: 109-16.

Feschotte, C., Swamy, L., and Wessler, S.R. 2003. Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* Miniature Inverted Repeat Transposable Elements (MITEs). *Genetics* 163: 747-758.

Feschotte, C., and Wessler, S.R. 2002. *Mariner*-like transposases are widespread and diverse in flowering plants. *Proc Natl Acad Sci U S A* 99: 280-5.

Frey, M., Stettner, C., and Gierl, A. 1998. A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS). *Plant Journal* 13: 717-721.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene rbcL. *Proc Natl Acad Sci U S A* 93: 10274-10279.

Ge, S., Sang, T., Lu, B.R., and Hong, D.Y. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci U S A* 96: 14400-5.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296: 92-100.

Izsvak, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., and Hackett, P.B. 1999. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *J Mol Evol* 48: 13-21.

Jiang, N., Bao, Z., Temnykh, S., Cheng, Z., Jiang, J., Wing, R.A., McCouch, S.R., and Wessler, S.R. 2002. *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* 161: 1293-305.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., and Wessler, S.R. 2003. An active DNA transposon family in rice. *Nature* 421: 163-167.

Jiang, N., and Wessler, S.R. 2001. Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* 13: 2553-2564.

Jordan, I.K., and McDonald, J.F. 1999. Comparative genomics and evolutionary dynamics of *Saccaromyces cerevisiae* Ty elements. *Genetica* 107: 3-13.

Kumar, A., and Bennetzen, J.L. 1999. Plant retrotransposons. *Annu Rev Genet* 33: 479-532.

Kunze, R., Saedler, H., and Loennig, W.-E. 1997. Plant transposable elements. *Adv Bot Res* 27: 331-470.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-81.

Li, W.-H. (1997). "Molcular Evolution," Sinauer Associates, Sunderland, MA.

Lorieux, M., Ndjiondjop, M.N., and Ghesquiere, A. 2000. A first interspecific *Oryza sativa* x *Oryza glaberrima* microsatellite-based genetic linkage map. *Theor Appl Genet* 100: 593-601.

Maheswaran, M., Subudhi, P.K., Nandi, S., Xu, J.C., Parco, A., Yang, D.C., and Huang, N. 1997. Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. *Theor Appl Genet* 94: 39-45.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A., and Wing, R.A. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged- connectors. *Genome Res* 10: 982-90.

McCouch, S.R., Kochert, G., Yu, Z.H., Khush, G.S., and Coffman, W.R. 1988. Molecular mapping of rice chormosomes. *Theor Appl Genet* 75: 815-829.

Meyers, B.C., Tingey, S.V., and Morgante, M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11: 1660-1676.

Morgan, G.T. 1995. Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J Mol Biol* 254: 1-5.

Oosumi, T., Garlick, B., and Belknap, W.R. 1996. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J Mol Evol* 43: 11-8.

Saedler, H., and Nevers, P. 1985. Transposition in plants: a molecular model. *EMBO Journal* 4: 585-590.

SanMiguel, P., and Bennetzen, J.L. 1998. Evidence that a

Stumm, G.B., Vedder, H., and Schlegel, J. 1997. A simple method for isolation of PCR fragments from silver-stained polyacrylamide gels by scratching with a fine needle. *Technical Tips Online* 1: 17.

Tarchini, R., Biddle, P., Wineland, R., Tingey, S., and Rafalski, A. 2000. The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* 12: 381-392.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11: 1441-52.

Tu, Z. 1997. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A* 94: 7475-80.

Turcotte, K., Srinivasan, S., and Bureau, T. 2001. Survey of transposable elements from rice genomic sequences. *Plant J* 25: 169-79.

Unsal, K., and Morgan, G.T. 1995. A novel group of families of short interspersed repetitive elements (SINEs)

Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E., and Schulman, A.H. 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769-1784.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23: 4407-14.

Yang, G., Dong, J., Chandrasekharan, M.B., and Hall, T.C. 2001. *Kiddo*, a new transposable element family closely associated with rice genes. *Mol Genet Genomics* 266: 417-24.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., and Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296: 79-92.

Zhang, Q., Arbuckle, J., and Wessler, S.R. 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. *Proc Natl Acad Sci U S A* 97: 1160-5.

Zhang, Q., and Kochert, G. 1997. Independent amplification of two classes of *Tourists* in some *Oryza* species. *Genetica* 101: 145-52.

Zhang, X., Feschotte, C., Zhang, Q., Jiang, N., Eggleston, W.B., and Wessler, S.R. 2001. P instability factor: an active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. *Proc Natl Acad Sci U S A* 98: 12572-7.

Zhang, X., Jiang, N., Feschotte, C., and Wessler, S.R. 2003. *PIF*- and *Pong*-like transposable elements: Distribution, evolution and relationship with *Tourist*-like Miniature Inverted-repeat Transposable Elements. *Genetics*: in press.