DNA-binding specificity of rice *mariner*-like transposases and interactions with *Stowaway* MITEs

Cédric Feschotte[1,2], Mark T. Osterlund[1], Ryan Peeler[1] and Susan R. Wessler[1,*]

[1]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA and [2]Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA

ABSTRACT

***Mariner*-like elements (MLEs) are DNA transposons found throughout the plant and animal kingdoms. A previous computational survey of the rice (*Oryza sativa*) genome sequence revealed 34 full length MLEs (*Osmars*) belonging to 25 distinct families. This survey, which also identified sequence similarities between the *Osmar* elements and the *Stowaway* superfamily of MITEs, led to the formulation of a hypothesis whereby *Stowaways* are mobilized by OSMAR transposases. Here we investigate the DNA-binding activities and specificities of two OSMAR transposases, OSMAR5 and OSMAR10. Like other *mariner*-like transposases, the OSMARs bind specifically to the terminal inverted repeat (TIR) sequences of their encoding transposons. OSMAR5 binds DNA through a bipartite N-terminal domain containing two functionally separable helix-turn-helix motifs, resembling the paired domain of Tc1-like transposases and PAX transcription factors in metazoans. Furthermore, binding of the OSMARs is not limited to their own TIRs; OSMAR5 transposase can also interact *in vitro* with TIRs from closely related *Osmar* elements and with consensus TIRs of several *Stowaway* families mined from the rice genome sequence. These results provide the first biochemical evidence for a functional relationship between *Osmar* elements and *Stowaway* MITEs and lead us to suggest that there is extensive cross-talk among related but distinct transposon families co-existing in a single eukaryote genome.**

INTRODUCTION

Transposable elements make up the largest fraction of many eukaryotic genomes. They are divided into two classes based on their mechanism of transposition (1–3). Class 1 elements (retrotransposons) transpose by means of an RNA intermediate in a reaction involving several enzymes, including reverse transcriptase and integrase. In contrast, class 2 elements (DNA transposons) move directly via DNA and the transposition reaction is catalyzed by an element-encoded enzyme called transposase. During transposition of class 2 elements, transposase molecules bind to the transposon in a sequence-specific manner and catalyze both the DNA cleavage and strand transfer steps of the so-called 'cut-and-paste' reaction (3).

Nonautonomous DNA transposons do not encode transposase, but can still undergo transposition using transposase encoded by an autonomous element [for a review, see (2)]. Because nonautonomous elements are often internal deletion derivatives of autonomous elements, they retain the terminal binding sites necessary to interact with transposase. A seemingly distinct class of nonautonomous elements, called miniature-inverted repeat elements (MITEs), was first identified in plant genomes (4–7) and was subsequently identified in a wide range of animal genomes, including those of nematodes, mosquitoes, sea squirt, zebrafish, frogs and humans [for a review, see (6)]. Elements structurally reminiscent of MITEs were also described in eubacteria and archaea (8–10). Like most other nonautonomous transposons, MITEs have no coding capacity and thus must rely on transposase encoded *in trans* by autonomous elements. MITE families are distinguished from other nonautonomous DNA transposons by their high copy numbers and structural homogeneity. These features are consistent with the initial amplification of one or a few MITEs.

Because MITEs do not encode transposase, their initial classification was based on similarities in noncoding sequences

---

including the terminal inverted repeat (TIR) and target site duplication (TSD). Using these criteria, most plant MITEs were assigned to one of two groups, *Tourist* or *Stowaway* [for a review, see (6,7)]. Subsequently, sequence similarity was detected between the TIRs and TSDs of MITEs and of transposase-encoding elements in the same genome. For example, *Tourist* MITEs share sequence similarity with the TIRs of the transpoase-encoding PIF/*Harbinger* superfamily (11–13), while the TIR and TSD of *Stowaway* MITEs are similar to the TIR and TSD of Tc1/*mariner* elements (5,14,15).

In a recent genome-wide analysis of rice, the sequences of virtually all *Stowaway* MITEs and *mariner*-like elements (MLEs) (called *Osmars*) were identified and compared (16). More than 22 000 *Stowaway* MITEs were classified into 36 families, while 34 different *Osmars* were found to group into three major clades based on their transposase and TIR sequences (Figure 1; clades A, B and C). Although sequence similarities between *Osmar* and *Stowaway* families were mainly restricted to their TIRs, these similarities were sufficient to associate most *Stowaway* families with one of the *Osmar* clades. This comparative analysis led to the formulation of a model for *Stowaway* amplification whereby *Stowaway* MITEs were mobilized by MLE transposases encoded by seemingly, distantly related elements.

Transposases of animal MLEs are known to specifically interact with the TIRs of the encoding transposon (17–19). Comparative genetic and biochemical studies of distinct MLE transposons revealed that minor sequence variations within the TIRs is sufficient to dramatically decrease binding (20–22). Because the binding of transposase to sequences at or near the TIR is usually the first step in transposition, failure to bind to subtle variant TIRs has been interpreted as a way to prevent the cross-mobilization of transposons from different families. However, it should be noted that the MLEs tested for cross-interactions in these studies were from different species and that their sequences had significantly diverged beyond their TIRs (20,22). Analyses of interactions between different MLEs within the same genome have not as yet been reported. With more than 25 different *Osmar* families (16), rice offers an ideal system to investigate the DNA-binding specificities of MLE transposases. Furthermore, with over 22 000 *Stowaway* elements, rice is also an ideal system to study the functional interactions between MITEs and transposase-encoding elements.

In this study we have reconstructed transposases from different *Osmar* families and have undertaken their functional characterization with emphasis on their DNA-binding specificities. Results gathered from *in vitro* assays and *in vivo* experiments in yeast indicate that a single *Osmar* transposase is able to recognize and bind its own TIRs as well as the TIRs of other *Osmar* families and those of several *Stowaway* MITEs co-existing in the rice genome.

Figure 1. Phylogenetic relationships of rice MLEs (*Osmars*) based on their transposase sequences. The neighbor-joining tree was generated from a multiple alignment of the entire transposase amino acid sequences of 34 *Osmars* and *Soymar1* from soybean, which served as outgroups (16). All *Osmars* are from the subspecies *japonica*, except those followed by 'i', which are from the subspecies *indica*. The elements can be divided into three clades (A, B and C), of which A and B are further divided into subclades (A1, A2, B1, B2 and B3). Every clade and subclade is supported by at least 75% bootstrap values (1000 replicates). Dots to the right of the element name represents transposon TIRs tested in this study (Figure 8B).

MATERIALS AND METHODS

Sequence and phylogenetic analyses

All *Osmar* sequences were previously described and their accession numbers are available in Feschotte *et al.* (16). *Osmars* were mined from two databases, one containing ∼360 Mb of BAC/PAC sequences from *O.sativa* ssp. *japonica* cv. Nipponbare (downloadable at http://rgp.dna.affrc.go.jp/cgi-bin/statusdb/seqcollab-assign.pl) and the second containing ∼430 Mb of contigs generated by whole-genome sequencing of *O. sativa* ssp. *indica* cv. 9311 (downloadable at http://btn.genomics.org.cn/rice/). *Stowaway* families were previously identified in the genomic sequence of *O. sativa* ssp. *japonica* cv. Nipponbare and their assignment to corresponding *Osmar* clades has been described (16). Consensus sequences for each *Stowaway* family are available through Repbase (http://www.girinst.org) or upon request. Predictions for helix-turn-helix motifs were carried out with the method of Dodd and Egan (23) through the Network Protein Sequence Analysis website (http://www.npsa-pbil.ibcp.fr). To generate the tree shown in Figure 8B, full-length *Osmar* transposase sequences were aligned with ClustalW using default parameters in MacVector 7.0 and the alignment was trimmed to preserve only the regions that align with the first 206 amino acids of *Osmar5* (see Figure 4). The resulting phylogenetic tree was generated with PAUP* Version 4.0b8 (http://paup.csit.fsu.edu/) using the neighbor-joining method with default parameters and rooted at the midpoint. All other sequence manipulations were performed with MacVector 7.0 or as described in Feschotte *et al.* (16).

Cloning of *OSMAR5* and *OSMAR10* ORFs

Both *OSMAR5* and *OSMAR10* transposases were cloned from genomic DNA using PCR. 5′ primers beginning at the start codon and 3′ primers ending with the stop codon of each predicted gene were used to amplify full-length transposases. Primer sequences are: OSMAR5-5′, GATGCACCATGTG-GCGCACA; OSMAR5-3′, GCTAGTTAGCCAAAGACTC; OSMAR10-5′, GATGCATGCCAACCATAGTATAGCCT-GCAT; OSMAR10-3′, ACTATGTATTGATGAAATCTAC-CGCATCTA. PCR products were cloned into the pCR2.1-TOPO cloning vector (Invitrogen). Introns were removed with PCR using complementary primers that spanned the introns. The sequences of those primers were: OSMAR5_intron1, AACAAGAAGAAACTGAACTAGATGACCATG-AAGCATTGCA; OSMAR5_intron2, TTTGGCCTTTTACT-AGGAAGGAACCAGCCCGAAGGAGAAG; OSMAR10_intron1, TAGTCGTAGTAGGGAGAGGGGAACCATGGA-CCTCAAGGCC. Individual exons were initially amplified from full-length transposase clones. The exons were then added together and subjected to 30 PCR cycles. The resulting PCR product, containing a full-length transposase ORF lacking introns, was cloned into the pCR2.1-TOPO cloning vector (Invitrogen) and the identity and integrity of the sequence was verified by sequencing. A chimeric *OSMAR10BA* clone was generated by combining exon 1 from *OSMAR10B* with exon 2 from *OSMAR10A*. The resulting plasmids were named TOPO(OSMAR5) and TOPO(OSMAR10BA).

Preparation of vectors and constructs for yeast one-hybrid assays

The yeast one-hybrid system was a modification of two previously reported systems. Plasmid pJG4-5 was used to express transposase as previously described (24). Reporter plasmid pLacZi-2μ was modified from pLacZi (Clontech) by introducing a 2μ origin of replication that was isolated by PCR amplification from pRS416 (New England Biolabs) with the primers 2μ-F (TATAATTAAATTGAAGCTCTAATTTGT-GAGTTTAGTATACACGAAGCATCTGTGCTTCAT) and 2μ-R (AGATGCGGCCAGCAAAACTAAAAAACTGTAT-TATAAGTAATATGATCCAATATCAAAGGA). The resulting fragment was gel-purified and introduced into the pLacZi vector through homologous recombination following the linearization of pLacZi at the Nsi1 site and co-transformation into yeast. Plasmid, pLacZi-2μ is thus a high copy autonomous vector.

The *OSMAR5* transposase coding sequence was amplified by PCR from TOPO(OSMAR5) and cloned into the pJG4-5 vector using EcoRI restriction site adapters, resulting in an inframe fusion with the B42 transcriptional activation domain. The resulting plasmid was named pJG4-5(OSMAR5). To create an inframe fusion of *OSMAR10* with the activation domain, we used PCR to amplify the OSMAR10BA ORF from TOPO(OSMAR10BA) with primers containing a 5′ EcoRI site and a 3′ XhoI site. Primers were: OSMAR10_5′ (ATGAATTCATGCATGCCAACCATAGTATAGCC) and OSMAR10_3′ Xho1 (ATCTCGAGACTATGTATTGATG-AAATCTACCG). The PCR product was digested after gel-purification and ligated into pJG4-5, resulting in the plasmid pJG4-5(OSMAR10). *OSMAR5* deletion constructs were generated by PCR using pJG4-5(OSMAR5) as a template with one

of two primers specific to the pJG4-5 vector (Bco1 and Bco2). An additional primer within the *OSMAR5* gene (del primers) was also used which contained an appropriate restriction site necessary to clone the resulting PCR amplicon into the pJG4-5 backbone. Bco1 (CCAGCCTCTTGCTGAGTGGAGATG) is a forward primer 5′ of the *OSMAR5* gene. Bco2 (GACAAGC-CGACAACCTTGATTGGAG) is a reverse primer 3′ of the *OSMAR5* gene. The del-1 clone (amino acids 207–502) was generated using primers DelF1 (ACGAATTCTCGCTAAA-GCCTTATTTGAAGGA) and Bco2. The del-2 clone (amino acids 115–502) was generated with DelF2 (ACGAATTCGC-TATTCCATTGCATCAAAGGA) and Bco2. The del-3 clone (amino acids 1–206) was generated with Bco1 and DelR1 (ACCTCGAGTCACCTTCAAATAAGGCTTTAGCGAG), which adds an inframe stop codon. The del-4 clone (amino acids 1–114) was generated with Bco1 primer and DelR2 (ACCTCGAGTCATCCTTTGATGCAATGGAATAGC), which adds an inframe stop codon. The PCR products for all deletion constructs were digested with EcoR1 and Xho1 and ligated to pJG4-5. The integrity of all pJG4-5 clones was verified by sequencing.

The pLacZi-2μ plasmids were constructed using synthetic complementary 30mer oligonucleotides (Integrated DNA Technologies) that were designed to generate 5′-AATT overhangs on each strand when annealed. Double-stranded DNAs were generated by mixing equal concentrations of oligos, heating to 95°C for 5 min and brought to room temperature. Double-stranded oligos were ligated to pLacZi-2μ previously linearized by digestion with EcoR1. Following transformation, clones were sequenced to verify inserts and orientation. The following oligonucleotides correspond with clones in Figures 2 and 3A: 'Osm10 TIR', CTCCCTCCGTTCCT-TAATATAGAGCGTGGC; 'Osm5 TIR' (also clone 'A'), CTCCCTCCGTCCCACAAAACATGACGTTTT; 'B', AAA-ACATGACGTTTTAAGGTTAGCAGCCAA; 'C', AAGGT-TAGCAGCCAAAATTAGCTGTTGTGC; 'D', CAGCCAA-AATTAGCTGTTGTGCAAAATGAC; 'E', ACATTTTTTC-TGGGACAATCCAGGGGCGGT; 'F', AAAACCTGCCGT-TTCACCGCCCCTGGATTG; 'G', CTCCCTCCGTCCCAC-AAAACCTGCCGTTTC.

Yeast one-hybrid assay

The yeast one-hybrid assay was performed as described by the manufacturer (Clontech, Matchmaker one-hybrid kit). Briefly, plasmids pJG4-5 and pLacZi-2 μ were co-transformed into yeast (strain EGY48) using standard transformation techniques. Transformants were selected on drop-out media lacking uracil and tryptophan. Colonies were grown overnight in liquid media containing glucose, and 5 μl of the culture was spotted onto galactose plates containing X-gal. Positive colonies appeared blue within 2–3 days. Liquid assays were performed according to the protocol provided by Clontech (Matchmaker 1-hybrid kit). ONPG was the substrate used for the liquid assays and eight replicates of each experiment were analyzed.
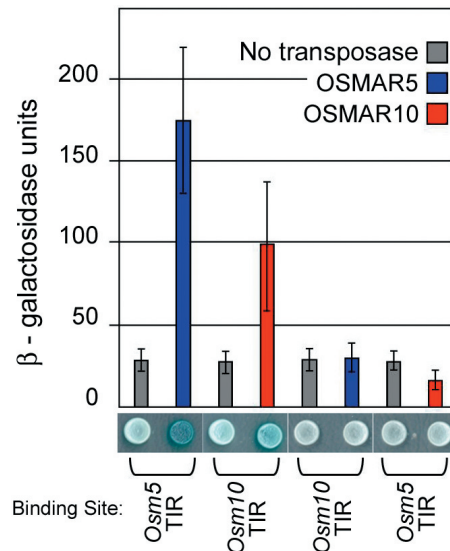
Protein expression and purification

Both full-length OSMAR5 and the truncated version OSMAR5N (amino acids 1–206) were cloned into the pMal vector from the pJG4-5 vectors [pJG4-5(OSMAR5) and del-3].
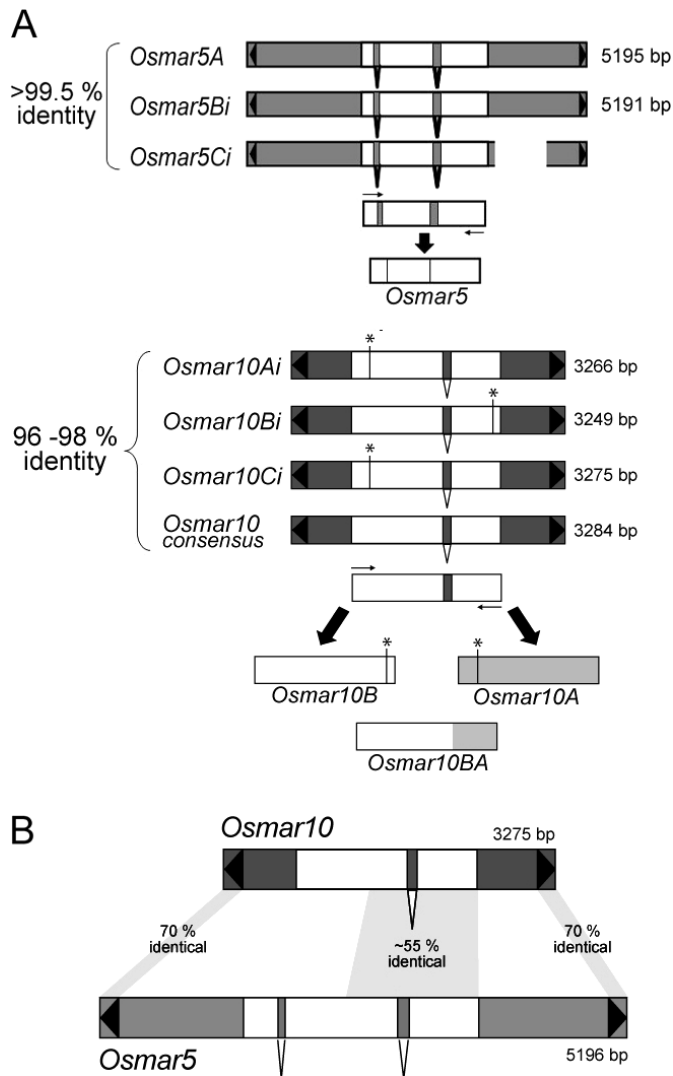
Figure 2. Analysis and construction of OSMAR5 and OSMAR10 transposases. (**A**) Cloning of OSMAR5 and OSMAR10 transposase constructs. Genomic copies of the *Osmar5* and *Osmar10* transposons and an *Osmar10* consensus sequence are represented. The extent of intrafamily pairwise sequence similarity (excluding gaps) is shown. *Osmar5Ci* lacks 763 bp of sequence downstream of the putative transposase gene due to a gap in the *indica* genome contig assembly. Asterisks above the *Osmar10* elements represent mutations that result in stop codons within the transposases. *Osmar5* and *Osmar10* transposase ORFs were generated by PCR amplification of exons, removing introns in the process. OSMAR5 was assembled from exons matching those of *Osmar5Bi*. OSMAR10 is chimeric in origin, generated by assembly of the first exon of *Osmar10B* with the second exon of *Osmar10A*, which results in an intact ORF OSMAR10BA. (**B**) Comparison between *Osmar5* and *Osmar10*. TIRs are represented by solid black triangles. Transposase coding sequences are depicted as white boxes, noncoding regions as solid gray and introns as open triangles below the element. Percentages of sequence similarity are indicated for the two transposons; all other regions could not be confidently aligned.

Figure 3. Yeast one-hybrid analysis of the DNA-binding of OSMAR5 and OSMAR10. Specific interaction between a fusion protein (OSMAR5 or OSMAR10BA transposase fused with the B42 transcriptional activation domain) and the DNA cloned upstream of a β-*galactosidase* reporter gene as quantified by β-GAL activity. DNA binding sites tested were the 30 bp *Osmar5* 5′ TIR (*Osm5* TIR) or the 30 bp *Osmar10A* 5′ TIR (*Osm10* TIR). Background controls of β-GAL activity were determined from experiments where an empty expression vector (without transposase sequences) was used (gray bars). Error bars indicate standard deviations from the average activity determined for eight replicates. Representative yeast colonies are shown below the graph.

transposase sequence and designated pMal(OSMAR5) and pMal(5N). Junctions between the two domains were verified by sequencing.

pMal(OSMAR5) and pMal(5N) were transformed into *Escherichia coli* strain BL21 for expression and purification of the corresponding recombinant proteins. One hundred ml log phase cultures containing each plasmid were induced with 0.3 mM IPTG for 3 h at room temperature. Cells were pelleted and resuspended in extraction buffer (20 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA, 10% glycerol and 1 mM PMSF) disrupted by sonication, centrifuged, and the supernatant mixed with amylose resin for 1 h at 4°C. After three 10 min washes with extraction buffer, purified protein was eluted with 500 ml of extraction buffer containing 2 mM DTT and 10 mM maltose. Purified proteins were analyzed by Coomassie staining and western blotting with an MBP monoclonal antibody (R29.6 from Santa Cruz Biotechnology) (data not shown).

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSAs) were performed using double-stranded oligonucleotides (Integrated DNA Technologies). The sequences of the oligonucleotides are shown in Figures 7, 8B and 9B. Complementary oligonulceotides were mixed at a concentration of 1.8 mM and end-labeled with γ-P[33] using T4 kinase (Invitrogen). The labeling reaction was stopped by heating to 95°C for 7 min, followed by slow cooling to allow the annealing of the oligos. EMSA reactions were carried out in 15 μl containing 1 μl of labeled DNA and 0.25 μl of purified protein. EMSA buffer was 15 mM Tris (pH 7.5), 0.1 mM EDTA, 1 mM DTT, 0.3 mg/ml BSA,

Full-length OSMAR5 and OSMAR5N were excised as EcoR1-Xho1 fragments from pJG4-5(OSMAR5) and del-3, respectively. After gel purification, the fragments were cloned into a pMal vector (pMal-c2X from New England Biolabs) previously digested with EcoR1 and Sal1. The resulting constructs contained inframe fusions between the sequence-encoding maltose-binding protein (MBP) and the

0.1% NP-40, 10% glycerol and 33 mg/ml single-stranded DNA. Reactions were incubated for 1 h at room temperature and samples were separated by electrophoresis on 6% native polyacrylamide gels. Gels were visualized using X-ray film or with a phosphorimager (Molecular Dynamics Inc.) and shifted DNA was quantified using Scion Image (Scion Corp., Frederick, MD) or with Image-quant software (Molecular Dynamics Inc.), respectively. Each EMSA was repeated three times independently in order to verify the consistency of the results shown in Figures 6–9.

RESULTS

As a first step in the functional analysis of OSMAR transposases and the interactions between *Osmar* and *Stowaway* elements, it was necessary to identify active or recently active rice elements. Almost all of the 34 *Osmar* elements in the rice genome encode inactive transposases as evidenced by the presence of stop codons, deletions and frameshift mutations in the coding regions [(16) and data not shown]. Among all element families, *Osmar5* and *Osmar10* are the best candidates for recent activity because each family is represented by nearly identical copies. Two copies of *Osmar5* were identified in the *indica* genome, while a survey of the *O.sativa ssp. japonica* genome sequence revealed a single *Osmar5* element located at a different position than the two *indica* elements (Figure 2A). The three *Osmar5* copies are >99.5% identical over their entire length. There are three copies of *Osmar10* in the *Oryza sativa ssp. indica* genomic sequence (with respect to chromosomal position) that share >96% nucleotide identity over their entire length (3.3 kb).

Sequence similarity between the TIRs and transposases of *Osmar5* and *Osmar10* suggests a common ancestry, although there is no significant homology between the elements in their subterminal noncoding regions (Figure 2B, light and dark gray regions). Based on a multiple alignment of the full-length transposases, phylogenetic analysis resolves the *Osmar* transposons into three clades [Figure 1 and (16)]. *Osmar5* falls within the B clade, which contains the largest number and diversity of *Osmar* elements, while *Osmar10* is one of only two members of the C clade. Thus, analysis of the specificity of transposase–DNA interactions for both *Osmar5* and *Osmar10* will permit, for the first time, inter-clade comparisons of related but distinct TE families co-existing in one genome.

Prior to the analysis of DNA binding, the open reading frames (ORFs) of *OSMAR5* and *OSMAR10* were subcloned and modified to generate potentially active proteins (Figure 2A). Because the three genomic copies of *Osmar5* contain almost identical exons (there are only two single nucleotide polymorphisms) and no nonsense mutations, the three exons of *Osmar5B* were amplified from *indica* genomic DNA and fused into a single ORF of 502 amino acids (Figure 2A). The three copies of *OSMAR10* transposase genes were more divergent, with each containing one premature stop codon due to point mutations (asterisks in Figure 2A). For *Osmar10*, an intact ORF was constructed by ligating the first exon of *Osmar10B* with the second exon of *Osmar10A*. The resulting ORF encodes a 449 amino acid protein that is ~98% identical to the consensus transposase of the three *Osmar10* copies.

DNA-binding specificity of OSMAR5 and OSMAR10

As a first step in the characterization of the DNA-binding activity of OSMAR, a yeast one-hybrid assay utilizing two plasmids was employed. One plasmid contains the DNA sequences to be tested for transposase binding cloned upstream of a β-*galactosidase* reporter gene (β-*gal*). A second plasmid contains the transposase ORF that has been translationally fused to a B42 transcriptional activation domain (24). In a yeast cell harboring both plasmids, binding of transposase to the DNA sequence of interest brings the activation domain near the reporter gene promoter, leading to β-*gal* expression and blue pigmented yeast colonies.

The reconstructed full-length ORFs of *Osmar5* and *Osmar10* were assayed independently for binding to their respective 30 bp TIRs. As shown in Figure 3, combining the OSMAR5 expression plasmid with a reporter plasmid harboring the *Osmar5* 5′ TIR resulted in a 4- to 5- fold increase in β-*gal* expression over background levels (expression plasmid without the transposase ORF). Similarly, in the presence of the 5′ TIR from *Osmar10A*, OSMAR10 induced β-GAL activity ~3-fold over background levels (Figure 3). However, there was no significant increase in β-GAL activity when the OSMAR5 plasmid was combined with the *Osmar10* TIR and *vice versa*.

DNA-binding sites of OSMAR5

The results above indicate that OSMAR5 and OSMAR10 can bind to their corresponding TIRs, but that there is no cross-interaction at the level of DNA-binding between OSMAR5 and OSMAR10, two elements that belong to different clades of *Osmar* elements (as defined by phylogenetic analysis of their transposases, Figure 1). To investigate the possibility of cross-interaction between *Osmar* elements from the same clade, the DNA-binding activity and specificity of OSMAR5 were further dissected. To map the binding sites of OSMAR5 on the *Osmar5* transposon, the yeast one-hybrid assay was used to assess the interaction of OSMAR5 with partially overlapping DNA fragments covering the terminal regions of *Osmar5*A (Figure 4A). The strongest β-GAL activity relative to background levels (~6-fold increase) was observed with a reporter plasmid containing the first or last 30 bp of *Osmar5A*, which corresponds to the TIR of the element (Figure 4A, constructs A and G). This indicates that both the 5′ and 3′ TIR contain binding sites for OSMAR5 and that the three nucleotide mismatches between the TIRs have no significant effect on binding (Figure 4: compare levels of β-GAL activity for A and G). A weaker yet significant increase in β-GAL activity was observed when sequences overlapping and slightly internal to the TIRs were tested (Figure 4A, clones B and F). However, β-*gal* induction was not observed with constructs C and D, which contain sequences more distant from the 5′ TIR. In contrast, an ~2.5-fold increase in β-GAL activity over background level was observed with a construct containing sequences that span the 30 bp located internal to the 3′ TIR (Figure 4A, clone E). This result suggests that there is an additional transposase-binding site in the 3′ subterminal region of *Osmar5*.

Alignment of the sequences for construct A with the reverse complement of constructs G and E revealed a well-conserved 17 bp motif (Figure 4B, denoted in blue). Each copy of the

motif differs from a consensus sequence (derived from the ends of all *Osmar5* copies) at only 1, 2 or 3 positions (Figure 4B, red nucleotides). Sequence analysis of *Osmar10* elements also
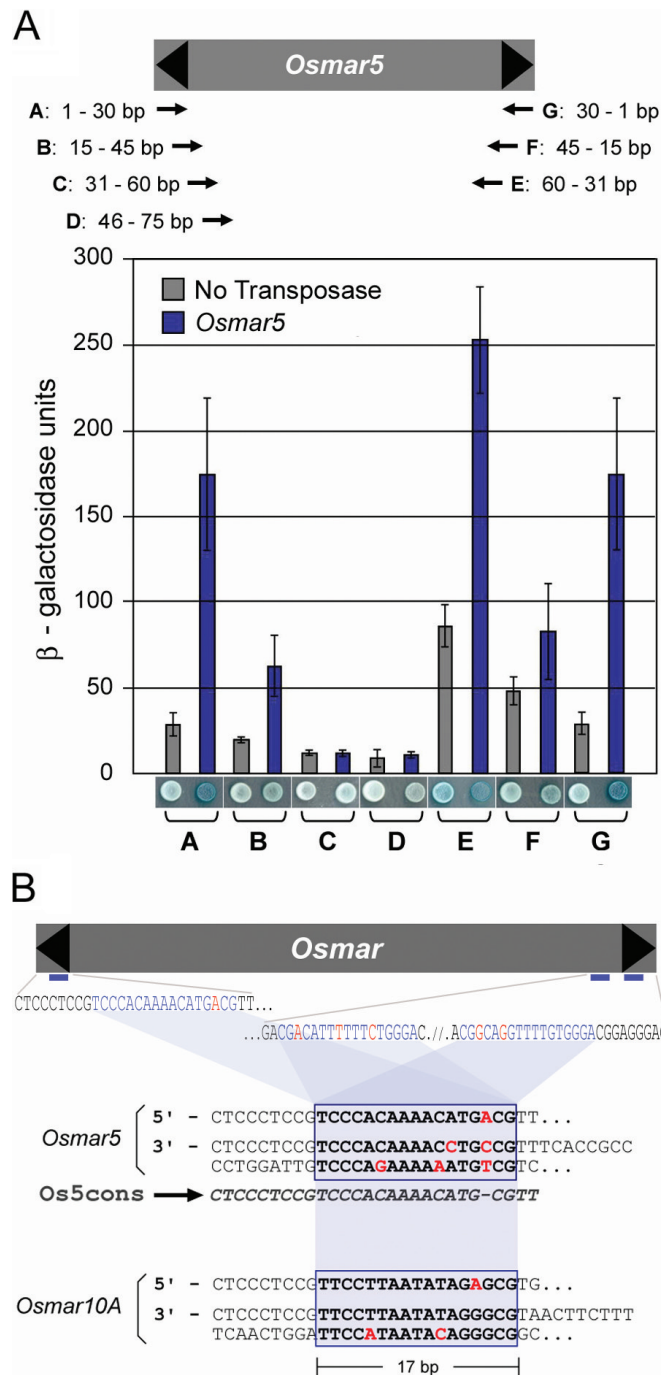

Figure 4. OSMAR5 binding sites in *Osmar5*. (**A**) Mapping of OSMAR5 binding sites using the yeast one-hybrid assay. See legend of Figure 3 for assay details. Positions of the 30 bp DNA sequences tested for interactions with OSMAR5 (relative to the *Osmar5* transposon ends) are shown in the diagram at the top. (**B**) The putative binding site sequences for OSMAR5. The diagram shows the position and organization of the sequences at the ends of a canonical *Osmar5* transposon. The 17 bp repeats are highlighted in blue and sequence variations within the repeats are in red. Arrows below the repeats indicate their relative orientation. Alignment of *Osmar5* and *Osmar10A* termini and a deduced *Osmar5* consensus TIR sequence (Os5cons) are shown.

revealed the presence of a conserved 17 bp motif between the 5′ and 3′ TIRs that is repeated upstream of the 3′ TIR (Figure 4B). In fact, comparison of the terminal regions of all B- and C-clade *Osmar* transposons led to the identification of a 17 bp 3′ subterminal repeat (data not shown). The organization of the repeats (location and spacing between the two 3′ sub-terminal repeats) is strongly conserved in all of these elements, but their sequences share variable degrees of relatedness, ranging from strong similarity within *Osmar* subclades (see below) to weak similarity between clades (for example compare the repeats of *Osmar5* and *Osmar10*, Figure 4B). In contrast, the subterminal 3′ repeats were conspicuously absent from A-clade *Osmar* transposons. Together these data indicate that *Osmar* elements of the B and C clades contain at least three transposase-binding sites, one within each TIR and a third in the 3′ subterminal region.

DNA-binding domain of OSMAR5 transposase

Computational analyses of full-length OSMAR transposases in the rice genome sequence identified conserved protein domains in their N-terminal and C-terminal halves (Figure 5A). A putative catalytic domain with a DD39D motif, typical of plant MLEs, was previously reported in the C-terminus (15,37). The N-terminal halves are more divergent than the C-termini, but secondary structure prediction programs identified two putative helix-turn-helix motifs (HTH1 and HTH2) in most of the OSMARS, including OSMAR5 and OSMAR10 (see Methods and Figure 5A).

The presence of HTH motifs in the N-terminal region is reminiscent of the structure of other Tc1/*mariner* transposases
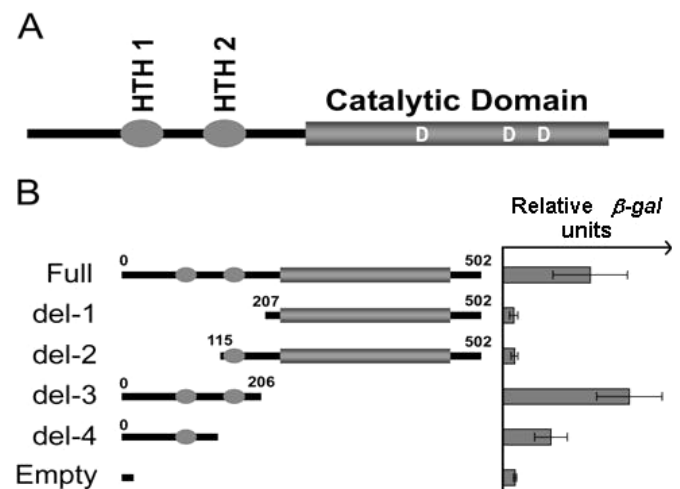
Figure 5. OSMAR5 binds DNA through a N-terminal domain containing two HTH motifs. (**A**) Computational predictions of the domain structure of OSMAR transposases. The N-terminal half of the protein is predicted to contain two helix-turn-helix motifs (HTH1 and HTH2) while the C-terminal half contains a putative catalytic domain with a DD (39) D motif. (**B**) DNA-binding analysis of deleted versions of OSMAR5 transposase as quantified in the yeast one-hybrid assay by β-GAL activity (details in Figure 3 legend). Full-length OSMAR5 and four deletion derivatives tested for DNA binding are shown along with the positions of their included amino acid residues relative to the full-length transposase. Error bars indicate standard deviations from the average values determined for eight replicates. The DNA used in the assays was the 30 bp *Osmar5* 5′ TIR (clone A in Figure 4A). 'Empty' denotes the use of an expression plasmid containing no transposase.

(18,26–28) and may indicate that OSMAR transposases bind DNA through an N-terminal domain. Using the yeast one-hybrid assay described above, four deletion constructs, encoding truncated versions of the OSMAR5 transposase, were tested for their ability to recognize and bind to the 5′ TIR of *Osmar5* (Figure 4A, construct A). As shown in Figure 5B, the C-terminal half of OSMAR5 (del-1; amino acids 207–502) did not induce a significant increase in β-GAL activity compared to an empty expression vector (background levels), whereas a construct containing only the first 206 amino acids of OSMAR5 induced β-GAL activity at similar or higher levels than full length OSMAR5 (Figure 5B, del-3). Results obtained with construct del-4 indicate that significant DNA-binding activity is retained but substantially reduced with a peptide consisting of the first 114 N-terminal residues and containing only HTH1. In contrast, results with construct del-2, which contains residues 115–502 and only HTH2, do not show β-*gal* induction above background levels. This suggests that specific DNA binding of OSMAR5 transposase requires the first 114 residues at the N-terminus, which includes HTH1, and is likely reinforced by the presence of additional features within region 115–206, such as HTH2.

Mutational dissection of the OSMAR5N-binding site

To complement and extend the *in vivo* DNA-binding studies presented above, full-length and C-terminal truncated OSMAR5 proteins were used for *in vitro* EMSAs. Both proteins were purified using N-terminal fusions with MBP (see Methods). The C-terminal truncated OSMAR5N includes the N-terminal 206 residues shown previously to be sufficient for DNA binding in yeast one-hybrid assays (Figure 5B; del-3). Mixing either full-length or truncated (OSMAR5N) versions of OSMAR5 with a radioactively labeled double-stranded oligonucleotide corresponding to the 30 bp 5′ TIR of *Osmar5A* results in a strong band shift on a native polyacrylamide gel (Figure 6). This interaction of OSMAR5N with labeled *Osmar5* 5′ TIR can be competed away with increasing amounts of unlabelled *Osmar5* 5′ TIR but not with unlabelled *Osmar10* TIR. In addition, no band shift was detected when OSMAR5N was combined with labeled *Osmar10* TIR or when purified MBP alone was used (Figure 6). Thus, both OSMAR5 and OSMAR5N display specific DNA-binding activities *in vitro*. OSMAR5N was used exclusively in the experiments that follow because it was easier to purify.

To define the nucleotides within the TIRs that are most important for OSMAR5 binding, EMSA was employed to assess interactions between OSMAR5N and a variety of mutant *Osmar5* TIRs. Twenty-one sequence variants, cumulatively covering the entire TIR, were tested for DNA binding, and the abundance of DNA shifted during electrophoretic migration was quantified using a phosphorimager (Figure 7). An *Osmar5* consensus TIR (Os5cons) was generated (see Methods section) that differs from the 5′ TIR by 1 bp and from the 3′ TIR by 2 bp.

When compared with the interaction of OSMAR5N with Os5cons, mutations within the first 9 nt and the last 2 nt of the TIR have little effect (Mut-1 through Mut-4 and Mut-22, Figure 7). In contrast, mutations in two specific TIR domains led to significant reductions in the relative levels of interaction
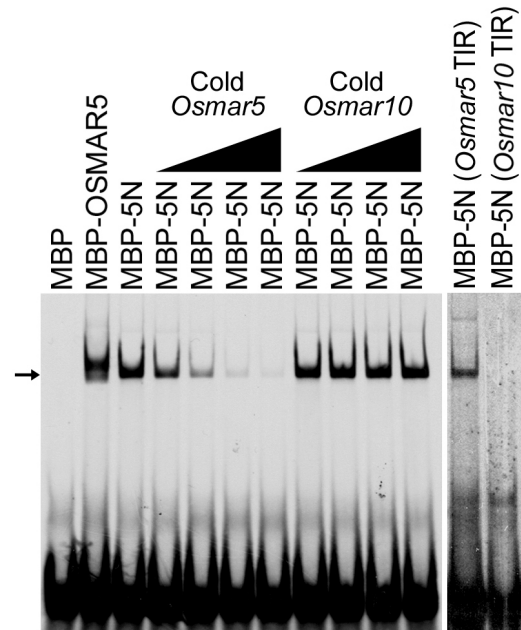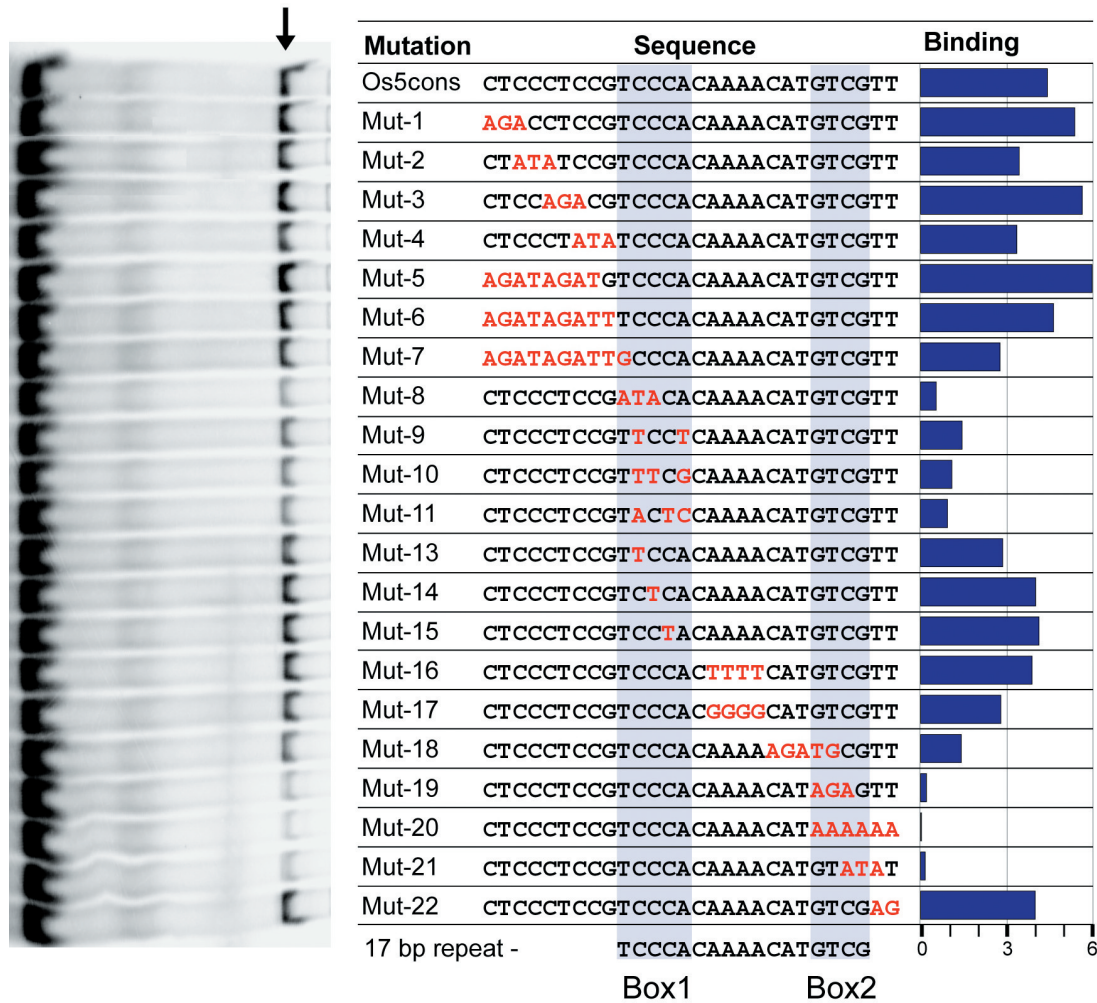

Figure 6. *In vitro* analysis of DNA binding of OSMAR5. The purified protein used in each reaction is indicated above the lane: MBP-OSMAR5: full-length OSMAR5 transposase fused to MBP; MBP-5N: a peptide consisting of the first 206 amino acids of OSMAR5 fused to MBP. The autoradiograph on the left shows EMSA with a radioactively labeled double-stranded 30 bp oligonucleotide corresponding to the *Osmar5* 5′ TIR. Black triangles above the lanes represent the amount of unlabelled *Osmar5* 5′ TIR ('Cold *Osmar5*') and unlabelled *Osmar10A* 5′ TIR ('Cold *Osmar10*') used in competition assays. The ratios of labeled to unlabelled DNA used in the competition assays are, from left to right, 1:2, 1:5, 1:10 and 1:20. Gel-shifted DNA is indicated by the arrow. The autoradiography on the right shows EMSA with an *Osmar5* 5′ TIR and an *Osmar10A* 5′ TIR.

with OSMAR5N (Figure 7, blue shading in Box1 and Box2). Although single base changes within these domains had little effect on OSMAR5N binding (Mut-13, Mut-14, Mut-15), two or more substitutions reduced the level of protein-DNA interaction as reflected by the intensity of the gel-shifted DNA (Figure 7, Mut-8 through Mut-11 and Mut-18 through Mut-21). These regions correspond precisely to the beginning and the end of the 17 bp binding motif that was identified using the one-hybrid assay (Figure 4). In contrast, the AT-rich spacer between the two boxes seems relatively insensitive to mutation. Replacing the central stretch of A nucleotides with G nucleotides only slightly reduces the amount of bound DNA, while replacing it with a stretch of T nucleotides has no significant effect (compare Mut-16 and Mut-17 to Os5cons). These results support the view that OSMAR5 transposase interacts with the *Osmar5* element through a 17 bp binding site, and suggest that specific protein-DNA contacts occurs predominantly within Box1 and Box2.

Interaction of OSMAR5N with the TIRs of other *Osmars*

Closely related *Osmar* transposons in the rice genome contain similar transposase sequences and similar TIRs, however, the subterminal domains of the elements typically show little or no homology. As an example, a schematic comparison of *Osmar5* with *Osmar2*, its closest relative in the available rice genome

Mutation	Sequence	Binding
Os5cons	CTCCCTCCGTCCCACAAAACATGTCGTT	
Mut-1	AGACCTCCGTCCCACAAAACATGTCGTT	
Mut-2	CTATATCCGTCCCACAAAACATGTCGTT	
Mut-3	CTCCAGACGTCCCACAAAACATGTCGTT	
Mut-4	CTCCCTATATCCCACAAAACATGTCGTT	
Mut-5	AGATAGATGTCCCACAAAACATGTCGTT	
Mut-6	AGATAGATTTCCCACAAAACATGTCGTT	
Mut-7	AGATAGATTGCCCACAAAACATGTCGTT	
Mut-8	CTCCCTCCGATACACAAAACATGTCGTT	
Mut-9	CTCCCTCCGTTCCTCAAAACATGTCGTT	
Mut-10	CTCCCTCCGTTTCGCAAAACATGTCGTT	
Mut-11	CTCCCTCCGTACTCCAAAACATGTCGTT	
Mut-13	CTCCCTCCGTTCCACAAAACATGTCGTT	
Mut-14	CTCCCTCCGTCTCACAAAACATGTCGTT	
Mut-15	CTCCCTCCGTCCTACAAAACATGTCGTT	
Mut-16	CTCCCTCCGTCCCACTTTTCATGTCGTT	
Mut-17	CTCCCTCCGTCCCACGGGGCATGTCGTT	
Mut-18	CTCCCTCCGTCCCACAAAAGATGCGTT	
Mut-19	CTCCCTCCGTCCCACAAAACATAGAGTT	
Mut-20	CTCCCTCCGTCCCACAAAACATAAAAAA	
Mut-21	CTCCCTCCGTCCCACAAAACATGTATAT	
Mut-22	CTCCCTCCGTCCCACAAAACATGTCGAG	
17 bp repeat -	TCCCACAAAACATGTCG	0 3 6

Box1   Box2

Figure 7. Binding of OSMAR5N with TIR variants. Sequences of an *Osmar5* consensus TIR (Os5cons) and 22 mutant TIRs used as substrates. Sequence variations within each mutant TIR relative to Os5cons are in red. EMSA was carried out with the OSMAR5N protein and each of the DNA substrates and the resulting autoradiograph is shown to the left of the corresponding sequence (an arrow denotes the position of gel-shifted DNA). Blue bars to the right of each sequence represent the amount of shifted DNA relative to free DNA for that sequence (the scale of 1 through 6 is arbitrary). The 17 bp consensus binding site identified in previous experiments along with Box1 and Box2 (see Figure 4B) are shown.

(see Figure 1), is shown in Figure 8A. While their transposase genes are 97% identical over their entire length (∼1.5 kb including introns), the terminal regions of the two transposons are ∼90 and ∼70% identical (excluding gaps) over the first ∼0.5 kb and the last ∼0.8 kb, respectively. The remaining, non-coding subterminal sequences share less than 60% nucleotide similarity and cannot be confidently aligned. It is noteworthy that the greatest similarity between the 3′ terminal regions is within the TIR and the internal 17 bp putative binding site, suggesting the possibility of cross-interaction between transposases and transposons that belong to the same clade.

This possibility of cross interaction was tested using EMSA to analyze the ability of OSMAR5N to bind oligonucleotides representing the TIRs of *Osmar2* and nine other *Osmar* elements (Figure 8B). In most cases, the TIR sequences are based on a consensus sequence derived from the alignment of the 5′ TIR, the 3′ TIR and the 17 bp 3′ subterminal repeat (if available) of all copies in the genome. Ambiguous positions were resolved as described in the Materials and Methods.

Also tested in these experiments were the two *Osmar5* TIRs (Os5-5′ and Os5-3′) as well as a synthetic sequence corresponding to the sequence of the 17 bp 3′ subterminal repeat of *Osmar5* fused to the first 9 canonical nucleotides of *Osmar* elements (Figure 8B, Os5-3′ sub). Results show that OSMAR5N can bind all three *Osmar5* TIR sequences at roughly equivalent levels consistent with prior results that a single nucleotide change within Box1 and Box2 of the OSMAR5-binding site has little effect on transposase binding. OSMAR5N also displays significant interactions with the TIRs of *Osmar2*, *Osmar24*, *Osmar17* and *Osmar21* and a weak interaction with the 5′ TIR of *Osmar28* (an element for which only the 5′ TIR was identified). Weak or no detectable interactions were seen with the TIRs of *Osmar1*, *Osmar4*, *Osmar9* and *Osmar14* (Figure 8B).

Interaction of OSMAR5N with *Stowaway* MITEs

A resemblance between the TIRs of *Stowaway* MITEs and *Osmar*s was previously noted (14,16). Although there is
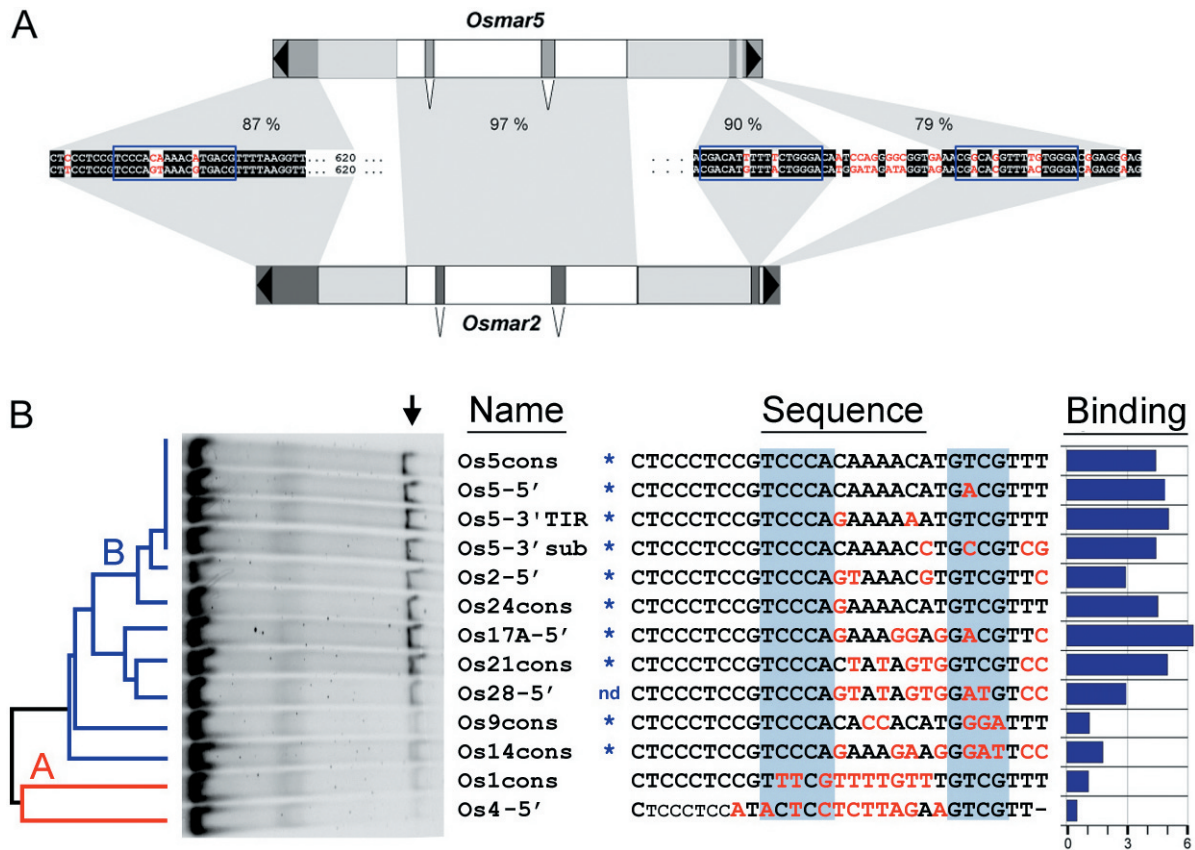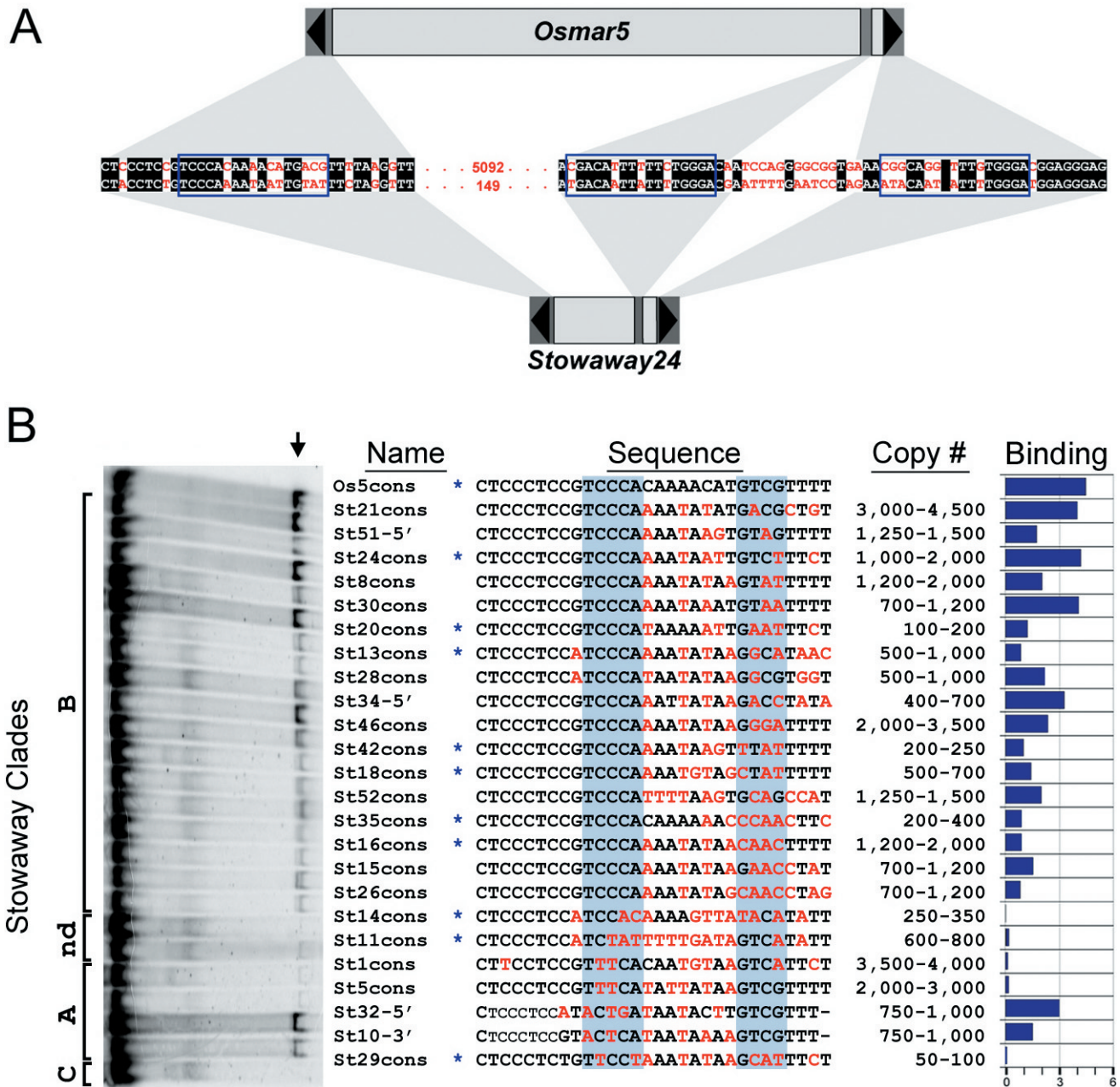
Figure 8. Binding of OSMAR5N with the TIRs of *Osmar* transposons. (**A**) Sequence and structural comparison of *Osmar5* and *Osmar2*. Regions with significant sequence similarity (>60%) are noted by gray shading. An alignment of the terminal sequences is shown, with identical nucleotides shaded in black. The putative 17 bp binding sites are in blue boxes. (**B**) EMSA of interactions between OSMAR5N and different *Osmar* TIRs. Thirteen TIR sequences, derived from 10 *Osmar* elements mined from the rice genome, are shown. If there was more than one element type in the genome, a consensus TIR sequence ('cons') was constructed based on a multiple alignment of the TIRs and the 3′ subterminal 17 bp repeat. Other substrates represent either one of the two TIRs or, in the case of Os5-3′ sub, the fusion of the first 9 bp of the TIR to the 3′ subterminal repeat of *Osmar5*. Sequence variations relative to Os5cons are in red and Box1 and Box2 (Figure 4) are shaded in blue. Blue asterisks denote *Osmar* transposons where a 3′ subterminal 17 bp repeat is recognizable (nd = not determined because the 3′ terminus was not available). The autoradiograph shows the results of EMSA with OSMAR5N and each of the 13 TIR sequences. Shifted DNA is indicated with an arrow and the amount of shifted DNA relative to free DNA is represented by the blue bars to the right of each sequence (the scale of 1 through 6 is arbitrary). The phylogenetic tree on the left side of the autoradiograph was generated from a multiple alignment of the DNA-binding domains of the corresponding *Osmar* transposases (see Methods). The A-clade (red) and B-clade (blue) correspond to the A and B clades shown in Figure 1.

no significant sequence similarity between *Osmars* and *Stowaways* outside of their TIRs, a recent comparative analysis of TIR sequences led to the assignment of most of the *Stowaway* MITE families to one of the three major clades of *Osmar* (16). For example, Figure 9A shows a comparison between *Osmar5* and a consensus sequence derived from the *Stowaway24* family. The TIRs of the *Stowaway24* consensus display greatest similarity (~70%) to those of *Osmar5* TIRs and as such it was proposed that this family may have been mobilized by OSMAR transposases from the B clade (16). In light of the results presented here, it can be seen that *Stowaway24* consensus also contains a 17 bp repeat sequence in its 3′ subterminal region that is similar to the binding site of *Osmar5* (Figure 9A, blue boxes). The spacer between the two 3′ subterminal repeats is conserved in length but is more divergent in sequence than the presumed binding sites. Further inspection of consensus sequences for 23 other *Stowaway* families revealed the presence of similar 17 bp repeats in the 3′ subterminal regions of nine families (denoted by blue asterisks in Figure 9B).

To test if OSMAR5 is capable of recognizing and binding to *Stowaway* elements, oligonucleotides that represent the TIRs of 24 different *Stowaway* families were tested using EMSA in the presence of OSMAR5N (Figure 9B). The EMSA results indicate that OSMAR5N strongly interacts *in vitro* with *St21cons* and *St24cons*, because the shifted DNA bands are of comparable intensity as those observed with the *Osmar5* consensus TIR (Os5cons, Figure 9B). OSMAR5N also displays significant binding with *St30cons*, *St34-5′* and *St32-5′*, although with somewhat reduced efficiency compared to Os5cons (Figure 9B). All other *Stowaway* TIR sequences tested displayed much weaker or no interaction with OSMAR5N (Figure 9B, e.g. *St28cons*, *St46cons*, *St29cons*).

DISCUSSION

While Tc1/*mariner* elements have been the subject of extensive genetic, evolutionary and biochemical characterization in animal species, their abundance and diversity in flowering

Figure 9. Binding of OSMAR5N with *Stowaway* TIRs. (**A**) Sequence comparison of *Osmar5* with the consensus sequence for the *Stowaway24* family of MITEs. Regions with significant sequence similarity (>60%) are noted by gray shading. An alignment of the terminal sequences is shown, with identical nucleotides between the elements in shaded boxes. The putative 17 bp binding sites are boxed in blue. (**B**) EMSA of the interactions between OSMAR5N and *Stowaway* TIRs. Twenty-four 30 bp oligonucleotides, each derived from the consensus TIR of a different *Stowaway* family, were tested by EMSA for interaction with OSMAR5N. See the legend of Figure 8 for description of 'cons' (here *Stowaway* not *Osmar* consensus) and Os5cons. Other sequences represent either one of the two TIRs. Sequence variations relative to Os5cons are indicated in red and Box1 and Box2 are shaded in blue. Blue asterisks mark *Stowaway* families where the 3' subterminal 17 bp repeat is recognizable. The autoradiograph shows the results of EMSA with OSMAR5N and each of the 25 TIR sequences. Shifted DNA is indicated with an arrow and the amount of shifted DNA relative to free DNA is represented by the blue bars to the right of each sequence (the scale of 1 through 6 is arbitrary). The approximate copy numbers of elements for each *Stowaway* family detected in the complete rice genome sequence (cv. Nipponbare) is given. The *Stowaways* tested were previously categorized into distinct clades (A, B and C), which are shown to the left (nd = not determined).

plants has only recently been appreciated (15). One of the distinguishing features of plant MLEs is their relationship with *Stowaway* MITES; an association restricted thus far to indirect evidence, such as sequence similarity of the TIRs and TSDs (6,14,16). In this study, we have exploited the output of our recent computational survey of all MLEs and *Stowaway* MITEs in rice (16) to initiate a functional characterization of plant MLEs and their encoded transposases and to test the hypothesis that MLE transposases interact with *Stowaway* MITEs.

DNA-binding activity of OSMAR5 and OSMAR10 transposases

A recent computational survey of the rice genome sequence revealed that MLEs are an ancient and diverse group of transposons (16). While 34 distinct *Osmar* elements were identified (Figure 1), none are known to be transpositionally active. *Osmar5* and *Osmar10* are among a few families that display signs of recent activity in the rice genome [(16), this study]. For each of the two families, we have reconstructed genes encoding possible active transposases based on a comparison of all copies identified in the databases (Figure 2). The resulting proteins, OSMAR10 and OSMAR5, were expressed in yeast cells and tested for their ability to recognize and bind 30 bp DNA sequences corresponding to the TIRs of their respective transposon. These experiments showed that OSMAR5 and OSMAR10 bind to TIR sequences from their own transposon. Furthermore, these interactions are specific because neither OSMAR5 nor OSMAR10 binds to the TIR sequence from the other transposon (Figure 3). These results indicate that the two reconstructed *Osmar* transposases display specific DNA-binding activities when expressed in a heterologous eukaryotic system. As such, yeast may be a suitable host for the functional study of plant MLEs, as it has been for studies of the *Ac/Ds* elements of maize (29,30).

OSMAR transposases contain a bipartite N-terminal DNA-binding domain

Having determined that both OSMAR10 and OSMAR5 display DNA-binding activities in yeast, we then sought to localize where binding occurred in these proteins. Computational predictions of the OSMAR5 and OSMAR10 secondary structures identified two HTH DNA-binding motifs (HTH1 and HTH2) in the N-terminal regions of each protein (Figure 5A). Yeast one-hybrid assays with deleted versions of OSMAR5 showed that a peptide consisting of the first 206 residues of OSMAR5 (OSMAR5N) and containing both HTH motifs is sufficient to confer a similar or even greater level of DNA-binding activity than the full-length transposase (Figure 5B, del-3). Conversely, deleting the first 206 residues from a full-length transposase construct completely abolished DNA-binding to the *Osmar5* TIR (Figure 5B, del-1). Furthermore, a peptide that contains the first 114 residues of OSMAR5 (and therefore only HTH1) still displays significant DNA-binding activity, although with a significantly reduced level compared to OSMAR5N (Figure 5B, del-4). Finally, deleting the first 114 residues from a full-length construct (and thus maintaining only HTH2) results in the complete loss of DNA-binding activity (Figure 5B, del-2). Together these results suggest that OSMAR5 binds DNA through a bipartite DNA-binding domain, which is located within the first 206 N-terminal residues and includes two functionally distinct HTH motifs. The first HTH appears necessary and sufficient for specific DNA binding, while the second HTH may strengthen the interaction. Because these two HTH motifs were identified in almost all OSMAR transposases in the rice genome (data not shown), it is likely that the binding features of OSMAR5 are representative of rice MLEs.

The structure and mode of action of the OSMAR5 DNA-binding domain is reminiscent of the so-called paired domain, which is found near the N-terminus of Tc1-like transposases and *Pax* transcription factors in metazoans (28,31–33). Like the DNA-binding domain of OSMAR5, the paired domain consists of two functionally separable HTH motifs. Functional and structural analyses have shown that specific DNA-protein contacts are usually mediated by the more N-terminal HTH motif and may be reinforced by a second HTH motif (25,28,34,35). In contrast to Tc1-like transposases, sequence and biochemical analysis of two insect *mariner*-like transposases (*Mos1* and *Himar1*) indicates that they contain and interact with DNA through a single HTH motif (17–19,26).

Tc1-like and *mariner*-like transposases have been identified in a broad range of animals and phylogenetic analyses have shown that they define two separate monophyletic groups within the Tc1/*mariner* superfamily (27,36). Phylogenetic analyses have revealed that plant *mariner*-like transposases form another distinct monophyletic group of sequences within the superfamily, which appears equally distant from Tc1-like and *mariner*-like transposases from animals (15,37,38). The plant proteins have been designated *mariner*-like transposases following the original classification of the soybean *Soymar1* and rice *Osmar1* as MLEs (39,40). This initial classification was based on the results of BLAST searches where the plant proteins would first hit animal *mariner*-like transposases (39) and on the fact that both groups display a conserved DDD motif instead of the canonical DDE catalytic triad found in Tc1-like transposases and most other recombinases of the so-called DDE 'megafamily' (note however that plant *mariner*-like transposases display a DD39D motif, while the motif is DD34D in typical animal *mariner*-like transposases) (38,41). The functional and structural similarity of the DNA-binding domain of OSMAR5 and the paired domain revealed by our analysis indicate that plant MLE transposases may in fact be evolutionarily closer to the Tc1-like group of transposases within the Tc1/*mariner* superfamily than to the animal *mariner*s.

Binding specificity of OSMAR5N

Analyses of the binding of OSMAR5N to multiple variants of the *Osmar5* TIR indicated that mutations in the first 9 bp had a minimal effect on binding (Figure 7, Mut-1 to Mut-6). In contrast, base substitutions in one of two domains (called Box1 and Box2) had a significant effect on binding, suggesting that the sequences of Box1 and Box2 provide most of the binding specificity to the *Osmar5* elements (Figure 7, Mut-8 to Mut-11). Box1 corresponds to a sequence motif in the TIR shown previously to be conserved within, but not between, the majors *Osmar* clades [see Figure 1 and (16)]. In addition, when Box2 is superimposed on the previously published alignment of *Osmar* TIRs (16), it becomes apparent that it corresponds to a second region of the TIR that is relatively well conserved within but not between *Osmar* clades (Figure 8B and data not shown). In contrast, the spacer region between Box1 and Box2 is much less conserved, even when closely related elements from the same clade are compared.

Within-clade conservation of Box1 and Box2 suggested that a given *Osmar* transposase might interact with the TIRs of other *Osmar* elements in the same clade. Indeed, OSMAR5N was able to bind to the consensus TIR sequences of all *Osmars* that were previously defined as members of the B1 subclade

(Figures 1 and 8), but not with the TIRs from elements in other *Osmar* clades and subclades (Figure 8B). Recall that clade assignments were based on phylogenetic analyses of *Osmar* transposases not their TIRs (Figure 1). In this study we show that these clades also have functional significance, that is, the transposase encoded by one clade member can bind to the TIRs of other clade members but not to the TIRs of elements in other clades. Taken together these results provide support for the view that *Osmar* transposases are coevolving with their binding sites in the TIRs (16).

OSMAR5N binding to *Stowaway* TIRs: implications for the mechanism of MITE amplification

Previous conclusions about the relationship between *Osmar* elements and *Stowaway* MITEs were derived solely from a comparative analysis of their sequences (16). About 22 000 *Stowaway* MITEs were classified into 36 families based on their TIR sequences and 34 *Osmars* were grouped into 3 major clades based on transposase sequence identity. As discussed above, the *Osmar* clade assignments also reflected TIR sequence similarity. Because sequence similarity between *Osmars* and *Stowaways* was found to be largely restricted to their TIRs, TIR sequence similarity was used to pair an *Osmar* clade with a *Stowaway* family. These pairings led us to propose that *Stowaway* MITEs were mobilized by *Osmar* transposases encoded by distantly related elements (7,16). Results from this study strengthen this model by demonstrating for the first time specific binding between an *Osmar* transposase-binding domain and *Stowaway* MITEs and that the strength of binding largely correlates with TIR similarity.

Osmar5 belongs to clade B (Figure 1). Comparison of consensus TIR sequences between *Osmar* and *Stowaway* elements had previously identified 23 *Stowaway* families that may have been mobilized by transposase(s) from this clade (16). Seventeen different TIR consensus sequences, derived from these 24 families, were assayed for binding with OSMAR5N and 4 were strongly bound by OSMAR5N (Figure 9B: St21cons, St24cons, St30cons, St34-5′). In contrast, most of the *Stowaway* TIRs that showed weaker interactions with OSMAR5N had identical Box1 sequences as Os5cons but differed in their Box2 sequence (Figure 9B, e.g. St16cons, St20cons and St26cons). Based on the binding of OSMAR5N with the 17 sequences, nothing can be concluded about the role of Box1 (because it was invariant). However variation in the Box2 sequence appeared to influence the strength of binding.

Surprisingly, interactions were detected between two *Stowaway* families that had previously been assigned to clade A. A robust interaction was detected with the consensus TIR of the *Stowaway32* family and a weaker interaction occurred with the TIR of the *Stowaway10* family (Figure 9B). These families were originally assigned to clade A based on the similarity of their Box1 sequences [see Figure 9B and (16)]. However, it is noteworthy that the Box2 of each family consensus sequence is identical to that of the *Osmar5* consensus TIR (Figure 9B). It may be that these families were previously misclassified and are in fact mobilized by clade B transposases, despite the substantial divergence of their Box1 motifs. As such, this result highlights the importance of experimental verification of hypothesis based solely on sequence comparisons.

Our results provide the first direct evidence for a physical interaction of a transposase with a plant MITE *in vitro*. Furthermore, we have shown that a single source of transposase (OSMAR5N) can potentially interact with MITE sequences representing several distinct families. Binding of transposase to the transposon ends is the first step of any transposition reaction and therefore it is a key event in determining the specificity of a transposase for its substrate (3). Future experiments will be required to determine whether the binding activity detected in our study are sufficient to engage the corresponding transposons into the subsequent steps of a full transposition reaction (including nicking and strand transfer activities).

This study reveals one mechanism by which MITEs could rapidly spread in a plant genome. The *Osmar5* family is one of the most recently active families in the rice genome and its encoded transposase has DNA-binding activity. The results of *in vitro* binding studies suggest that this transposase could interact *in vivo* with many *Stowaway* elements currently in the rice genome. With such a large population of *Stowaway* elements to potentially interact with, the introduction of an active OSMAR element into a genome (or the activation of a silenced element) could be accompanied by the amplification of one or more *Stowaway* founder elements into new MITE subfamilies. A further outcome of this model is a 'snowball effect', where each new wave of MITE amplification enlarges the reservoir of substrates poised to exploit the next source of active transposase that emerges in the genome.

REFERENCES

1. Capy,P., Bazin,C., Higuet,D. and Langin,T. *Dynamics and Evolution of Transposable Elements*. Springer-Verlag, Austin, Texas.
2. Feschotte,C., Jiang,N. and Wessler,S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.*, **3**, 329–341.
3. Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (2002) *Mobile DNA II*. American Society for Microbiology Press, Washington, DC.
4. Bureau,T.E. and Wessler,S.R. (1992) *Tourist*: a large family of inverted-repeat element frequently associated with maize genes. *Plant Cell*, **4**, 1283–1294.
5. Bureau,T.E. and Wessler,S.R. (1994) *Stowaway*: a new family of inverted-repeat elements associated with genes of both monocotyledonous and dicotyledonous plants. *Plant Cell*, **6**, 907–916.
6. Feschotte,C., Zhang,X. and Wessler,S. (2002) Miniature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. American Society for Microbiology Press, Washington, DC, pp. 1147–1158.

7. Jiang,N., Feschotte,C., Zhang,X.Y. and Wessler,S.R. (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.*, **7**, 115–119.

8. Mazzone,M., De Gregorio,E., Lavitola,A., Pagliarulo,C., Alifano,P. and Di Nocera,P.P. (2001) Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic Neisseriae. *Gene*, **278**, 211–222.

9. Redder,P., She,Q. and Garrett,R.A. (2001) Non-autonomous mobile elements in the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol.*, **306**, 1–6.

10. Buisine,N., Tang,C.M. and Chalmers,R. (2002) Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic Neisseria sp. *FEBS Lett.*, **522**, 52–58.

11. Kapitonov,V.V. and Jurka,J. (1999) Molecular paleontology of transposable elements from Arabidopsis thaliana. *Genetica*, **107**, 27–37.

12. 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. USA*, **98**, 12572–12577.

13. Zhang,X., Jiang,N., Feschotte,C. and Wessler,S.R. (2004) Distribution and evolution of *PIF*- and *Pong*-like transposons and their relationships with *Tourist*-like MITEs. *Genetics*, **166**, 971–986.

14. Turcotte,K., Srinivasan,S. and Bureau,T. (2001) Survey of transposable elements from rice genomic sequences. *Plant J.*, **25**, 169–179.

15. Feschotte,C. and Wessler,S.R. (2002) *Mariner*-like transposases are widespread and diverse in flowering plants. *Proc. Natl Acad. Sci. USA*, **99**, 280–285.

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* MITEs. *Genetics*, **163**, 747–758.

17. Lampe,D.J., Churchill,M.E. and Robertson,H.M. (1996) A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J.*, **15**, 5470–5479.

18. Zhang,L., Dawson,A. and Finnegan,D.J. (2001) DNA-binding activity and subunit interaction of the mariner transposase. *Nucleic Acids Res.*, **29**, 3566–3575.

19. Auge-Gouillou,C., Hamelin,M.H., Demattei,M.V., Periquet,G. and Bigot,Y. (2001) The ITR binding domain of the Mariner Mos-1 transposase. *Mol. Genet. Genomics*, **265**, 58–65.

20. De Aguiar,D. and Hartl,D.L. (1999) Regulatory potential of nonautonomous mariner elements and subfamily crosstalk. *Genetica*, **107**, 79–85.

21. Auge-Gouillou,C., Hamelin,M.H., Demattei,M.V., Periquet,M. and Bigot,Y. (2001) The wild-type conformation of the Mos-1 inverted terminal repeats is suboptimal for transposition in bacteria. *Mol. Genet. Genomics*, **265**, 51–57.

22. Lampe,D.J., Walden,K.K. and Robertson,H.M. (2001) Loss of transposase-DNA interaction may underlie the divergence of mariner family transposable elements and the ability of more than one mariner to occupy the same genome. *Mol. Biol. Evol.*, **18**, 954–961.

23. Dodd,I.B. and Egan,J.B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.*, **18**, 5019–5026.

24. Finley,R.L.,Jr and Brent,R. (1994) Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators. *Proc. Natl Acad. Sci. USA*, **91**, 12980–12984.

25. Jun,S. and Desplan,C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639–2650.

26. Pietrokovski,S. and Henikoff,S. (1997) A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Mol. Gen. Genet.*, **254**, 689–695.

27. Plasterk,R.H.A., Izsvák,Z. and Ivics,Z. (1999) Resident aliens: the Tc1/*mariner* superfamily of transposable elements. *Trends Genet.*, **15**, 326–332.

28. Izsvak,Z., Khare,D., Behlke,J., Heinemann,U., Plasterk,R.H. and Ivics,Z. (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. *J. Biol. Chem.*, **277**, 34581–34588.

29. Weil,C.F. and Kunze,R. (2000) Transposition of maize Ac/Ds transposable elements in the yeast *Saccharomyces cerevisiae*. *Nature Genet.*, **26**, 187–190.

30. Yu,J., Marshall,K., Yamaguchi,M., Haber,J.E. and Weil,C.F. (2004) Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **24**, 1351–1364.

31. Franz,G., Loukeris,T.G., Dialektaki,G., Thompson,C.R. and Savakis,C. (1994) Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc. Natl Acad. Sci. USA*, **91**, 4746–4750.

32. Gehring,W.J. (2002) The genetic control of eye development and its implications for the evolution of the various eye-types. *Int. J. Dev. Biol.*, **46**, 65–73.

33. Pichaud,F. and Desplan,C. (2002) Pax genes and eye organogenesis. *Curr. Opin. Genet. Dev.*, **12**, 430–434.

34. Xu,W., Rould,M.A., Jun,S., Desplan,C. and Pabo,C.O. (1995) Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell*, **80**, 639–650.

35. van Pouderoyen,G., Ketting,R.F., Perrakis,A., Plasterk,R.H. and Sixma,T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA. *EMBO J.*, **16**, 6044–6054.

36. Robertson,H.M. (1995) The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.*, **41**, 99–105.

37. Shao,H. and Tu,Z. (2001) Expanding the diversity of the IS630-Tc1-*mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics*, **159**, 1103–1115.

38. Robertson,H.M. (2002) Evolution of DNA transposons. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. American Society for Microbiology Press, Washington, DC, pp. 1093–1110.

39. Jarvik,T. and Lark,K.G. (1998) Characterization of *Soymar*1, a *mariner* element in soybean. *Genetics*, **149**, 1569–1574.

40. 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**, 381–391.

41. Chandler,M. and Mahillon,J. (2002) Insertion sequences revisited. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. American Society for Microbiology Press, Washington, DC, pp. 305–366.