Supporting Information

Lorenzen *et al*. 10.1073/pnas.0800444105

Fig. S1. DNA sequence and conceptual translation of the *Tc1–1* element in the M^1 insertion. The sequence begins and ends with the duplicated TA target motif in bold. The 35-nt inverted terminal repeats (ITRs) are underlined, with palindromic regions shaded. Authentic (TAA) and premature (TGA) stop codons are in bold. The otherwise fully intact ORF is interrupted by a single C→T transition that converts R to the latter stop codon at nucleotide 781. The diagnostic residues of the catalytic DD34E motif are boxed.

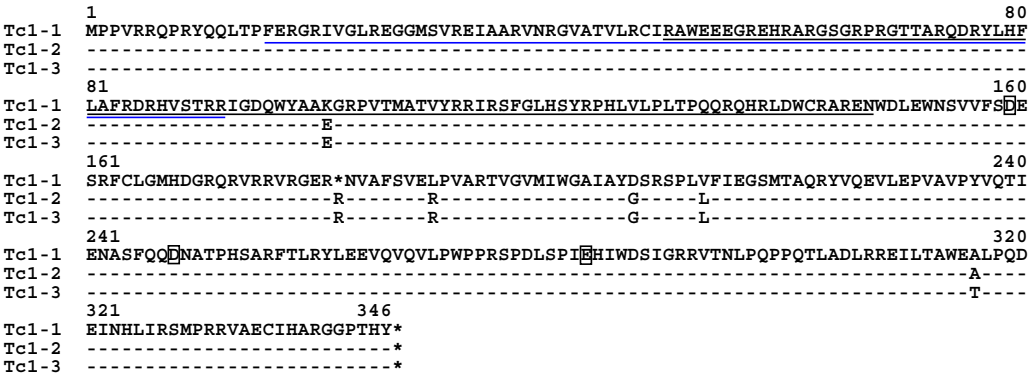

Fig. S2. Alignment of Tc1 proteins. Tc1–1: translation of transposase gene in the M^1-associated *Tc1* insertion (LG3). Tc1–2 and Tc1–3 are closely related elements found in a non-M^1 (GA2) genome. Tc1–2: translation of transposase gene found within an intron of EST DT789639 (GLEAN_00717; LG2). Tc1–3: translation of transposase gene in the *Tc1* element corresponding to GLEAN_02110 (LG unknown). *, stop codon. Dashes indicate identity with Tc1–1. Note premature stop codon in Tc1–1 at residue 182. Boxed residues: Tc1-DD34E motif [Shao H, 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]. Black underline: transposase_5 DNA-binding domain (pfam01498). Blue underline: COG3415 DNA binding domain.

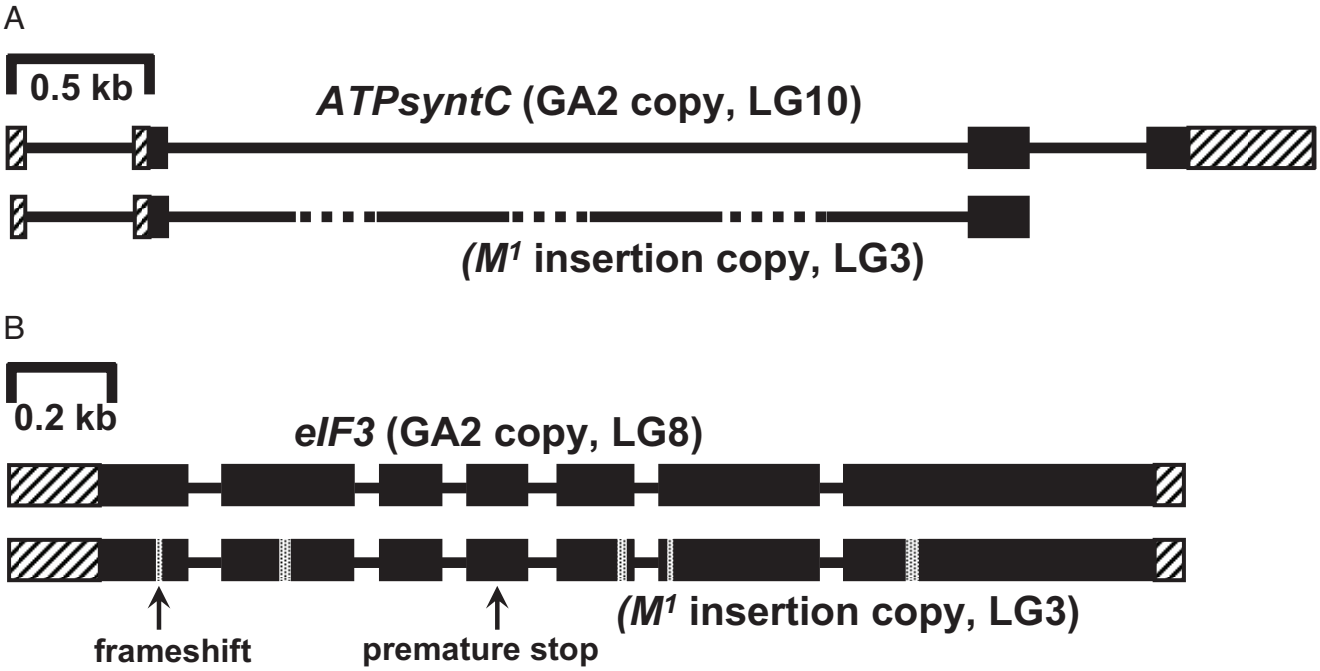

Fig. S3. Two gene fragments from the M^1-associated *Tc1* insertion and their functional counterparts. (*A*) An apparently functional copy of the *ATP synthase subunit C* (*ATPsyntC*) gene on LG10 of the non-*Medea* strain GA2 (*Upper*) and the truncated copy from the M^1 insertion (*Lower*). (*B*) An apparently functional copy of *elongation initiation factor 3* (*eIF3*) on LG8 of the GA2 strain and the full-length but defective copy from the M^1 insertion. Thick bars represent exons, and thin bars represent introns. Diagonal cross-hatching indicates UTRs. Shaded portions of exons are deleted in the M^1 copy of *eIF3*. Dashed lines indicate deleted portions of the large intron in the M^1 copy of *ATPsyntC*.

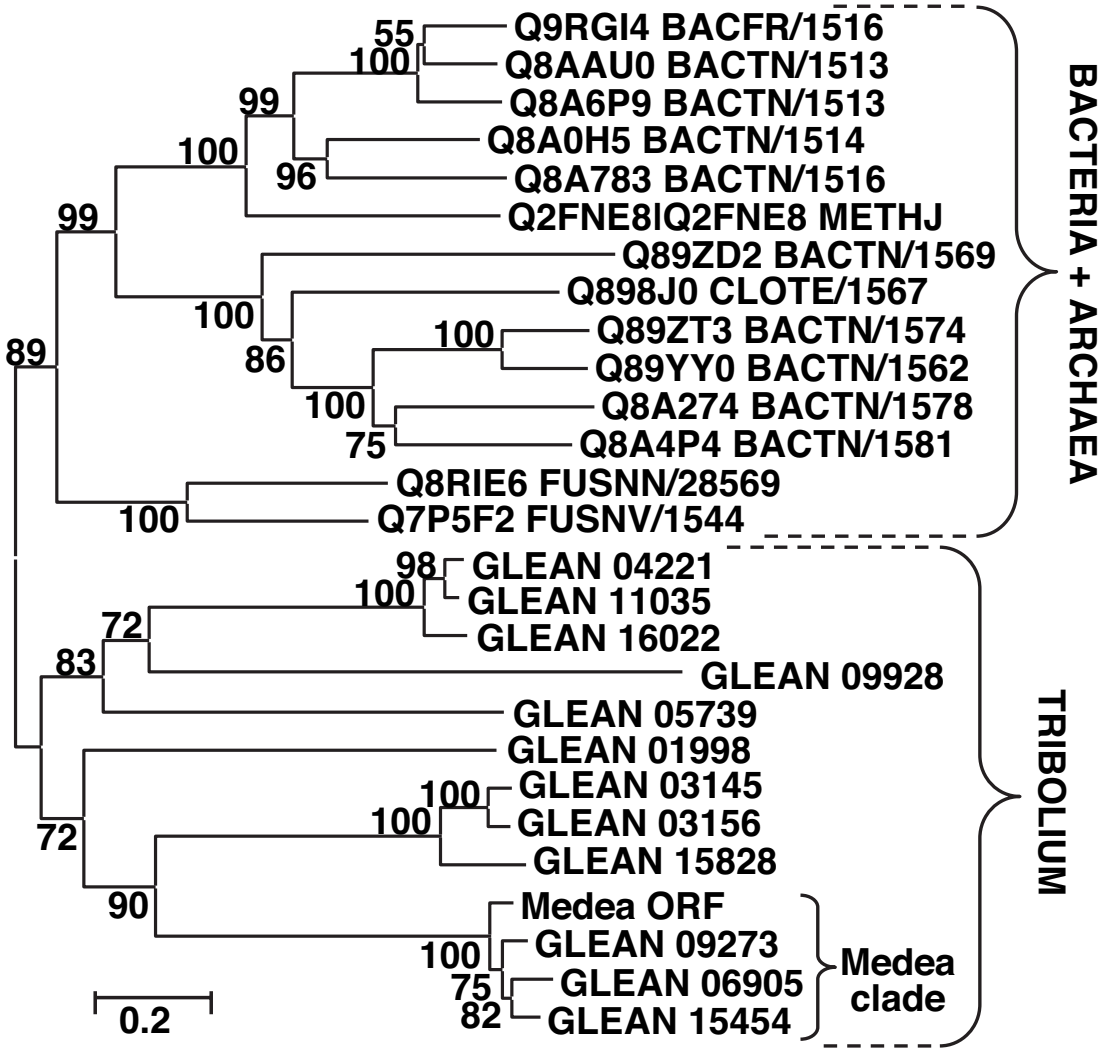

Fig. S4. Phylogenetic analysis of DUF1703 proteins. Shown are unrooted neighbor-joining trees based on multiple alignment of complete amino acid sequences with bootstrap values for 1,000 replications indicated. Accession numbers are given for the representative 14 bacterial/archeael proteins, and GLEAN numbers are given for *Tribolium* proteins. Scale indicates distance, rate of change per amino acid sequence.

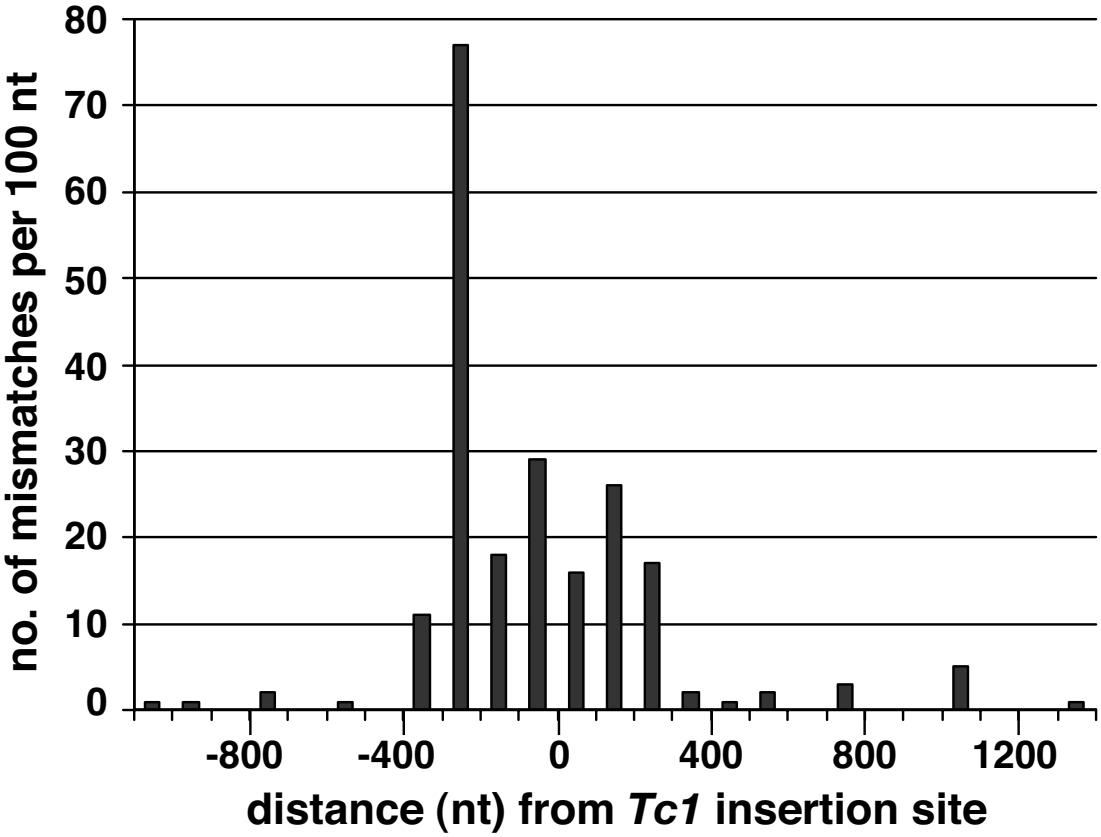

Fig. S5. M^1-associated sequence variation flanking the *Tc1* insertion site. Homologous M^1 and non-M^1 (GA2) sequences were aligned by using AlignX. The *x* axis is the distance in nucleotides from the Tc1 insertion point. The *y* axis represents the number of nonidentical residues in each contiguous 100-nt segment. The large strain difference in the segment 200–300 nt centromeric (to the left) of the insertion point reflects the copy number difference in the tandem repeat (five copies in non-M^1, approximately two copies in M^1).