CMLS, Cell. Mol. Life Sci. 59 (2002) 2097–2111
1420-682X/02/122097-15

CMLS Cellular and Molecular Life Sciences

Review

Dictyostelium mobile elements: strategies to amplify in a compact genome

T. Winckler[a,*], T. Dingermann[a] and G. Glöckner[b]

[a] Institut für Pharmazeutische Biologie, Universität Frankfurt/M. (Biozentrum), Marie-Curie-Strasse 9, 60439 Frankfurt am Main (Germany), Fax +49 69 798 29662, e-mail: Winckler@em.uni-frankfurt.de
[b] IMB Jena, Department of Genome Analysis, Beutenbergstrasse 11, 07745 Jena (Germany)

Received 10 May 2002; received after revision 10 June 2002; accepted 12 June 2002

Abstract. *Dictyostelium discoideum* is a eukaryotic microorganism that is attractive for the study of fundamental biological phenomena such as cell-cell communication, formation of multicellularity, cell differentiation and morphogenesis. Large-scale sequencing of the *D. discoideum* genome has provided new insights into evolutionary strategies evolved by transposable elements (TEs) to settle in compact microbial genomes and to maintain active populations over evolutionary time. The high gene density (about 1 gene/2.6 kb) of the *D. discoideum* genome leaves limited space for selfish molecular invaders to move and amplify without causing deleterious mutations that eradicate their host. Targeting of transfer RNA (tRNA) gene loci appears to be a generally successful strategy for TEs residing in compact genomes to insert away from coding regions. In *D. discoideum*, tRNA gene-targeted retrotransposition has evolved independently at least three times by both non-long terminal repeat (LTR) retrotransposons and retrovirus-like LTR retrotransposons. Unlike the nonspecifically inserting *D. discoideum* TEs, which have a strong tendency to insert into preexisting TE copies and form large and complex clusters near the ends of chromosomes, the tRNA gene-targeted retrotransposons have managed to occupy 75% of the tRNA gene loci spread on chromosome 2 and represent 80% of the TEs recognized on the assembled central 6.5-Mb part of chromosome 2. In this review we update the available information about *D. discoideum* TEs which emerges both from previous work and current large-scale genome sequencing, with special emphasis on the fact that tRNA genes are principal determinants of retrotransposon insertions into the *D. discoideum* genome.

Key words. *Dictyostelium*; retrotransposon; mobile element; tRNA; integration; genomics.

Transposable elements: a brief introduction

'This thing was moving around' – B. McClintock, 1948

It was in the 1940s when Barbara McClintock developed the concept of 'controlling elements'. They were defined as genetic units that associated with and thereby controlled the expression of genes [1]. Instability of effects of such controlling elements on gene expression were shown by McClintock to be caused by instability of the controlling elements in terms of their genomic positions. These studies led to the concept of transposable elements (TEs) as genetic entities that migrate through the genome of a host cell. The discovery of TEs in a broad range of organisms has brought about an intense debate over the probable sense or nonsense of this sort of mobile DNA. TEs can be regarded as molecular parasites that invade genomes and replicate themselves autonomously to sustain active populations within a given host [2, 3]. Although it seems true that TEs are selfish pieces of DNA, there is overwhelming evidence suggesting that TEs deeply influence their host cell genomes, both detrimen-

* Corresponding author.

tally and beneficially. The most drastic consequence of TE action we could imagine is the eradication of the host cell through disastrous insertion of TEs into essential genes. However, TEs may also exert more subtle beneficial effects on their host genomes, e.g. by offering sites for unequal cross-overs or by shuffling of genomic sequences to create new genes [4–6]. The host's genome flexibility may be significantly improved if the parasitic TEs are allowed to 'live' and expand to a certain extent, such that the host can adapt to environmental changes more efficiently than its competitors [4, 5, 7–9]. In addition, there are several examples of 'molecular domestication' of TE-encoded cis-regulatory proteins into cellular functions [10], e.g. the V(D)J recombination-mediating proteins RAG1 and RAG2 in the development of the vertebrate immune system [11], components of nuclear spliceosomes in eukaryotic RNA processing [12] or telomerase in chromosome maintenance [13]. The interaction of the molecular parasites with their hosts appears to be of such evolutionary significance that TEs have been widespread and may even be ubiquitous inhabitants of all prokaryotic and eukaryotic cells.

Classification of eukaryotic TEs

TEs are classified into two major groups according to their transposition mechanisms [14]. The class I, or retroelements, are – in the broadest sense – genetic entities originally derived by reverse transcription of RNA intermediates into DNA by means of an RNA-directed DNA polymerase (reverse transcriptase, RT). By contrast, class II elements move directly from DNA to DNA locus by a cut-and-paste mechanism catalyzed by an encoded transposase. Eukaryotic retrotransposons can be divided into autonomous and nonautonomous elements. Retrotransposons are considered autonomous if they encode intact proteins for reverse transcription and integration. The autonomous retrotransposons are classified according to the presence or absence of long terminal repeats (LTRs) bracketing the retroelement-encoded genes (LTR and non-LTR retrotransposons, respectively). Nonautonomous retroelements are mainly represented by the short interspersed nuclear elements (SINEs). SINEs have no protein-coding capacities and are probably molecular hitchhikers that mimick the 3′ ends of non-LTR retrotransposons and thereby ensure their trans-mobilization by borrowing the required proteins from residing autonomous non-LTR retrotransposons ([9], but see also discussion by [15]).

DNA transposons

Eukaryotic DNA transposons are rather simply structured TEs that consist of a transposase-encoding gene framed by inverted terminal repeat sequences (ITRs). The transposase mediates the excision of the transposable element from one chromosomal locus and integration of the mobilized element at a new site. DNA transposons are widespread in eukaryotes and can cause a variety of traits by random insertions into their host genomes. For example, the *Drosophila melanogaster P* elements are responsible for causing a syndrome of traits collectively known as hybrid dysgenesis [16]. The syndrome is usually only observed in the progeny of males with active *P* elements and females lacking *P* elements. The dysgenic traits can be explained by an initial burst of transposition of *P* elements that causes disastrous genomic rearrangements in developing germ cells. Typically the transposition activity of DNA transposons rapidly declines shortly after invasion of a new host and an initial burst of transposon amplification. This is because the transposase does not discriminate intact from defective elements, which successively leads to the accumulation of defective transposons. The nematode *Caenorhabditis elegans* contains the Tc1 transposon that is similar to the mariner element discovered in *Drosophila mauritiana* (reviewed in [17]). It seems that members of the Tc1/mariner superfamily are found in all animals. Whereas transposition of *P* elements is restricted to their natural host, *Drosophila*, Tc1/mariner elements are active in many eukaryotic species and may integrate any recombinant piece of DNA into a target genome independent of additional host factors, provided that the gene of interest is flanked by ITRs and the transposase is expressed in trans in the same cell [17, 18]. One particular Tc1/mariner element, Sleeping Beauty, has recently been used to introduce a marker gene into the mouse germ line [19].

LTR retrotransposons

The LTR retrotransposons encode a gag gene whose product is a nucleocapsid, RNA-binding protein. Enzymatic activities required for retrotransposition are encoded by the pol gene and include RT, RNase H, integrase (IN) and protease (PRO) [20]. The gag and pol genes are usually encoded in separate reading frames that are translated into single proteins by ribosomal frameshifting. This ensures high-level production of the GAG protein and low abundance of enzymes and regulatory proteins required for retrotransposition. The fusion proteins are processed by the encoded protease [21]. LTR retrotransposons can be divided into six clades based upon phylogenetic analysis of conserved RT/RNAse H domains [22]. One of these clades represents the vertebrate retroviruses, which can be regarded as specialized mobile LTR retrotransposons that have acquired the ability to spread in a host cell population and eventually invade new host species (horizontal transfer). Interestingly, the capability of infecting new cells is linked to the acquisition of only one extra gene called envelope (env). Typical

env genes encode transmembrane proteins that are inserted into the outer membranes of the retrovirus particles and function as host receptor-binding proteins. Most env genes are highly diverged and hard to trace back to their origins. It has been noted that LTR retrotransposons of various different phylogenetic clades and many different organisms may eventually pick up env genes of viral origin, thus mimicking the potential of these viruses to bind to and infect cells [22].

Retroviruses and LTR retrotransposons share a common retrotransposition mechanism [23, 24]. Reverse transcription of viral RNA is primed by a cellular tRNA that is also packed into the viral particles. Priming occurs downstream of the 5′ LTR, thus producing a short first reverse transcription product covering part of the 5′ LTR (strong-stop minus-strand DNA). The strong-stop minus-strand DNA is transferred to the 3′ LTR and primes the full-length synthesis of the minus DNA strand using the plus-strand RNA as template. Meanwhile, the RNase H removes parts of the RNA from the 3′ LTR region, leaving a polypurine tract as primer to start plus-strand DNA synthesis. The plus-strand strong-stop DNA again switches to the 5′ LTR and initiates the synthesis of plus-strand DNA. Hence, the end products of reverse transcription of either retroviruses or LTR retrotransposons are double-stranded DNA fragments that are inserted into the host genomes by means of the retroelement-encoded IN [21, 24]. The catalytic cores of retroviral integrases are similar to transposases and include a critical set of aspartic acid and glutamic acid residues (the $DD_{35}E$ motif) that form an active site able to break and join phosphodiester bonds to perform a transient DNA double strand break [25, 26].

Non-LTR retrotransposons

The non-LTR retrotransposons, also referred to as LINEs (long interspersed nuclear elements), contain one or two open reading frames (ORFs) that are flanked by short untranslated regions (UTRs). Active promoter elements have been detected in the 5′ UTRs of many non-LTR retrotransposons, e.g. I, jockey and F elements (all from *D. melanogaster*) [27–29], in mouse and human L1 [30–33], and in TRE5-A from *Dictyostelium discoideum* [34]. The RT of non-LTR retrotransposons is readily detectable by sequence alignment and can be used to establish phylogenetic relationships [35–37]. This has led to the hypothesis that non-LTR retrotransposons are derived from prokaryotic group II introns and are ancient to LTR retrotransposons and retroviruses [37]. In non-LTR retrotransposons with two nonoverlapping ORFs, the derived proteins are translated separately from bicistronic RNAs by ribosomal termination/reinitiation or internal ribosomal entry, as experimentally shown for the *Drosophila I* factor and the human L1 element [38, 39]. This separates the non-LTR retrotransposons from retroviral elements, which usually produce fusion proteins by ribosomal frameshifting [24], and is consistent with the apparent absence in non-LTR retrotransposons of retrovirus-like protease domains responsible for cleavage of protein precursors.

Experiments on the enzymatic properties of the *Bombyx mori* R2-encoded protein have been crucial for the development of a relatively simple model describing the retrotransposition of non-LTR retrotransposons as a coupled reverse transcription/integration reaction, the 'target-primed reverse transcription' (TPRT) [40, 41]. The endonuclease (EN) of the multifunctional R2Bm protein cuts one strand of the target DNA, thus generating a free 3′-hydroxyl group that can be used by RT to start reverse transcription of the R2Bm RNA. Once reverse transcription is in progress, the EN cleaves the opposite DNA strand. It is not yet clear how joining of the cDNA 3′ end to the genomic DNA is accomplished. A model derived from in vitro analysis of *B. mori* R2 proposes that the RT performs a template switch to copy the second target DNA strand to a certain extent, and this connects the first strand complementary DNA (cDNA) of the retrotransposon to the genomic DNA by base pairing [42]. Second-strand cDNA synthesis and gap repair most likely involves host proteins. The TPRT mechanism of retrotransposition may have its roots in common ancestors of eukaryotic non-LTR retrotransposons with prokaryotic group II introns [43]. Since the first strand cDNA of non-LTR retrotransposons is covalently attached to the target DNA upon reverse transcription, the TPRT model explains very elegantly the occurrence of 5′-truncated versions of non-LTR retroelements as being simply the products of abortive reverse transcription reactions.

TEs in *D. discoideum*

Genome organization

D. discoideum is a widely used microbial model for the study of developmentally regulated gene expression, cell differentiation and morphogenesis and has recently been included into the list of relevant model organisms for biomedical research (http://www.nih.gov/science/models/). *D. discoideum* amoebae have, depending on strain properties, a 34–40-Mb haploid genome that is organized in six chromosomes ranging from 4 to 9 Mb in size. Previous studies, performed in the 'pregenome' era, roughly estimated the gene content of the *D. discoideum* genome at 8–14,000 genes [44–46]. Today, sequencing of the 34-Mb genome of strain AX4 is nearly finished [47–52]. Hallmarks of the recently announced complete DNA sequence of the 8-Mb chromosome 2, which makes up about 25% of the genome, are the very high A+T content and unexpectedly short intergenic regions.

The averaged A+T content is 78% but may reach 98% in intergenic regions. It was calculated from chromosome 2 data that the *D. discoideum* genome contains one gene per 2.6 kb on average, and that there are 10,000–11,000 genes in total [53]. The relatively high load with interspersed movable elements (10% of the genome), often organized in large complex clusters, was a challenge to the assembly of the *D. discoideum* genome into large contigs, leaving several gaps flanked by TEs that could not be closed [52–54].

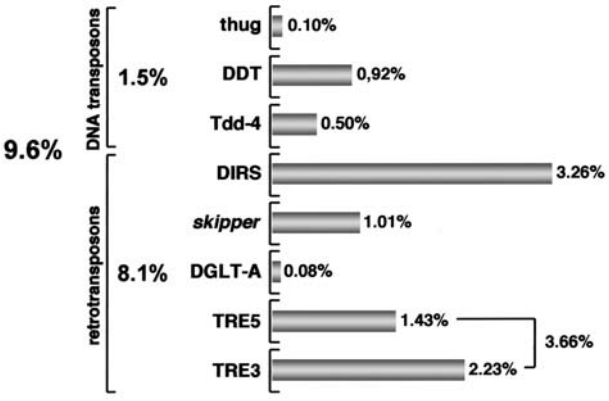

Figure 1. Complex repeat content of the *D. discoideum* genome. Data are from [54]. Refer to the text for informations on the individual TEs.

Nomenclature and abundance of *D. discoideum* TEs

Movable DNA elements in the *D. discoideum* genome were first described in the early 1980s. Since then, many new TEs have been discovered, and the assignment of element names is a bit of a mess. The first element was independently discovered by two laboratories and consequently got two names, DIRS-1-1 and Tdd-1 [55, 56]. Since additional repetitive elements were discovered in the laboratory that also identified Tdd-1, these elements were named Tdd-2 and Tdd-3 regardless of their nature [57]. Today we know that Tdd-1 is an LTR retrotransposon (see below), while Tdd-2 and Tdd-3 are non-LTR retrotransposons associated with tRNA genes [58, 59]. Studies on tRNA gene expression led to the discovery of two tRNA-gene-associated retrotransposons named DRE [60] and H3R [61]. DRE is structurally related to Tdd-2 and Tdd-3, whereas H3R has properties of a solitary LTR derived from an LTR retrotransposon [62]. Meanwhile, a putative DNA transposon was discovered and given the name Tdd-4. In the process of sequencing the entire *D. discoideum* genome new TEs were discovered, one of which is a Tdd-4-related element named Tdd-5. On the other hand, new relatives of DRE and Tdd-3 were discovered and also deserved new names.

Now that the sequencing of the *D. discoideum* genome is nearly finished, we can list all discovered complex repeat structures and think about a new nomenclature. At first glance *D. discoideum* seems to be devoid of DNA transposons. When the available sequence data of the *Dictyostelium* Genome Project are scanned for known transposases such as Tc1/mariner, hobo, Ac or bacterial IS, no hits are obtained ($P < 10^{-5}$). However, there are several repetitive sequences with characteristics of DNA transposons that do not share similarities with amino acid sequences of transposases. These elements can be grouped into the Tdd-4, DDT and thug families and comprise about 1.5% of the *D. discoideum* genome [54] (fig. 1).

Retrotransposons make up about 8.1% of the *D. discoideum* genome [54] (fig. 1). All *D. discoideum* LTR retrotransposons are derived from Ty3/*gypsy*-like ancestors as deduced from amino acid alignments of the encoded RTs (reviewed in [62]). The DIRS-1-derived sequences by far outnumber the two other LTR retrotransposons, *skipper* and DGLT-A (fig. 1).

About 3.7% of the *D. discoideum* genome is composed of non-LTR retrotransposons (fig. 1). Notably, all these elements are found in close vicinity to tRNA genes [54]. We therefore proposed to use the collective name tRNA gene-targeted retrotransposable element (TRE) [58]. Hence TRE5 denotes elements found upstream (5′) of tRNA genes, while TRE3 is used for elements found downstream of tRNA genes. After introduction of this nomenclature we became aware that H3R is indeed a solitary LTR of an unusual LTR retrotransposon that is also found upstream of tRNA genes. This element has been assembled from sequence data and a full-length element is hypothetical, but the protein encoded in this element has similarity to RT sequences of Ty3/*gypsy*-like LTR retrotransposons and was therefore renamed *Dictyostelium gypsy*-like transposon, DGLT-A [54].

Putative DNA transposons in the *D. discoideum* genome

Tdd-4

The putative DNA transposon Tdd-4 is unique since it encodes a gene whose 707 amino acid protein product shares significant homology in an ~130-amino acid motif to retroviral integrase (IN) core domains [63]. Yet no RT sequence can be identified in the Tdd-4-encoded protein. A strong argument that Tdd-4 is indeed a DNA transposon and not a retroelement comes from the fact that the ‘transposase’ gene is interrupted by six introns (fig. 2A). No other proteins are encoded in Tdd-4, except for a slightly shorter IN protein derivative produced from the same gene by alternative splicing (fig. 2A). BLAST searches of the GenBank/SwissProt databases do not highlight *D. discoideum* LTR retroelements as the closest relatives of Tdd-4 IN, which may argue for horizontal transfer of a Ty3/*gypsy*-like IN domain into the *D. dis-*

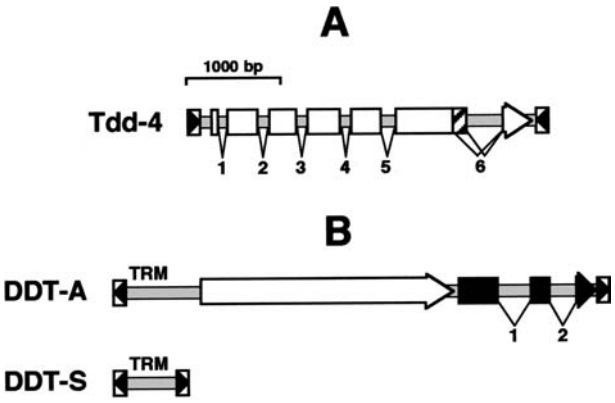

Figure 2. Structures of putative *D. discoideum* DNA transposons. (*A*) The Tdd-4 element encodes a single ORF interrupted by six introns (indicated by numbers). Alternative splicing produces two alternative protein products. Black triangles denote the orientation of the ITR sequences. (*B*) Structure of the autonomous DDT-A element and its nonautonomous DDT-S derivative. DDT-A encodes two proteins indicated by the white and black arrow, respectively. The latter is interrupted by two introns (indicated by numbers). TRM, tandemly repeated motifs, refer to [54] for details.

coideum genome. Tdd-4 has 145-bp nearly perfect ITRs. The element starts with 5′-TG… and ends with …CA-3′, a motif typically found in transposon ITRs. However, the 5′-TG…CA-3′ sequence also terminates the LTRs of retrotransposons and is essential for retrotransposition [24]. Hence Tdd-4 may represent an unusual, intermediate structural concept of DNA transposons and retroelements in which a retroviral IN-like core domain of an ancient LTR retrotransposon was 'trapped' by inverted repeats, resulting in a TE that moves by a cut-and-paste mechanism rather than by retrotransposition. Genomic Tdd-4 copies are flanked by 5-bp TSD sequences. About 40 Tdd-4 elements are estimated to reside in the haploid *D. discoideum* genome [54].

Tdd-5

Tdd-5 was discovered by the *Dictyostelium* Genome Sequencing Consortium [54]. Only about five Tdd-5 elements may be present in the *D. discoideum* genome. The element could not be completely reconstructed from the current data set, suggesting that functional elements may not exist in the modern *D. discoideum* genome. It seems likely that Tdd-4 and Tdd-5 diverged from a common ancestor after invading the *D. discoideum* genome, since fragmented Tdd-5 ORF sequences show significant similarity to Tdd-4 [54].

DDT elements

DDT elements represent a novel class of putative DNA transposons whose encoded proteins are neither similar to known transposases nor to *D. discoideum* Tdd-4 or other known protein sequences. DDT-A and DDT-B are 5168 and 5521 bp in length, respectively, and have short ITRs that meet the consensus 5′-AC…TG-3′ (fig. 2B). The first gene has coding capacity for 813 (DDT-A) and 815 (DDT-B) amino acids, respectively. The second ORF is interrupted by two introns. The spliced transcripts code for proteins with 264 (DTT-A) and 256 (DDT-B) amino acids that have no known orthologues in other organisms. ORF1 terminates with a putative polyadenylation site, and ORF2 starts with new ATG translation initiation codon, suggesting that both genes are translated separately from bicistronic transcripts. DDT-S is an 857-bp nonautonomous derivative of either DDT-A or DDT-B that lacks both ORFs (fig. 2B). It was estimated from the data set of the *Dictyostelium* genome project that about 20 copies of both DDT-A and DDT-B and ~130 DDT-S copies exist in the *D. discoideum* genome [54]. DDT insertions are flanked by 2-bp TSDs indicative of true transposition events.

Thug elements

The thug elements are a poorly defined group of low-copy (≤20) repetitive elements [54]. The two identified thug elements are 2192 bp (thug-S) and 1132 bp (thug-T) in length and are highly A+T rich. They are flanked by 5′-TG…CA-3′ terminal repeat sequences. Thug elements do not have any coding capacity. Although the ITR structures of thug-S and thug-T are unrelated to one another or to other TEs in the *D. discoideum* genome, it may be possible that this class of elements is trans-mobilized by TE-derived enzymes recognizing the ITR sequences of the thug elements.

LTR retrotransposons in the *D. discoideum* genome

DIRS-1

DIRS-1 was the first mobile element to be discovered in the *D. discoideum* genome. DIRS-1 is unique in that it has LTR-like structures arranged as inverted LTRs (fig. 3A). The element encodes an RT/RNase H domain in a separate ORF that is distantly related to corresponding domains of Ty3/*gypsy*-like LTR retrotransposons [22]. The *D. discoideum* genome contains some 200 DIRS-1-related sequences, most of which are derived from element-internal recombinations that resulted in defective remnants [54]. The DIRS-1 element is expressed at high levels in growing *D. discoideum* cells, both in sense and antisense orientation. There is experimental evidence that double-stranded RNA retrotransposition intermediates occur, but the significance of these RNA species for DIRS-1 retrotransposition and evolution is unclear [J. Oberstrass, personal communication].

Skipper

The *skipper* element is a bona fide Ty3/*gypsy*-like LTR retrotransposon with some unique features. It is strongly expressed in growing *D. discoideum* cells, which allowed

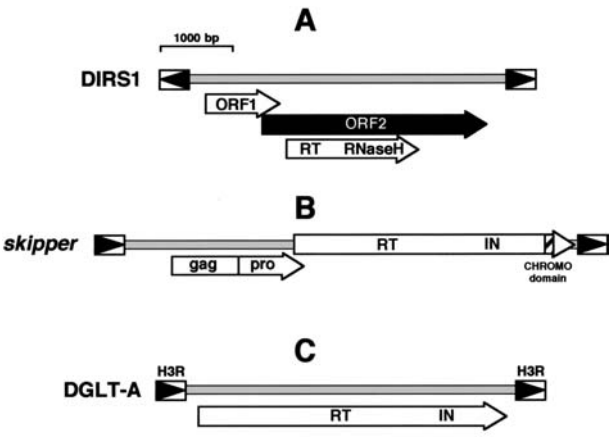

Figure 3. Structures of *D. discoideum* LTR retrotransposons. Black triangles denote the orientations of LTR sequences. Refer to the text for details. (*A*) The DIRS-1 element contains three ORFs and the RT/RNaseH separated from ORF2, whose function is unknown. (*B*) Note that the *skipper* element encodes the GAG and PRO proteins in a single ORF, but separated by a single stop codon. (*C*) DGLT-A contains a single ORF in which only RT and IN motifs can be detected by BLAST searches. Note that single LTR sequences in the *D. discoideum* genome are referred to as H3R.

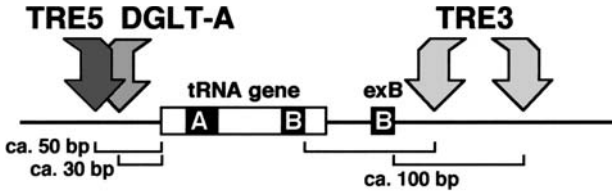

Figure 4. tRNA genes are the landmarks for integration of *D. discoideum* TEs. As in other eukaryotes, *D. discoideum* tRNA genes contain internal promoter elements denoted by A box and B box. In addition, most *D. discoideum* tRNA genes show external B-box sequences ~30 bp downstream, identified by the consensus GTTCRANNC (in which R is adenine or guanine and N stands for all four nucleotides). tRNA genes are targeted by one LTR retrotransposon (DGLT-A) ~30 bp upstream of the tRNA coding region, and three non-LTR retrotransposons ~50 bp upstream (the TRE5s) and four non-LTR retrotransposons ~100 bp downstream (the TRE3s). Note that TRE3s have may use either a tRNA gene-internal B box or an external B box as integration target.

the isolation of a nearly complete 6-kb cDNA [64]. The *skipper* element has a somewhat peculiar arrangement of genes encoding functional proteins such as GAG, PRO, RT/RNase H and IN (fig. 3B). There are two ORFs arranged in different reading frames as in other LTR retrotransposons. However, the pro gene is fused to the gag gene but is separated from it by a single translation stop codon. The presence of a nearby pseudoknot structure suggests that a GAG/PRO precursor is produced by ribosomal stop codon suppression and is subsequently cleaved into functional GAG and PRO proteins [64]. At the carboxy-terminal end of the IN domain is a CHROMO (chromatin organization modifier) domain detectable that is also present in many other members of the Ty3/*gypsy*-like LTR retrotransposons. The CHROMO domain is thought to play a role in recognition of chromosomal integration sites [65]. *Skipper* seems to use an uncommon way to prime minus-strand DNA synthesis, which is usually achieved by binding of a cellular tRNA to a primer binding site immediately downstream of the left LTR. This tRNA primer binding site is missing in *skipper* [64]. Although *skipper* is strongly expressed, only about 50 full-length and 10 solitary LTR elements have been estimated to reside in the *D. discoideum* genome [54].

DGLT-A

During the analysis of tRNA gene upstream regions to identify TRE5 elements, we noticed about 10 years ago a short 260-bp repetitive and highly A+T-rich sequence that was flanked by canonical 5′-TG…CA-3′ LTR consensus sequences. Since this DNA fragment contained a Hind III restriction site, it was named H3R [61]. We later speculated that H3R may represent a solitary LTR of a not yet identified larger LTR retrotransposon [62]. This has recently been confirmed by the *Dictyostelium* genome project. Indeed, a hypothetical 5051-bp element could be reconstructed from the sequence data set that contains a single ORF with coding capacity for a protein with 1437 amino acids (fig. 3C). Within this protein a Ty3/*gypsy*-like RT core domain and an IN core domain can be identified in BLAST searches. In the full-length consensus element, named DGLT-A, the ORF is flanked by H3R sequences arranged as direct repeats (fig. 3C). About five nearly complete DGLT-A sequences and 15 H3R solitary LTR sequences are thought to be present in the *D. discoideum* genome [54].

Non-LTR retrotransposons in the *D. discoideum* genome: the TREs

It is intriguing that all non-LTR retroelements identified in the *D. discoideum* genome are clustered near tRNA genes. The TRE family comprises seven members that can be further divided into elements integrating either ~50 bp upstream (TRE5) or ~100 bp downstream (TRE3) of tRNA genes (fig. 4).

The TRE5 group

There are three TRE5 elements detectable in the *D. discoideum* genome, the best studied of which is TRE5-A. This element was first identified in the late 1980s and named DRE [60]. A full-length TRE5-A.1 spans 5657 bp of genomic DNA. It encodes two ORFs arranged in separate reading frames [66] that are flanked by modular UTRs (fig. 5). ORF1 encodes a 51-kDa protein that has no similarities to other proteins. It has been observed that recombinant ORF1 protein forms large complexes in the cytosol of *D. discoideum* cells [67], but RNA binding in

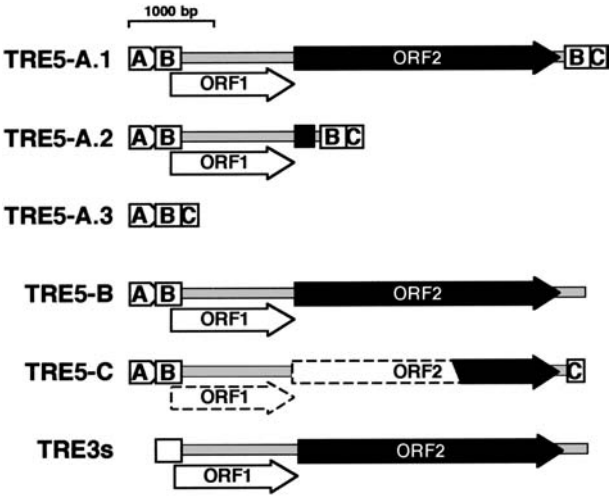

Figure 5. Structures of the *D. discoideum* TREs. Details are outlined in the text.

this complex formation is speculative. In analogy to what is known about ORF1 proteins in other non-LTR retrotransposons [6], the TRE5-A ORF1 may bind to TRE5-A transcripts to form ribonucleoparticles that are somehow carried to the genomic DNA for reverse transcription and integration. Endogenous ORF1 protein can be detected by Western blotting of cell extract proteins of *D. discoideum* cells treated with antimycin A to increase the TRE5-A-derived transcript level [67]. TRE5-A ORF2 protein is not detectable in growing *D. discoideum* cells by biochemical/immunological means, but it must be present at catalytic amounts because the TRE5-A elements are actively retrotransposing in these cells (discussed below). ORF2 protein contains three distinct domains that share homologies with RT, EN and cysteine-rich domains [59], similar to corresponding sequences in other non-LTR retrotransposons [6, 37, 68, 69]. No unequivocal nuclear import signal sequences can be identified in either TRE5-A ORF1 or ORF2 protein. So how do these proteins enter the nucleus for reverse transcription and integration? It must kept in mind that TRE5-A elements reside in continuously dividing cells. Retrotransposon-derived protein/RNA complexes may approach the genomic DNA while the nuclear envelope is broken down during mitosis, and therefore no obvious evolutionary constraint to invent nuclear transportation strategies may exist when residing in a rapidly growing microorganism such as *D. discoideum*.

The A module is a 200-bp sequence at the 5′ end of TRE5-A (fig. 5) that functions as an element-internal promoter, which means that plus-strand (sense) RNA synthesis is initiated at the first nucleotide of the TRE5-A element [34]. Genome sequencing has uncovered TRE5-A elements with up to seven consecutive A modules. In this respect TRE5-A resembles mouse L1 elements, which also contain short multimer blocks in their 5′ UTRs [6, 32]. In the mouse 5′ UTR a single monomer is insufficient to support expression of reporter genes, but multimerization increased reporter gene activity proportional to the number of monomers [32]. In contrast, a solitary A module of the *D. discoideum* TRE5-A element acts as a strong promoter [34]. It is not known whether several consecutive A modules would even increase this promoter activity. In the human L1 element the 5′ UTR monomers contain a 5′-GCCATCTT-3′ binding site for the Yin Yang-1 (YY-1) transcription factor [32] that is important for efficient L1 transcription [70, 71]. This binding site is conserved in many other retrotransposons as well [32]. A putative YY-1 binding site is also present in the A module (nucleotide positions 111–118 of TRE5-A). However, it is not known whether a YY-1-like protein with the same sequence recognition specificity as in mammals is present in the *D. discoideum* genome. The putative YY-1 binding site is mutated to 5′-GCCATTTT-3′ in the A module of TRE5-C and is apparently absent in TRE5-B, thus questioning a role of this particular DNA sequence in the regulation of A module promoter activity in TRE5 elements.

Downstream of the A module we defined a B module, a 290-bp sequence that is duplicated and moved to the 3′ end of the TRE5-A element (fig. 5). The function of the B module for TRE5-A mobilization is unclear; however, it contains the ATG translation codon that is used to express ORF1 protein. As discussed below, it is tempting to speculate that a B-module-related sequence may represent an ancient promoter sequence whose function was no longer required after acquisition of the A module, a new strong promoter, into the TRE5 elements.

The C module is a unique 134-bp sequence located at the very 3′ end of TRE5-A (fig. 5). It is a weak promoter that is responsible for the expression of TRE5-A-derived minus-strand RNAs whose transcription is initiated within the 3′ terminal oligo(A) tract of the element [34].

TRE5-A-derived sense and antisense RNA levels are drastically increased in cells grown in the presence of the respiratory chain blocker antimycin A [34]. Antimycin seems to specifically affect the stability of TRE5-A RNAs rather than transcription rates of the retroelement promoters. Minus-strand transcripts may play a role in the regulation of the plus-strand RNAs that are critical to express the retroelement-encoded proteins as a prerequisite of retrotransposition. In fact, antisense RNA-regulated control of TRE5-A expression would be an attractive way to control TRE5-A retrotransposition rates by host-encoded transcription regulators. Such a host factor may be CMBF, a protein that binds to the C-module of TRE5-A in vitro [72, 73]. CMBF is an essential host protein involved in several aspects of gene regulation during growth and development of *D. discoideum* cells [74]. Mutant *D. discoideum* cells severely underexpressing CMBF

[74] show diminished TRE5-A sense and antisense RNA expression that can be restored by expression of a CMBF-encoding cDNA in the mutant cells [T. Winckler, unpublished data from our laboratory].

TRE5-A shows complex structural variations. Quite common is a large deletion of ORF2 from the TRE5-A.1 element, which along with several other minor mutations makes up a lineage of TRE5-A derivatives referred to as TRE5-A.2 (fig. 5). Intraelement recombination among the B modules may lead to the deletion of both ORFs to produce 'ABC' minielements known as TRE5-A.3 (fig. 5). The majority of TRE5-As, as well as TRE5-B and TRE5-C elements, become severely 5′-truncated upon retrotransposition. These elements are nonfunctional since they lack A modules as internal promoters. There is also evidence for interelement recombination events. For example, we noticed a clone derived from the *Dictyostelium* genome project (JC2d24f04) containing a hybrid ABC element in which the A module is derived from TRE5-C, while the BC modules are derived from TRE5-A.1 [T. Winckler, unpublished observation].

Inspection of genomic TRE5-A insertions has revealed a remarkable integration preference. We recently analyzed 35 genomic TRE5-A copies and found that they inserted at 44–54 bp distance upstream of the associated tRNA genes in an orientation-specific manner. In fact, 33 of 35 analyzed elements are positioned even more precisely at 48 ± 2 bp 5′ of the tRNA gene coding regions [61, 75]. All bona fide TRE5-A insertions are flanked by target site duplications (TSDs) of 12–16-bp length [61, 76, 77]. Since all tRNA genes in the *D. discoideum* genome have unique flanking DNA sequences, it is not likely that conserved genomic sequences upstream of tRNA genes determine the sites for TRE5-A integration. We prefer the hypothesis that the strict position specificity of integration achieved by mobile TRE5-A elements depends on direct protein interactions of TRE5-A-encoded proteins with the RNA polymerase III (pol III) transcription apparatus, namely transcription factor TFIIIB. This factor is thought to occupy the space at the 5′ end and upstream regions of transcribed tRNA genes, thereby possibly relaxing the local chromatin structure in the vicinity of the tRNA genes in a way that allows TRE5-A integration complexes to access to the DNA immediately upstream of the TFIIIB binding sites.

TRE5-B and TRE5-C have been less successful than TRE5-A in colonizing the *D. discoideum* genome. Only about 15 TRE5-B and 5 TRE5-C copies are estimated to reside in the *D. discoideum* genome [54]. TRE5-B has an AB-module structure similar to TRE5-A in its 5′ UTR, but lacks both the B-module duplication and the C module of TRE5-A (fig. 5). However, close inspection of AB-module sequences showing up in the *Dictyostelium* genome databases revealed an AB-module variant (sequence IIAFP1D14094) that neither belonged to TRE5-A nor to TRE5-B. The newly identified AB-module sequence shows a phylogenetic relationship to TRE5-A and TRE5-B very similar to the TRE5-C element when comparing the highly conserved ORF2 sequences. It is therefore plausible to presume that the new AB-module sequence represents the missing 5′ end of TRE5-C. Yet no complete TRE5-C element could be reconstructed. All genomic TRE5-B and TRE5-C insertions produced variable length TSDs similar to those of TRE5-A. The distances of TRE5-B and TRE5-C elements to the corresponding tRNA genes have been analyzed based on chromosone 2 sequence data. Of the four TRE5-B elements found on chromosome 2, three copies were inserted at 45–50 bp upstream of a tRNA gene, while one copy showed a unique 96-bp distance to the first nucleotide of the associated tRNA gene. Five TRE5-C copies were detected on chromosome 2, four of which were associated with a tRNA gene at 44–46 bp distance. One TRE5-C copy was moved away from its tRNA gene target by insertion of a H3R LTR derived from a mobile DGLT-A.

TRE5-A elements are continuously expanding in the modern *D. discoideum* genome

Retrotransposition events are rare. For example, although ~45% of the modern human genome is derived from mobile elements [78], 'a principle investigator simply cannot count on seeing a natural de novo retrotransposition within the standard five-year period of a typical research grant' [15], suggesting that almost all TEs in our genome have gone extinct (with the exception of a certain subgroup of L1 elements that can cause heritable diseases, see [6]). Are *D. discoideum* TREs expanding in the modern *D. discoideum* genome? Unfortunately, the TREs do not insert into structural genes whose disruption could lead to phenotypic manifestation of resulting mutations. To make TREs mutagenic, i.e. to coax them to insert into a structural gene whose disruption can be selected for, we designed a 'TRE trap' that consisted of a bait tRNA gene positioned within an intron in the *D. discoideum* UMP synthase gene [75]. Mutant cells lacking a functional UMP synthase gene acquire resistance to the cytostatic drug 5-fluoroorotic acid, thus allowing us to positively select for mutants that experienced fresh UMP synthase gene disruption events caused by TRE retrotransposition [75]. TRE insertion mutations within the artificial UMP synthase gene could be picked up easily: about 100 clones were isolated in one selection procedure starting with 10^7 cells [75]. We could show that the presence of the bait tRNA gene within the UMP synthase reporter gene increased the mutation rate about 50-fold when compared with a tRNA gene-less UMP synthase gene derivative, and 96% of all isolated clones contained TRE insertions [75]. Interestingly only TRE5-A copies were picked up in our screenings, suggesting that TREs other

than TRE5-A expand at very low rates, if any. About half of the newly inserted TRE5-A copies were derived from nonautonomous TRE5-A.2 elements, suggesting that they are able to amplify by trans-mobilization, i. e. by borrowing the ORF2-encoded enzymes from full-length TRE5-A.1 elements [75]. Both TRE5-A subtypes are highly prone to 5′ deletion upon retrotransposition, thus generating defective element copies at high rate [75]. Phylogenetic analysis using ORF1 sequences suggest that the TRE5-A.2 lineage was generated only very recently in the evolution of the TREs in the *D. discoideum* genome [unpublished observation]. Thus, rapid amplification of TRE5-A.2 to copy numbers similar to their full-length TRE5-A.1 parents [76] may be explained by a selection advantage of TRE5-A.2 due to the reduced element size (2.6 kb compared with 5.7 kb), greatly enhancing the probability of TRE5-A.2 elements to escape abortive reverse transcription and to generate actively transcribed progeny with intact 5′ ends [75].

The TRE3 group

The four TRE3 elements show a comparable structural organization and less variability than the TRE5s (fig. 5). On chromosome 2, three full-length TRE3-A elements with intact ORFs are present. One full-length copy of each TRE3-B and TRE3-C are also found on chromsome 2, but these elements show multiple mutations, rendering the elements nonfunctional. It must be noted, however, that many more full-length TREs are expected to reside on chromosome 2, since several gaps remained after assembly of the available data set due to the presence of almost identical TRE copies whose sizes exeeded the insert sizes of the gene libraries. Although the ORFs of all four TRE3 elements are quite conserved, their 5′ UTR and 3′ UTR are highly diverged. TRE3-A contains a functional promoter present within the first 350 bp of the element [T. Winckler, unpublished data]. About 150 TRE3 copies are present in the *D. discoideum* genome, 70% of which belong to TRE3-A and TRE3-B [54].

In previous studies it was noted that neither TRE3-A nor TRE3-B produce TSDs when integrating at new genomic locations [58, 59]. This conclusion was obviously hampered by the limited number of elements analyzed. Inspection of 35 TRE3 insertions on chromosome 2 showed that both TRE3-A and TRE3-B may or may not produce TSDs: in only about half of the bona fide integration events we found TSDs ranging from 8 to 12 bp in length [T. Winckler, unpublished analysis]. By contrast, all TRE3-C and TRE3-D copies on chromosome 2 being accessible to this analysis were flanked by TSDs of highly variable lengths (15–19 bp and 14–40 bp, respectively).

Most (but not all) TRE3s on chromosome 2 are associated with tRNA genes. As noted in previous studies [58, 59], the positions of inserted genomic TRE3s relative to their target tRNA genes vary considerably. It appears that a tRNA gene-internal B-block promoter sequence is the primary target for integration. The B block is the binding site for the RNA polymerase III transcription factor TFIIIC. In the *D. discoideum* genome about 80% of the tRNA genes are flanked by so-called extra B blocks located immediately downstream of the tRNA gene-coding regions [79] (fig. 4). These extra B blocks are recognized by TFIIIC in vitro [80], suggesting that in vivo the TRE3-specific integration proteins may interact with TFIIIC molecules bound to either the gene-internal B blocks or the external B blocks. Hence TRE3 insertions can occur at two alternative positions ~40–150 bp downstream of the respective tRNA genes ([58, 59]; see fig. 4).

The TRE family: common origin but different position preferences

The origin of the TREs remains a mystery. BLAST searches using the RT domains of the TRE3 and TRE5 lineage reveals that some vertebrate retrotransposons, namely the teleost fish element Swimmer-1 and the human L1 elements, are the most similar to the TREs. There are no TRE sequences detectable in *Dictyostelium* species closely related to *D. discoideum* [T. Winckler, unpublished]. As put forward by T. Eickbush, non-LTR retrotransposons are thought to be excluded from horizontal transfer among species and can thus be spread within a species only by germline transmission [22, 37]. Since *Dictyostelium* has a sexual life cycle that occasionally leads to the fusion of compatible strains followed by meiosis [81], we may hypothesize that TREs invaded the *D. discoideum* genome by horizontal transfer from a yet unidentified *Dictyostelium* carrying a TRE ancestor. Structural and phylogenetic analysis of the proteins encoded by the seven members of the TRE family strongly argues that they evolved from a common ancestor whose integration preference remains elusive.

First, both the ORF1 and ORF2 proteins of the TREs have retained a significant degree of similarity. Phylogenetic trees using a partial ORF2 amino acid sequences (fig. 6) or combined ORF1/ORF2 datasets strongly support a common origin of the TREs, which diverged into two different clades representing the family members with integration preferences upstream and downstream of tRNA genes, respectively. Second, TREs share a common structural organization. Full-length elements are nearly 6 kb long and encode two genes organized in separate, but partially overlapping ORFs (fig. 5). In TRE3-A, TRE3-B and TRE3-C the translation stop codon of ORF1 overlaps with a translation initiation codon of ORF2 [58]. In TRE3-D the ORF2 translation start is located 7 bp upstream of ORF1 termination. This suggests that TRE3 RNAs are translated by a mechanism proposed by Thomas and Capecchi [82]: ORF1 translation is termi-

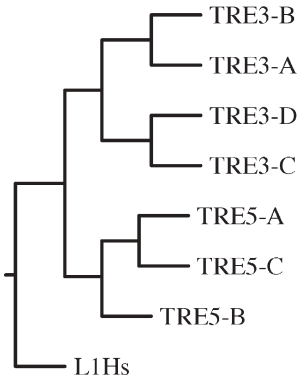

Figure 6. Molecular phylogeny of the *D. discoideum* TREs. Shown is a maximum likelihood analysis using the ProtML program of the Phylip software package (http://bioweb.pasteur.fr/seqanal/interfaces/prot_nucml.html). Approximately 300 amino acid positions from the carboxy-ternimal one third of ORF2 encoding the cysteine-rich domain were used for tree calculation. The tree was rooted with the corresponding sequence of human L1, which clusters in the same clade as TRE5-A [37]. The tree topology shown in this figure matches a previously performed parsimony tree [54]. This topology is also strongly supported by neighbor-joining trees calculated from CLUSTAL-X alignments of combined ORF1/ORF sequences ranging from 1438 to 1677 amino acid in total length, but omitting TRE5-C, whose full-length sequence is unknown (not shown).

nated when the ribosomes reach the ORF1 stop codon, then ribosomes scan back and forth several nucleotides, and initiate ORF2 translation if an ATG codon is located nearby. The typical TRE3-like arrangement of the ORF1/ORF2 overlaps was modified in the TRE5 lineage: in TRE5-B, a 109-bp polyadenine stretch has been inserted just after the translation termination codon of ORF1. This polyadenine stretch is immediately followed by a translation initiation codon for ORF2. This argues that TRE5-B transcripts are bicistronic like the TRE3 RNAs but that ORF2 is translated by internal ribosomal entry. In TRE5-A the situation is mysterious, since no ATG codon is present within the first 300 bp of ORF2, and ORF1 and ORF2 overlap by some 70 bp with no homoadenine stretch separating them. The downstream ATG codon in ORF2 is unlikely to be used to initiate ORF2 translation since this would eliminate about one third of the encoded EN domain. A short stem loop structure may be formed in TRE5-A transcripts ~30 bp upstream of the ORF1 translation stop codon, and it has been shown experimentally by expession of artificial *lacZ* reporter constructs that this stem loop supports a –1 ribosomal frameshifting [83]. This suggests that TRE5-A may use a multifunctional ORF1/2 fusion protein to catalyse its retrotransposition. It is unclear how and why this apparent ribosomal frameshifting in TRE5-A translation evolved. Unfortunately, the full-length sequence of TRE5-C, the closest relative to TRE5-A [54], still awaits its discovery and represents a missing link that could add critical further information to solve this mystery.

Third, inspection of the 3′ UTR sequences reveals both similarities and differences between TRE3 and TRE5 elements. TRE3-C and TRE3-D contain 30–45-bp homoadenine stretches immediately downstream of their ORF2 translation termination codons and have otherwise completely unrelated downstream 3′ UTR sequences. The 3′ UTRs of TRE3-A and TRE3-B lack these homoadenine insertions and are otherwise unrelated to each other or the 3′ UTRs of the other TREs. Importantly, however, the arrangement of homoadenine stretches downstream of ORF2 is also found in all TRE5 elements, which have again profoundly different 3′ UTRs downstream of the poly(A) tracts. Hence the TREs have repeatedly adopted new 3′ UTRs by taking along flanking genomic sequences upon retrotransposition. Shuffling of 3′-flanking sequences is also evident in human L1 retrotransposition and is called L1-mediated 3′ transduction [84]. Hence, we prefer to use the term TRE-mediated 3′ transduction to describe the shuffling of genomic *D. discoideum* sequences by mobile TREs. It is worth emphasizing that, unlike in human L1 evolution, the shuffled *D. discoideum* sequences became integral parts of the individual TRE lineages that were conserved and retrotransposed along with the retroelement in multiple independent retrotransposition events.

From the above-listed arguments it appears that despite the similarities in the overall structural organization of the TREs, there may have some kind of integration preference switch which occurred when we assume that the unidentified ancestor of the TREs had only one of the two position preferences upstream or downstream of tRNA genes. On the other hand, the two target preferences may have evolved independent of a nonspecific ancestor. Can we pin down certain changes in the overall structural organization of the individual ancient TREs to account for this profound switch of target preference? We could speculate that a major change in the structures of the TRE5 elements compared with TRE3s was the acquisition of the Amodule, and therefore that the A module may somehow contribute structural information for the insertion upstream of tRNA genes. However, it is hard to imagine how the 5′ end of a non-LTR retroelement would participate in retrotransposition by the TPRT mechanism that is thought to depend solely on certain 3′ end sequences of the mobile non-LTR retrotransposons. In the future we can address this question experimentally by testing the mobility of artificial, nonautonomous TRE5-A elements in the TRE retrotransposition assay mentioned above [75].

tRNA gene-targeted integration of mobile elements: a successful evolutionary strategy to colonize small genomes

When we look over the genomes of various organisms, it appears that the load of mobile elements increases with

the overall genome size [85]. This may be interpreted as follows: that in small genomes there is a high risk for active TEs to hit essential host genes, thus dooming themselves and their hosts to extinction. As a consequence, only few and highly position-specific TEs may be tolerated in such organisms. In other words, TEs inhabiting compact genomes must to some extent develop strategies to integrate away from genes. These include (i) integration of mobilized TEs into preexisting chromosomal TE copies and (ii) targeting of certain locations that do not contain coding sequences. The mechanisms mobile TEs use to target other preexisting TE loci may involve both true integration events or recombination. Evidence that homology-guided gene conversion is used for the insertion of mobile LTR retrotransposons comes, for example, from experiments with yeast Ty and Tf2 elements [86–88]. It seems plausible that homology-guided targeting can only be applied to TEs that generate free double-stranded DNA transposition intermediates, as do the DNA transposons and LTR retrotransposons. By contrast, non-LTR retrotransposons may be excluded from homology-mediated integration/recombination since no free DNA intermediates are produced during the target-primed reverse transcription process. As a consequence, non-LTR retroelements may have problems settling in compact genomes unless they invent strategies to integrate at positions not occupied by functional genes.

Targeting to certain chromosomal loci other than preexisting TE clusters may be achieved by specific recognition of DNA sequences or by interaction with DNA-bound proteins. Specific targeting of chromosomal loci via protein-protein interactions is best exemplified by analysis of the yeast genome, which is highly packed with genes and tolerates only five distinct LTR retrotransposons (designated Ty1–Ty5). The distribution of Ty elements in the yeast genome is clearly nonrandom: more than 80% of Ty1, Ty2, Ty3 and Ty4 elements were found to be scattered within a 750-bp distance to tRNA genes or other genes transcribed by pol III [89]. Ty1 exhibits regional specificity to integrate near pol III-transcribed genes. This specificity may be mediated by interaction of Ty1 integration complexes with pol III transcription factors and/or other chromatin components [85, 90–92]. Ty3 has a strong preference for integrating 2 bp upstream of tRNA genes. Target site selection of mobile Ty3 elements depends in vitro on direct interaction of the Ty3 preintegration complex with transcription factors bound to RNA polymerase III-transcribed genes [93–95]. Ty5 has a strong bias for integration sites in silent chromatin and telomeric regions [96]. This integration preference is accomplished by direct interaction of Ty5 proteins with Sir4p bound to silent chromatin fractions [97].

The international consortium employed with the sequencing of the *D. discoideum* genome recently announced the completion of the 8-Mb chromosome 2 sequence, representing about 25% of the entire genome [53]. This allows a close-up look at the genomic distribution of *D. discoideum* TEs based on a significant portion of the *D. discoideum* genome. It is evident that both homology-guided integration and position-specific integration near tRNA genes have been positively selected as strategies to sustain an appreciable TE content in the compact *D. discoideum* genome. LTR retrotransposons with no particular target site selection, namely DIRS-1 and *skipper*, have been largely excluded from colonizing the central part of chromosome 2. Instead, they were forced to integrate into loci of complex TE clusters at the ends of the chromosome to prevent interference with normal chromosome and gene functions.

In contrast to the LTR retrotransposons of the *D. discoideum* genome discussed above, the TREs have managed to scatter throughout the entire genome. The assembled 6.5-Mb linkage group covering about 86% of chromosome 2 contains 65 scattered tRNA gene loci (fig. 7). It is intriguing how precisely the distribution of TEs on chromosome 2 follows that of the tRNA genes (fig. 7). We identified 61 TE loci and 92 individual TE copies on chromosome 2, 80% of which belong to the TRE or DGLT-A families and are associated with tRNA genes. The TREs occupy 48 of the 65 tRNA gene loci found in the 6.5-Mb linkage group, and in most cases a single tRNA gene is associated with multiple TREs (fig. 7). This is also indicated by previous studies [60, 61, 98] and shows that former association of a TRE3 or TRE5 with a tRNA gene does not prevent integration of additional elements at the same tRNA gene locus. In fact we never observed integration of one TRE into another TRE, although numerous TRE::TRE tandem insertions can be observed that resulted from integration of a mobile TRE near a target tRNA gene in a tDNA::TRE situation. Thus, the position of the tRNA gene-associated proteins rather than the presence of another TRE attracts mobile TREs to their tRNA gene targets. Chromosome 2 also contains two examples of a previously integrated TRE5 being dissociated from its target tRNA gene by insertion of a solitary H3R derived from a mobile DGLT-A. Conversely, many examples exist in the *D. discoideum* genome where a H3R associated with a tRNA gene was disrupted by insertion of a TRE5 at the same tRNA gene [61]. This is explained by the different distance preferences of the elements relative to their targets (compare fig. 4). As shown in figure 8, clusters of TRE elements exist which cannot be entirely explained by consecutive insertions of single elements. Rather, deletions can occur which result in abnormal spacing between tRNA and TRE element (the A modules in fig. 8). The presence of nonunique DNA in the cluster shown may be due to TRE-mediated 3′ transduction during retrotransposition.

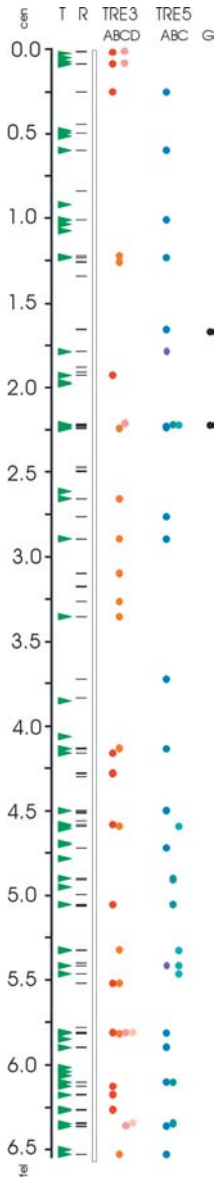

Figure 7. Distribution of tRNA genes and TREs on chromosome 2. Repetitive element loci (R) were detected by BLAST analysis using a database of *D. discoideum* complex repeats. TRNAScan-SE detected the tRNA species in the sequence (T). Naming conventions for TRE elements are as outlined in Szafranski et al. [58]. 'G' refers to DGLT-A elements.

Conclusions

Eukaryotic microorganisms have highly compacted, often haploid, genomes that demand special adaptations by molecular invaders to successfully establish and maintain active populations within their hosts. The degree to which mobile elements are tolerated in compact genomes largely depends on the genome size and degree of genome compaction. For example, the genome of the intracellular parasite *Encephalitozoon cuniculi* is so highly packed with genes (1 gene per 1 kb of genomic DNA) that it does not contain any retrotransposable elements [99]. The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, whose genomes are completely sequenced and which have gene densities comparable to that of cellular slime mold *D. discoideum*, tolerate LTR retrotransposons only at 3 and 0.4% of their total genome size, respectively [89, 100]. The challenge of colonizing such inhospitable environments has brought about several strategies by which retrotransposons try to establish and maintain active populations. tRNA genes are superior targets for integration of mobile elements because they are numerous, scattered throughout the host genomes and usually devoid of functional genes in their close neighborhoods. In the genome of *S. cerevisiae* tRNA gene-target integration was invented at least twice, since Ty1 and Ty3 obviously use different modes of tRNA gene recognition and target site selection. tRNA gene-targeted LTR retrotransposons may have independently evolved in the genomes of other yeast species, such as *Candida albicans* [101]. In *S. cerevisiae* the LTR retrotransposon Ty5 has invented yet another strategy to avoid integration into genes by interacting with a protein specifically located on silent chromatin [97]. In *S. pombe*, the only tolerated LTR retrotransposons Tf1 and Tf2 also show a strong integration bias, albeit not to tRNA genes. Tf1/Tf2 elements prefer genomic regions 100–400 bp upstream of genes, possibly by recognizing unique chromatin structures at AT-rich DNA sequences, and thereby manage to spread over the entire *S. pombe* genome [102, 103].

In *D. discoideum*, the LTR retrotransposon DGLT-A has invented tRNA gene-mediated integration independent of

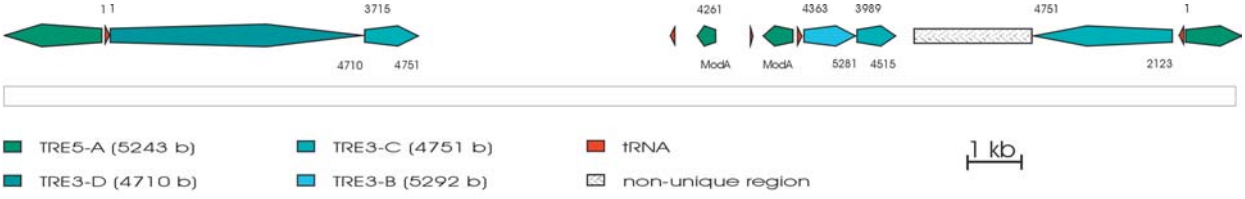

Figure 8. Scheme of a clustered tRNA gene locus associated with repetitive elements on chromosome 2. Only TRE3-D is represented as complete element. The TRE5-A elements at the borders of this locus cannot entirely be assembled due to their overall homogeneity throughout the genome. Subsequent integrations of other TRE3 elements shifted TRE3-C elements away from tRNA genes. A modules may be remnants of deletions of TRE5-A elements indicated by the dilated distances to tRNA genes. The nonunique region occurs twice in the *D. discoideum* genome.

TRE5 elements since both belong to separate groups of retrotransposons. TRE3s, on the other hand, show strong target preference downstream of tRNA genes and thus differ from TRE5. Hence tRNA gene-targeted integration was invented at least three times in *D. discoideum*. What may be the advantage of inserting mobile TE copies near tRNA genes instead of using preexisting TE clusters as landmarks for safe integration loci? Large TE clusters can easily be removed from the host genome by simple recombination events among adjacent TE copies without greatly influencing the integrity and functionality of the host genome. On the other hand, tRNA gene-associated TEs are loners in the sea of genomic DNA and seldom meet other homologous TEs that would force their elimination by recombination. In addition, elimination of large genomic portions of genomic DNA by recombination between distant TE copies may be strongly selected against due to the simultaneous loss of essential genes that ultimately causes cell death.

Acknowledgements. We are grateful to J. Oberstrass (University of Kassel, Germany) for communicating data prior to publication. The *Dictyostelium* Genome Project can be accessed on the World Wide Web at http://genome.imb-jena.de/dictyostelium or http://www.uni-koeln.de/dictyostelium/

1 McClintock B. (1948) Mutable loci in maize. Carnegie Inst. Wash. Year Book **47:** 155–169
2 Doolittle R. and Sapienza C. (1980) Selfish genes, the phenotypic paradigm and genome evolution. Nature **284:** 601–603
3 Orgel L. E. and Crick F. H. C. (1980) Selfish DNA: the ultimate parasite. Nature **284:** 604–607
4 Kazazian H. H. and Moran J. V. (1998) The impact of L1 retrotransposons on the human genome. Nat. Genet. **19:** 19–24
5 Kazazian H. H. (2000) L1 retrotransposons shape the mammalian genome. Science **289:** 1152–1153
6 Ostertag E. M. and Kazazian H. H. (2001) Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. **35:** 501–538
7 Smit A. F. A. (1999) Interspersed repeats and other momentos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. **9:** 657–663
8 Esnault C., Maestre J. and Heidmann T. (2000) Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. **24:** 363–367
9 Okada N., Hamada M., Ogiwara I. and Ohshima K. (1997) SINEs and LINEs share common 3′ sequences: a review. Gene **205:** 229–243
10 Miller W. J., McDonald J. F., Nouaud D. and Anxolabéhère D. (1999) Molecular domestication – more than a sporadic episode in evolution. Genetica **107:** 197–207
11 Agrawal A., Eastman Q. M. and Schatz D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature **394:** 744–751
12 Sharp P. A. (1991) Five easy pieces. Science **254:** 663
13 Nakamura T. M., Morin G. B., Chapman K. B., Weinrich S. L., Andrews W. H., Lingner J. et al. (1997) Telomerase catalytic subunit homologs from fission yeast and human. Science **277:** 955–959
14 Berg D. E. and Howe M. M. (1989) Mobile DNA, American Society for Mircrobiology, Washington, DC
15 Weiner A. M. (2000) Do all SINEs lead to LINEs? Nat. Genet. **24:** 332–333
16 Engels W. R. (1996) P elements in *Drosophila*. Curr. Top. Microbiol. Immunol. **204:** 103–123
17 Plasterk R., Izsvak Z. and Ivics Z. (1999) Resident aliens: the Tc1/*mariner* superfamily of transposable elements. Trends Genet. **15:** 326–332
18 Izsvak Z., Ivics Z. and Plasterk R. (2000) *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. J. Mol. Biol. **302:** 93–102
19 Dupuy A. J., Clark K., Carlson C. M., Fritz S., Davidson A. E., Markley K. M. et al. (2002) Mammalian germ-line transgenesis by transposition. Proc. Natl. Acad. Sci. USA **99:** 4495–4499
20 Doolittle R. F., Feng D. F., Johnson M. S. and McClure M. A. (1989) Origins and evolutionary relationships of retroviruses. Quart. Rev. Biol. **64:** 1–30
21 Varmus H. and Brown P. (1989). Retroviruses. In: Mobile DNA, pp. 53–108, Berg E. B. and Howe M. M. (eds), American Society for Microbiology, Washington, DC
22 Malik H., Henikoff S. and Eickbush T. H. (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. **10:** 1307–1318
23 Boeke J. and Chapman K. (1991) Retrotransposition mechanisms. Curr. Opin. Cell Biol. **3:** 502–507
24 Boeke J. and Corces V. (1989) Transcription and reverse transcription of retrotransposons. Ann. Rev. Microbiol. **43:** 403–434
25 Polard P. and Chandler M. (1995) Bacterial transposases and retroviral integrases. Mol. Microbiol. **15:** 13–23
26 Capy P., Vitalis R., Langin T., Higuet D. and Bazin C. (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common acnestor? J. Mol. Evol. **42:** 359–368
27 McLean C., Bucheton A. and Finnegan D. (1993) The 5′ untranslated region of the *I* factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression. Mol. Cell. Biol. **13:** 1042–1050
28 Mizrokhi L., Georgieva S. G. and Ilyin Y. V. (1988) *jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell **54:** 685–691
29 Minchiotti G. and Di Nocera P. (1991) Convergent transcription initiates from oppositely oriented promoters within the 5′ end regions of *Drosophila melanogaster F* elements. Mol. Cell. Biol. **11:** 5171–5180
30 Severynse D. M., Hutchison C. A. and Edgell M. H. (1992) Identification of transcriptional regulatory activity within the 5′ A-type monomer sequence of the mouse LINE-1 retrotransposon. Mamm. Genome **2:** 41–50
31 Adey N. B., Tollefsbol T. O., Sparks A. B., Edgell M. H. and Hutchison C. A. (1994) Molecular resurrection of an extinct ancestral promoter in mouse L1. Proc. Natl. Acad. Sci. USA **91:** 1569–1573
32 DeBerardinis R. J. and Kazazian H. H. (1999) Analysis of the promoter from an expanding mouse retrotransposon subfamily. Genomics **56:** 317–323
33 Swergold G. (1990) Identification, charaterization, and cell specificity of a human LINE-1 promoter. Mol. Cel. Biol. **10:** 6718–6729
34 Schumann G., Zündorf I., Hofmann J., Marschalek R. and Dingermann T. (1994) Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the *Dictyostelium* Repetitive Element, from *Dictyostelium discoideum*. Mol. Cell. Biol. **14:** 3074–3084
35 Xiong Y. and Eickbush T. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:** 3353–3362
36 Eickbush T. H. (1992) Transposing without ends: the non-LTR retrotransposable elements. New Biol. **4:** 430–440

37 Malik H. S., Burke W. D. and Eickbush T. H. (1999) The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. **16:** 793–805

38 Bouhidel K., Terzian C. and Pinon H. (1994) The full-length transcript of the *I* factor, a LINE element of *Drosophila melanogaster*, is a potential bicistronic RNA messenger. Nuc. Acids Res. **22:** 2370–2374

39 McMillan J. P. and Singer M. F. (1993) Translation of the human LINE-1 element, L1Hs. Proc. Natl. Acad. Sci. USA **90:** 11533–11537

40 Luan D. D., Korman M. H., Jakubczak J. L. and Eickbush T. H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72:** 595–605

41 Luan D. D. and Eickbush T. H. (1995) RNA template requirements of target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. **15:** 3882–3891

42 George J. A., Burke W. D. and Eickbush T. H. (1996) Analysis of the 5′ junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. Genetics **142:** 853–863

43 Zimmerly S., Guo H., Perlman P. S. and Lambowitz A. M. (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. Cell **82:** 545–554

44 Firtel R. (1972) Changes in the expression of single-copy DNA during the development of the cellular slime mold *Dictyostelium discoideum*. J. Mol. Biol. **66:** 363–377

45 Blumberg D. D. and Lodish H. F. (1980) Complexity of nuclear and polysomal RNA's in growing *Dictyostelium discoideum* cells. Dev. Biol. **78:** 268–284

46 Blumberg D. D. and Lodish H. F. (1980) Changes in the messenger RNA population during differentiation of *Dictyostelium discoideum*. Dev. Biol. **78:** 285–300

47 Loomis W. F. and Insall R. H. (1999) Box 1: the *Dictyostelium* genome project. Nature **401:** 440–442

48 Loomis W. F. and Insall R. H. (1999) Cell biology – a cell for all reasons. Nature **401:** 440–441

49 Kay R. R. and Williams J. G. (1999) The *Dictyostelium* genome project – an invitation to species hopping. Trends Genet. **15:** 294–297

50 Sucgang R., Shaulsky G. and Kuspa A. (2000) Toward the functional analysis of the *Dictyostelium discoideum* genome. J. Eukaryt. Microbiol. **47:** 334–339

51 Kuspa A., Sucgang R. and Shaulsky G. (2001) The promise of a protist: the *Dictyostelium* genome project. Funct. Integr. Genomics **1:** 279–293

52 Glöckner G. (2000) Large scale sequencing and analysis of AT rich eukaryotic genomes. Curr. Genomics **1:** 289–299

53 Glöckner G., Eichinger L., Szafranski K., Pachebat J., Dear P., Lehmann R. et al. (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature **418:** 79–85

54 Glöckner G., Szafranski K., Winckler T., Dingermann T., Quail M., Cox E. et al. (2001) The complex repeats of *Dictyostelium discoideum*. Genome Res. **11:** 585–594

55 Zuker C., Cappello J., Chisholm R. L. and Lodish H. F. (1983) A repetitive *Dictyostelium* gene family that is induced during differentition and by heat shock. Cell **34:** 997–1005

56 Rosen E., Sivertsen A. and Firtel R. A. (1983) An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in *Dictyostelium*. Cell **35:** 243–251

57 Poole S. J. and Firtel R. A. (1984) Genomic instability and mobile genetic elements in regions surrounding two discoidin I genes of *Dictyostelium discoideum*. Mol. Cell. Biol. **4:** 671–680

58 Szafranski K., Glöckner G., Dingermann T., Dannat K., Noegel A. A., Eichinger L. et al. (1999) Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* transfer RNA genes. Mol. Gen. Genet. **262:** 772–780

59 Winckler T., Tschepke C., deHostos E. L., Jendretzke A. and Dingermann T. (1998) Tdd-3, a tRNA gene-associated poly(A) retrotransposon from *Dictyostelium discoideum*. Mol. Gen. Genet. **257:** 655–661

60 Marschalek R., Brechner T., Amon-Böhm E. and Dingermann T. (1989) Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science **244:** 1493–1496

61 Hofmann J., Schumann G., Borschet G., Gosseringer R., Bach M., Bertling W. M. et al. (1991) Transfer RNA genes from *Dictyostelium discoideum* are frequently associated with repetitive elements and contain consensus boxes in their 5′-flanking and 3′-flanking regions. J. Mol. Biol. **222:** 537–552

62 Winckler T. (1998) Retrotransposable elements in the *Dictyostelium discoideum* genome. Cell. Mol. Life Sci. **54:** 383–393

63 Wells D. J. (1999) Tdd-4, a DNA transposon of *Dictyostelium* that encodes proteins similar to LTR retroelement integrases. Nucl. Acids Res. **27:** 2408–2415

64 Leng P., Klatte D., Schumann G., Boeke J. and Steck T. (1998) Skipper, an LTR retrotransposon of *Dictyostelium*. J. Mol. Biol. **26:** 2008–2015

65 Malik H. S. and Eickbush T. H. (1999) Modular evolution of the integrase domain in the Ty3/*gypsy* class of LTR retrotransposons. J. Virol. **73:** 5186–5190

66 Marschalek R., Hofmann J., Schumann G., Gosseringer R. and Dingermann T. (1992) Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* tRNA genes. Mol. Cell. Biol. **12:** 229–239

67 Hentschel U., Zündorf I., Dingermann T. and Winckler T. (2001) On the problem of establishing the subcellular localization of *Dictyostelium* retrotransposon TRE5-A proteins by biochemical analysis of nuclear extracts. Anal. Biochem **296:** 83–91

68 Feng Q. H., Moran J. V., Kazazian H. H. and Boeke J. D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87:** 905–916

69 Yang J., Malik H. S. and Eickbush T. H. (1999) Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. Proc. Natl. Acad. Sci. USA **96:** 7847–7852

70 Minakami R., Kurose K., Etoh K., Furuhata Y,. Hattori M. and Sakaki Y. (1992) Identification of an internal *cis*-element essential for the human L1 transcription and a nuclear factor(s) binding to the element. Nucleic Acids Res. **20:** 3139–3145

71 Becker K. G., Swergold G. D., Ozato K. and Thayer R. E. (1993) Binding of the ubiquitous nuclear ranscription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum. Mol. Genet. **2:** 1697–1702

72 Geier A., Horn J., Dingermann T. and Winckler T. (1996) Nuclear protein factor binds specifically to the 3′-regulatory module of the long-interspersed-nuclear-element-like *Dictyostelium* repetitive element. Eur. J. Biochem. **241:** 70–76

73 Horn J., Dietz-Schmidt A., Zündorf I., Garin J., Dingermann T. and Winckler T. (1999) A *Dictyostelium* protein binds to distinct oligo(dA) · oligo(dT) DNA sequences in the C-module of the retrotransposable element DRE. Eur. J. Biochem. **265:** 441–448

74 Winckler T., Trautwein C., Tschepke C., Neuhäuser C., Zündorf I., Beck P. et al. (2001) Gene function analysis by *amber* stop codon suppression: CMBF is a nuclear protein that supports growth and development of *Dictyostelium* amoebae. J. Mol. Biol. **305:** 703–714

75 Beck P., Dingermann T. and Winckler T. (2002) Transfer RNA gene-targeted retrotransposition of *Dictyostelium* TRE5-A into a chromosomal UMP synthase gene trap. J. Mol. Biol. **318:** 273–285

76 Marschalek R., Hofmann J., Schumann G. and Dingermann T. (1992) Two distinct subforms of the retrotransposable DRE element in NC4 strains of *Dictyostelium discoideum*. Nucleic Acids Res. **20:** 6247–6252

77 Marschalek R., Hofmann J., Schumann G., Bach M. and Dingermann T. (1993) Different organization of the transfer RNA-gene-associated repetitive element, DRE, in NC4-derived strains and in other wild-type *Dictyostelium discoideum* strains. Eur. J. Biochem. **217:** 627–631

78 Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J. et al. (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860–920

79 Marschalek R. and Dingermann T. (1991) Structure, organization and function of transfer RNA genes from the cellular slime mold *Dictyostelium discoideum*. Adv. Gene Technol. **2:** 103–143

80 Bukenberger M., Dingermann T., Meissner W., Seifart K. H. and Winckler T. (1994) Isolation of transcription factor IIIC from *Dictyostelium discoideum*. Eur. J. Biochem. **220:** 839–846

81 Kessin R. H. (2001) *Dictyostelium* – Evolution, Cell Biology and the Development of Multicellularity, Cambridge University Press, Cambridge

82 Thomas K. R. and Capecchi M. R. (1996) Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature **324:** 34–38

83 Reindl N. (1991) Expression modifizierter RNA Polymerase II und III Gene in *Dictyostelium*, thesis, University of Erlangen/Nürnberg, Germany

84 Moran J. V., DeBerardinis R. J. and Kazazian H. H. (1999) Exon shuffling by L1 retrotransposition. Science **283:** 1530–1534

85 Boeke J. D. and Devine S. E. (1998) Yeast retrotransposons: finding a nice quiet neighborhood. Cell **93:** 1087–1089

86 Roeder G. S. and Fink G. R. (1982) Movement of yeast transposable elements by gene conversion. Proc. Natl. Acad. Sci. USA **79:** 5621–5625

87 Sharon G., Burkett T. J. and Garfinkel D. J. (1994) Efficient homologous recombination of Ty1 element cDNA when integration is blocked. Mol. Cell. Biol. **14:** 6540–6551

88 Hoff E. F., Levin H. L. and Boeke J. D. (1998) *Schizosaccharomyces pombe* retrotransposon Tf2 mobilizes primarily through homologous cDNA recombination. Mol. Cell. Biol. **18:** 6839–6852

89 Kim J. M., Vanguri S., Boeke J. D., Gabriel A. and Voytas D. F. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res. **8:** 464–478

90 Voytas D. and Boeke J. (1993) Yeast retrotransposons and tRNAs. Trends Genet. **9:** 421–427

91 Voytas D. (1996) Retroelements in genome organization. Science **274:** 737–738

92 Curcio M. J. and Morse R. H. (1996) Tying together integration and chromatin. Trends Genet. **12:** 436–438

93 Kirchner J., Conolly C. M. and Sandmeyer S. B. (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science **267:** 1488–1491

94 Yieh L., Kassavetis G., Geiduschek E. P. and Sandmeyer S. B. (2000) The Brf and TBP subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the *gypsy*-like element Ty3. J. Biol. Chem. **275:** 29800– 29807

95 Aye M., Dildine S. L., Claypool J. A., Jourdain S. and Sandmeyer S. B. (2001) A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. Mol. Cell. Biol. **21:** 7839–7851

96 Zou S. and Voytas D. F. (1997) Silent chromatin determines target preference of the retrotransposon Ty5. Proc. Natl. Acad. Sci. USA **94:** 7412–7416

97 Xie W., Gai X., Zhu Y., Zappulla D. C., Sternglanz R. and Voytas D. F. (2001) Targeting of the yeast Ty5 retrotransposon to silent chromatin is medaited by interactions between integrase and Sir4p. Mol. Cell. Biol. **21:** 6606–6614

98 Marschalek R., Borschet G. and Dingermann T. (1990) Genomic organization of the transposable element Tdd-3 from *Dictyostelium discoideum*. Nucleic Acids Res. **18:** 5751–5757

99 Katinka M. D., Duprat S., Cornillot E., Méténier G., Thomarat F., Prensier G. et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Enzephalitozoon cuniculi*. Nature **414:** 450–453

100 Wood V., Gwilliam R., Rajandream M. A., Lyne M., Stewart A., Sgouros J. et al. (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature **415:** 871–880

101 Perreau V., Santos M. and Tuite M. (1997) *beta*, a novel repetitive DNA element associated with tRNA genes in the pathogenic yeast *Candida albicans*. Mol. Microbiol. **25:** 229–236

102 Behrens R., Hayles J. and Nurse P. (2000) Fission yeast retrotransposon Tf1 integration is targeted to 5′ ends of open reading frames. Nucleic Acids Res. **28:** 4709–4716

103 Singelton T. L. and Levin H. L. (2002) A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukar. Cell **1:** 44–55