Molecular Analysis of the LTR Retrotransposon Ylt1 from the Genome of Dimorphic Fungus *Yarrowia lipolytica*

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt
der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Diplom-Biologe Andriy Kovalchuk

geboren am 27.02.1980 in Lviv, Ukraine

Gutachter:
Prof. Dr. rer. nat. habil. G. Barth
Prof. Dr. rer. nat. habil. A. Hyman
PD Dr. rer. nat. habil. R. Schaffrath

Eingereicht am: 15.08.2005

Tag der Verteidigung:

Contents

List of abbreviations

A	acetate
amp^R	β-lactamase-encoding gene, which confers resistance to ampicilline
APS	ammonium persulfate
ARS	autonomously replicating sequence
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
CA	capsid protein
cDNA	DNA molecule produced from RNA template during reverse transcription
CEN	centromere sequence
C-end	carboxy terminal end
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	2´-desoxynucleoside-5´-triphosphate
E	ethanol
EDTA	ethylenediaminetetraacetatic acid
En	endonuclease
G	glucose
Gag	structural protein of LTR retrotransposons and retroviruses
HA	hemagglutinin epitope
ICL1	gene encoding isocitrate lyase
IN	integrase
kb	kilobase (1000 bp)
kDa	kilodalton
lacZ	gene encoding ß-galactosidase from *E. coli*
Leu	leucine
LEU2	gene encodig β-isopropylmalate dehydrogenase
LINE	long interspersed nuclear element(s)
LTR	long terminal repeat
M	minimal medium
Met	methionine
mRNA	messenger RNA
msDNA	multiple-copy single-stranded DNA

MW	molecular weight
NC	nucleocapsid protein
N-end	amino terminal end
OD	optical density
ORF	open reading frame
ORI	origin of replication
PAGE	polyacrylamide gel electrophoresis
PBS	primer-binding site
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulfonylfluoride
pol	polyprotein-encoding gene
Pol	polyprotein
PPT	polypurine tract
PR	protease
PVDF	polyvinylidene fluoride
rDNA	gene encoding ribosomal RNAs
RH	ribonuclease H
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecylsulfate
SINE	short interspersed nuclear element(s)
TAE	Tris-Acetat-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
TSD	target site duplication
Ura	uracil
URA3	gene encoding orotidine-5´-phosphate decarboxylase
VLP	virus-like particle(s)
WT	wild type
Y	glycerol

1 Introduction

1.1 The overview of retroelements

One of the most abundant genes in cellular organisms encodes reverse transcriptase (RT), the enzyme which creates double-stranded DNA from an RNA template [Frame *et al.*, 2001]. Transposable elements containing reverse transcriptase are abundant in the eukaryotic genomes, whereas the prokaryotic ones are only sparsely inhabited with retroelements. Due to their abundance, retroelements often have a significant impact on the genome evolution, being involved in such diverse processes as genome rearrangements, regulation of gene expression and even telomere maintenance. Conventionally, the following main groups of retroelements are recognized, the LTR retrotransposons, the vertebrate retroviruses, the non-LTR retrotransposons, the retroplasmids, the retrointrons and the prokaryotic "retrons", as well as hepadnaviruses (also known as mammalian pararetroviruses) and caulimoviruses (plant pararetroviruses) [Boeke and Stoye, 1997].

1.1.1 Retroelements of prokaryotes and eukaryotic organelles

Retroelements described from the prokaryotic genomes are among the most unusual and the less understood ones. There are so-called retrons, which in some cases were shown to be the part of larger mobile elements, strongly resembling cryptic prophages and usually referred as "retronphages" [Garfinkel, 1992; Boeke and Stoye, 1997]. Their presence was shown in *Myxococcus* species [Yee *et al.*, 1984], several strains of *Escherichia coli* [Lampson *et al.*, 1989; Lim and Maas, 1989; Sun *et al.*, 1989] and some *Rhizobium* species [Rice *et al.*, 1993].

The retrons contain a single ORF, which encodes their own reverse transcriptase. Interestingly, there are no further genes whose products would provide the element movement within host genome. A retron is transcribed by host RNA polymerase and the originating mRNA is then processed by host enzymes to a shorter form. This form further serves as a template for reverse transcriptase. The priming mechanism is highly unusual as the same RNA molecule is both primer and template. Further on, 2′-OH group of the internal guanosine residue serves as a priming group rather than the usual 3′-OH group. The reverse transcription results in the formation of the so-called msDNA molecule (multi-copy single-stranded DNA), consisting of RNA and DNA strands linked via a 2′-5′ phosphodiester bond [Boeke and Stoye, 1997]. msDNAs are abundant satellite molecules, those copy number can reach over 500 copies per genome [Dhundale *et al.*, 1988]. They do not confer either any selective advantages or even any noticeable phenotype on their host.

As mentioned before, retrons are often a part of much larger prophages. The last ones can mediate the transfer of retrons among bacterial strains. So, the retron transmission mediated by P4 phage of *E. coli* was shown experimentally [Inouye *et al.*, 1991]. On the other hand, the direct mobility of retrons itself has never been demonstrated. Taking into account the absence of any effect on host-cell phenotype, it remains unclear how retrons have maintained themselves within certain bacterial lineages so successfully.

Other highly unusual groups of retroelements inhabit both the genomes of eukaryotic organelles (mitochondria and chloroplasts) and prokaryotes. There are so-called retroplasmids and retrointrons [Boeke and Stoye, 1997]. Both groups were shown to be phylogenetically closely related to bacterial retrons [Xiong and Eickbush, 1990].

Retroplasmids were described from the mitochondria of the filamentous fungus *Neurospora crassa*. Two types of retroplasmids are known, Mauriceville and Varkud plasmids, named after the strains where they were isolated from [Nargang *et al.*, 1984; Nargang, 1986]. Both are highly similar and contain a single RT-encoding ORF. In the course of transcription, a long transcript that exactly covers the entire element is produced [Akins *et al.*, 1988]. Interestingly, there is an unusual tRNA-like structure at the 3´ end of the transcript. Similar structures were reported for several plant RNA viruses [Miller *et al.*, 1986].

The reverse transcription process uses this transcript as a template. The reverse transcriptase of retroplasmids is unique among DNA polymerases studied so far as it lacks an absolute primer requirement [Wang and Lambowitz, 1993]. It recognizes the tRNA-like structure of the transcript and initiates primer-independent reverse transcription. RT produces full-length linear DNA [Kennell *et al.*, 1994], whereas the mechanism of the final circularization step is currently unknown.

Strains containing retroplasmids usually show no phenotypic changes. However, it was demonstrated that insertions of retroplasmids into mitochondrial DNA may result in a senescent phenotype [Akins *et al.*, 1986]. Further on, growth-impaired mutants with increased copy number of retroplasmids were reported [Akins *et al.*, 1986; Akins *et al.*, 1989].

Retrointrons, often referred as mobile group II introns, are primary found in bacterial and organellar genomes. They exhibit properties of both catalytic RNA and transposable element and possess a single ORF encoding a protein with reverse transcriptase (RT) and DNA endonuclease (En) activities.

Retrointrons use an exciting mobility mechanism. The intron RNA catalyzes its own splicing from a pre-mRNA, resulting in an excised intron lariat with a 2´-5´ phosphodiester bond. This structure reverse splices directly into a DNA target site (this process is mediated by the

En activity) and is then reverse transcribed by the intron-encoded protein [Lambowitz and Zimmerly, 2004]. The autocatalytic splicing of the retrointrons from mRNAs produced from targeted genes allows them to minimize host damage.

Retrointrons are capable of both retrohoming (movement to a cognate DNA allele that lacks the intron) and retrotransposition to new loci [Curcio and Derbyshire, 2003]. The retrohoming is characterized by high efficiency, resulting in the occupation of ~90 % of the progeny alleles in the crosses between yeast strains containing different combinations of mitochondrial introns [Lambowitz and Zimmerly, 2004]. Conversely, retrotransposition to novel sites occurs with low frequency (10^{-5}-10^{-4}) [Dickson *et al.*, 2001; Ichiyanagi *et al.*, 2002]. Even in this case integration target still should resemble the normal homing site. Such strict specificity is explained by the mechanism of the target site recognition, primarily involving the base-pairing interactions of intron RNA with target DNA, although the endonuclease does show some specificity for the nucleotides around the cleavage site as well [Lambowitz and Zimmerly, 2004].

Group II introns are the proposed ancestors of both nuclear spliceosomal introns and nuclear non-LTR retrotransposons [Cavalier-Smith, 1991; Lambowitz and Zimmerly, 2004]. It is suggested that the separation of the processes of transcription and translation in eukaryotic cells may favor the using of host-encoded proteins, thus leading to the development of the common splicing machinery and to the loss of intron-encoded proteins. On the other hand, retrointrons, which by chance had not inserted within genes, could lose their autocatalytic properties and thus give rise to non-LTR retrotransposons.

Retroelements are sometimes considered as remnants of an ancient “RNA world”, where RNA molecules served both as the genetic material and the biocatalysts [Boeke and Stoye, 1997]. Thus, both retroplasmids and retrointrons could be regarded as descendants of this proposed RNA-dominated era, retaining such unique features as primer-independent DNA synthesis (retroplasmids) or RNA-catalyzed self-splicing (retrointrons). Finally, they are probable ancestors of the present-day retrotransposons [Lambowitz and Zimmerly, 2004].

1.1.2 Non-LTR retrotransposons

Retroplasmids and retrointrons inhabit genomes of eukaryotic organelles, but they are, with few exceptions, absent from eukaryotic nuclear genomes [Lambowitz and Zimmerly, 2004]. The nuclear genomes of modern eukaryotes are invaded primarily by their more successful offspring. In the rare cases when retrointrons are still found in the eukaryotic chromosomes, they are integrated in the fragments of mitochondrial DNA [Lin *et al.*, 1999]. Thus, they were apparently transferred to the nucleus together with the last ones. So, the nuclear integration of the fragments

of mitochondrial DNA could be regarded as one of the processes leading to the transfer of group II introns to the nucleus. Such events may have led finally to the development of modern nuclear retroelements. Among them the non-LTR retrotransposons are currently considered to be the most ancient ones [Malik *et al.*, 1999]. At the same time, they are one of the most important retroelements, taking into account their high number, structural diversity and distribution among eukaryotes.

Non-LTR retrotransposons are occasionally termed also poly(A) or TP-retrotransposons [Boeke and Stoye, 1997; Curcio and Derbyshire, 2003]. Other names for this group are retroposons or long interspersed nuclear elements (LINEs). The first name reflects the one of the structural features of these elements – the absence of direct long terminal repeats, so characteristic for the members of another group, LTR retrotransposons. However, with the recent description of DIRS-like retrotransposons as another class of retroelements that lacks canonical LTRs this name could not be no longer accepted as the reliable one [Curcio and Derbyshire, 2003]. However, it is still widely used in the literature. For this reason it will be preserved in this work, too. The second name, poly(A) retrotransposons, is given for the presence of 3´ poly(A), oligo(A) or similar sequences (e.g., $[TAA]_n$) in the majority of these elements [Boeke and Stoye, 1997]. Again, there are some exceptions from this scheme. Therefore, a new name was proposed based on their mechanism of mobility: target-primed retrotransposons, or TP-retrotransposons [Curcio and Derbyshire, 2003].

TP-retrotransposons are represented in virtually all investigated groups of eukaryotes (baker's yeast *Saccharomyces cerevisiae* being a notable exception) [Boeke and Stoye, 1997]. They were especially successful in mammalian genomes. For example, there are about 850,000 copies of autonomous LINEs in human genome, and their fraction in the genome is as high as 21 %. The most abundant of them, L1 element, is present in more than 500,000 copies, accounting for 16 % of the human genome [Lander *et al.*, 2001]. Non-LTR retrotransposons are generally less abundant in plant genomes [Schmidt, 1999], but there are still some examples of element's extreme amplification, as in the case of del2 retrotransposon, which comprises 4 % of the genome of *Lilium speciosum* [Leeton and Smyth, 1993].

It was mentioned already, that non-LTR retrotransposons lack direct terminal repeats. Moreover, most of them are completely devoid of any kind of terminal repeats, except for the target site duplications [Boeke and Stoye, 1997]. On the other hand, they often have poly(A) or A-rich sequence at their 3´ end. Non-LTR retrotransposons are generally expressed from their own promoters. The organization of coding sequences and the nature of encoded proteins slightly vary among the non-LTR retroelements [Malik *et al.*, 1999]. The most primitive members of this

group contain a single ORF encoding for a protein with reverse transcriptase and restriction-enzyme-like site-specific endonuclease activities. The majority of non-LTR retrotransposons, however, acquired a second ORF encoding for a protein with nucleic-acid binding activity. This acquisition was accompanied by the appearance of the AP (apurinic-apyrimidinic)-endonuclease activity, encoded in the one reading frame with RT. A small number of TP-retrotransposons additionally encode RNaseH activity. Finally, phylogenetically distinct class of presumed TP-retrotransposons, so-called Penelope-like elements, possesses endonucleases of recently described GIY-YIG type, which were not previously associated with retroelements [Lyozin *et al.*, 2001; Arkhipova *et al.*, 2003; Curcio and Derbyshire, 2003].

The life cycle of non-LTR retrotransposons was studied primarily on R1Bm and R2Bm elements of the silkworm *Bombyx mori* and on the human L1 element. In contrast to LTR-retroelements, some details of their retrotransposition mechanism are still poorly understood. The majority of TR-retroelements uses their own promoter sequences in the 5′ region of the element. The resulting transcripts are translated into element's proteins and are used as a template for reverse transcription. A complex of full-length RNA and reverse transcriptase/endonuclease (RT/En) initiates the transposition by cleaving an antisense strand of target DNA. The resulting 3′-OH group serves as a primer for reverse transcription [Luan *et al.*, 1993; Curcio and Derbyshire, 2003]. Further on, the transposition machinery of some non-LTR retrotransposons can use preexisting DNA nicks and double-strand breaks to initiate reverse transcription [Morrish *et al.*, 2002]. In any case, RT performs the synthesis of cDNA on the full-length RNA of the element. Notably, the synthesis rarely reaches the 5′ end of the RNA, and the large fraction of non-LTR retrotransposons in eukaryotic genomes are truncated at the 5′ end. It is still unclear whether the second-strain synthesis is performed by the element's reverse polymerase or by host DNA-repair enzymes. Similarly, the details of the cDNA attachment to the upstream region target of target DNA are not fully understood [Curcio and Derbyshire, 2003].

It should be noted that, although the majority of non-LTR retrotransposons shows no specific integration preferences, some more primitive elements encoding site-specific endonuclease activity are tightly associated with certain targets. Thus, R1Bm and R2Bm elements of *B. mori* are adapted to highly conserved regions of the 28S rDNA [Xiong and Eickbush, 1988], whereas CRE-like elements of trypanosomes are site-specific for mini-exons arrays of the host genome [Aksoy *et al.*, 1987].

The non-LTR retrotransposons are believed to be responsible for the retrotransposition of the non-autonomous SINEs (short interspersed nuclear elements) and for the creation of the processed pseudogenes [Boeke and Stoye, 1997]. The first ones are short sequences (100-

400 bp) harboring an internal polymerase III promoter and encoding no proteins. Interestingly, 3′ ends of many SINEs are similar to the 3′ ends of LINEs, providing an evidence for LINE machinery-dependent transposition. One of their best known representatives is human *Alu* element. Promoter regions of SINEs are derived either from tRNA or 7SL RNA sequences. Although they do not encode their own transposition machinery, they are remarkably successful. For instance, there are more than 1,500 thousands copies of SINEs in human genome, accounting for 13 % of the genome. Alone the *Alu* element is present in more than one million copies [Lander *et al.*, 2001].

Processed pseudogenes differ from their "parent" genes by loss of introns, presence of 3′ poly(A) track and lack of the native external promoter sequence. Taken together, these features strongly suggest their origin via the reverse transcription of an mRNA intermediate.

Some of the modern-day non-LTR retroelements are still transpositionally active. Their expression is generally tightly regulated and is often restricted to germ line cells [Boeke and Stoye, 1997]. At the same time, the transposition of mammalian L1 elements was observed in some tumor cell (e.g., breast and colon cancer) [Morse *et al.*, 1988; Miki *et al.*, 1992]. The consequences of such transposition are best studied for mammalian L1-like elements. Thus, the mutations caused by human L1 element in the factor VIII and in the dystrophin genes resulted in hemophilia A [Kazazian *et al.*, 1988] and in Duchenne muscular dystrophy [Narita *et al.*, 1993], correspondingly.

Non-LTR retrotransposons were the first elements that successfully invaded eukaryotic chromosomes. It is generally accepted that they gave rise to the next group of retroelements, LTR retrotransposons. On the other hand, the RT sequence of LINE-like elements shows a remarkable similarity to telomerase, that maintains the integrity of telomeres of eukaryotic linear chromosomes [Boeke and Stoye, 1997; Curcio and Derbyshire, 2003]. However, the phylogenetic relationships of both enzymes are currently not fully understood [Eickbush, 1997; Nakamura and Cech, 1998].

Finally, two very special TP-retrotransposons of fruit flies *Drosophila* should be mentioned here. They are the HeT-A and TART elements, which are responsible for *Drosophila* telomere maintenance. Unlike the majority of eukaryotes, fruit flies lack the telomerase enzyme and characteristic telomeric repeats. Instead, the repeated transposition of HeT-A and TART elements to the chromosomal ends provides their stability [Levis *et al.*, 1993].

1.1.3 LTR and DIRS1-like retrotransposons

LTR retrotransposons are characterized by the most complex structure among retroelements. At the same time, the details of their life cycle are mostly well-understood. Our knowledge of the biology of LTR retrotransposons benefits significantly from the study of the transposable elements residing in the genomes of the well-known model organisms, the budding yeast *S. cerevisiae* and the fruit fly *Drosophila melanogaster*.

The LTR retrotransposons are closely related to vertebrate retroviruses; they lack, however, an *env* gene, and their RNA-containing virus-like particles (VLPs) are non-infectious (there some known exceptions, which will be discussed later) [Boeke and Stoye, 1997].

LTR retrotransposons are widely distributed in the eukaryotic kingdom. They were found virtually in all major groups with the exception of some species of ancient origin as *Giardia lamblia* and trypanosomes [Malik and Eickbush, 2001]. Interestingly, the LTR retrotransposons are the most abundant in the genomes of insects, gymnosperms and flowering plants, whereas genomes of vertebrates are only sparsely occupied by the members of this group. The most striking examples of enormous expansion of LTR retrotransposons can be found in grass genomes (family *Poaceae*). Thus, their fraction in the maize genome is at least 50 % [San Miguel and Bennetzen, 1998], whereas in the genome of wheat there are, by different estimations, up to 75-80 % of LTR retroelements [Wicker *et al.*, 2001; San Miguel *et al.*, 2002]. LTR retrotransposons are abundant in the genomes of other higher plants, too. So, elements with copy numbers $>10^4$ were reported from the genomes of *Lilium* [Sentry and Smith, 1989] and *Pinus* species [Kriebel, 1985; Kossack and Kinlaw, 1999].

The most common structural feature of LTR retrotransposons is the presence of direct long terminal repeats (LTRs) flanking the both ends of the element. They contain regulatory sequences responsible for the regulation of the retrotransposon expression, including RNA polymerase II promoter, termination and polyadenylation signals and some additional regulatory motifs. Notably, the LTRs of retrotransposons and retroviruses commonly share the consensus terminal sequences TG…CA [Boeke and Stoye, 1997].

Other sequences important for the replication of LTR retrotransposons are the so-called primer-binding site (PBS) and the polypurine tract (PPT). PBS is located immediately downstream from 5´ LTR. It is usually complementary to a specific cellular tRNA used as a primer for the synthesis of the minus-strand cDNA ((-)-cDNA) It should be noted, however, that some known LTR retrotransposons have evolved alternative mechanisms of priming and do not depend on cellular tRNAs for their reverse transcription. The polypurine tract is a short purine-rich

sequence stretch adjacent to 3′ LTR. It serves as a primer for the synthesis of the plus-strand cDNA ((+)-cDNA).

The protein-coding sequences of LTR elements are located between the both LTRs. Their organization varies between LTR elements, but generally there are two main regions called *gag* and *pol*. They can be either arranged in one reading frame or in different frames separated by -1 or +1 frameshifting [Boeke and Stoye, 1997; Gao *et al.*, 2003]. Some invertebrate elements have acquired an additional reading frame showing some similarity to the *env* gene of vertebrate retroviruses. As this acquisition confers them infectional properties, these elements are occasionally called invertebrate retroviruses. It was shown, however, that several independent acquisition events occurred in different lineages of LTR retrotransposons, and the *env*-like genes were acquired from various viral sources [Malik *et al.*, 2000].

The *gag* coding region encodes for structural protein with RNA-binding activity, so-called Gag (group-specific antigen) protein. It is a main component of virus-like particles of LTR retrotransposons. Gag proteins of different LTR retrotransposons show very limited sequence similarity. In fact, the only well-conserved region is a Zn-finger motif responsible for the binding of nucleic acids. It contains one histidine and three cysteine residues and can be shown schematically as $CX_2CX_4HX_4C$ [Covey, 1986].

A polyprotein (Pol) encoded by the *pol* region includes following enzymatic activities: protease (PR), reverse transcriptase (RT), RNaseH (RH) and integrase (IN). Proteases encoded by LTR retrotransposons belong to the class of aspartic proteases. They are characterized by the presence of certain conservative amino acids in the active center including an absolutely invariable aspartic acid residue: $(\text{hydrophobic residues})_2$-D-T/S-G-A/S. The protease is produced as a part of the polyprotein. First, it releases itself by specifically cutting polypeptide chain at both ends of its sequence. Then it performs a specific cleavage of the remaining fragments to yield a reverse transcriptase and integrase as well as mature forms of Gag protein [Dunn *et al.*, 2002].

The enzymatic activities of reverse transcriptase and RNaseH remain on the same polypeptide chain after the processing of the polyprotein. The first one has a key function in the replication of LTR retroelements, namely, it carries out the process of reverse transcription. The amino acid sequence of reverse transcriptase was shown to be the most conserved part of the retroelements [Xiong and Eickbush, 1990]. Therefore, it was widely used for the analysis of phylogenetic relationships between and inside different groups of retroelements. One of the characteristic motifs of the reverse transcriptase is the catalytic core YXDD box containing one tyrosine and two aspartic acid residues. RNaseH is responsible for the degradation of the RNA strand in the

heteroduplex DNA-RNA during the reverse transcription. Its characteristic feature is a presence of the conserved TDAS motif.

Finally, an integrase performs an integration of cDNA in the host genome. All integrases encoded by LTR retrotransposons belong to the class of DDE-transposases, named after the characteristic catalytic DD(35)E motif. Therefore, LTR retrotransposons are newly named DDE-retroelements [Curcio and Derbyshire, 2003]. Another important integrase motif is HH-CC, located in the N-terminal domain and involved in the binding to LTR sequences. Furthermore, some integrases of LTR retrotransposons have acquired additional modules, which are believed to be important for the specificity of cDNA integration. Among them are the GPY/F domain found in numerous Ty3/*gypsy* retrotransposons and the chromodomain [Malik and Eickbush, 1999].

An important requirement for the efficient transposition of LTR retrotransposons is a certain ratio of expressed Gag and Pol proteins. The Gag protein should be expressed in molar excess of Pol to ensure the protein maturation and VLPs assembly. Changes in their stoichiometry can severely reduce transposition [Xu and Boeke, 1990; Kirchner *et al.*, 1992; Kawakami *et al.*, 1993; Farabaugh, 1995]. LTR retrotransposons evolved different strategies allowing them to reach the necessary Gag:Pol ratio [Gao *et al.*, 2003]. Many of them express Gag and Pol proteins in different reading frames. In this case ribosomal frameshift (either -1 or +1) is needed for the expression of the Pol protein, and the frameshift frequency determines the Gag:Pol ratio. Alternatively, *gag* and *pol* regions may be organized in one reading frame but separated by stop codon, so that stop codon read-through ensures Pol protein expression. Finally, certain LTR retrotransposons have a single ORF and utilize posttranslational regulation mechanisms to achieve required Gag:Pol ratio instead, e.g. through preferential Pol degradation. The recent report of Gao *et al.* [2003] suggests that the type of the regulation strategy used by retroelements is significantly influenced by their host organisms. So, the elements of *Arabidopsis thaliana* preferentially have their *gag* and *pol* regions arranged in a single ORF whereas the majority of the insect (*D. melanogaster* and *B. mori*) LTR retrotransposons utilize -1 translational frameshift for their expression. Finally, an equal percentage of elements with *gag* and *pol* in a single frame or in -1 or +1 overlapping frames were found in fungi. Interestingly, *gag* and *pol* of the most of the yeast retrotransposons are separated by +1 frameshift, which rarely was found outside this group.

The classification of LTR retrotransposons is primarily based on the amino acid sequence of their reverse transcriptase [Xiong and Eickbush, 1990]. Currently, three main groups are recognized, Ty1/*copia*, Ty3/*gypsy* and BEL retrotransposons [Frame *et al.*, 2001]. The fourth

group, DIRS1-like elements, was placed previously close to the canonical LTR retrotransposons, but now they are recognized as a distinct class of retroelements [Goodwin and Poulter, 2001a]. Their features will be discussed at the end of this chapter. Three groups of viruses, retroviruses, hepadnaviruses and caulimoviruses, replicate through the process of reverse transcription. They are grouped together with LTR retrotransposons by the analysis of their RT sequences [Xiong and Eickbush, 1990], but due to their infectivity their biological properties are quite different and they will be briefly described in a separate section.

It was already mentioned, that three groups of canonical LTR retrotransposons are defined largely by their RT sequence similarity. The Ty1/*copia* group is named after its prominent members, Ty1 retrotransposon of *S. cerevisiae* and *copia* element of *D. melanogaster*. This group is believed to be the most ancient one among LTR retrotransposons [Xiong and Eickbush, 1990; Malik and Eickbush, 2001]. Its members occupied the genomes of fungi (including numerous yeast species), plants and animals. The characteristic feature of all known Ty1/*copia* elements is the domain arrangement in their *pol* gene: PR-IN-RT-RH [Boeke and Stoye, 1997]. Until recently, it was believed to be an exclusive attribute of the Ty1/*copia* group. Surprisingly, Goodwin and Poulter [2002] described recently a peculiar group of deuterostome Ty3/*gypsy*-like retrotransposons with Ty1/*copia*-like *pol*-domain orders. This finding emphasizes again that our current knowledge of retroelements diversity and evolution are still incomplete, and the investigation of further eukaryotic genomes may bring unexpected results.

Again, the Ty3/*gypsy* group receives its name for the well-studied Ty3 retrotransposon of *S. cerevisiae* and *gypsy* element of *D. melanogaster*. It includes a great variety of retroelements occurring in all main groups of eukaryotes. Its members show a high similarity to vertebrate retroviruses, differing from them mainly in the absence of *env* gene. Both Ty3/*gypsy*-like retrotransposons and retroviruses share an order of the coding regions in *pol* gene: PR-RT-RH-IN (with the exception of *Gmr1*-like retrotransposons mentioned above) [Boeke and Stoye, 1997]. Another feature common for the majority of Ty3/*gypsy* elements is the presence of the so-called GPY/F domain in the integrase (it should be mentioned that GPY/F domain was obviously lost in some lineages of these retrotransposons) [Malik and Eickbush, 1999].

The group of BEL-like retrotransposons was characterized recently [Frame *et al.*, 2001]. All its known members are restricted in their distribution to animal genomes (BEL element itself was described from *D. melanogaster* genome). The arrangement of coding regions in *pol* gene of BEL-like retrotransposons (PR-RT-RH-IN) would suggest their placement among Ty3/*gypsy* elements. However, the comparison of RT sequences unambiguously placed them outside both the Ty1/*copia* and the Ty3/*gypsy* groups. They are characterized by significant structural

heterogeneity. There are only few common features like the presence of a triplet of Zn-finger motifs near the C-terminal end of Gag region, and a single long ORF. Notably, the sequences of reverse transcriptase and aspartic protease active sites of BEL-like elements show some striking differences from the other groups of LTR retrotransposons.

Another group of retrotransposons, DIRS1-like elements, displays a similarity with LTR-containing elements in their RT sequence. However, several details of their structural organization and their life cycle were shown to differ significantly from those of the canonical LTR retrotransposons, placing them in a separate class, DIRS1 or Y-retrotransposons [Goodwin and Poulter, 2001a]. So, what are the main differences between DIRS1 and LTR retrotransposons? First of all, DIRS1-like elements encode neither DDE-type integrase nor aspartic protease. The integration of their cDNA is performed by the Y-transposases instead. These enzymes are related to the site-specific recombinase of the bacteriophage lambda. They are proposed to insert the circular DNA intermediate into the host genome [Curcio and Derbyshire, 2003]. Thus, their mechanism of integration differs significantly from the one described for LTR retrotransposons, those integrases mediate the insertion of linear molecules of cDNA. Further on, DIRS1 retrotransposons lack characteristic LTRs and do not produce target site duplications. They are bordered either by inverted repeats or by "split" direct repeats instead. These repeats are likely to have dual functions – namely, to produce full-length cDNA from terminally truncated RNAs and serve as recombination sites for the Y-transposase. The details of the reverse transcription and integration of DIRS1 elements are largely unknown. It is suggested, however, that the circular cDNA molecule produced in the course of reverse transcription is further integrated by the mechanism related to the one described for bacteriophage lambda.

The number of known DIRS1-like elements grows continuously since their characterization as a separate class. They were found in the genomes of slime mold (*Dictyostelium discoideum*) [Cappello *et al.*, 1985], fungi [Ruiz-Perez *et al.*, 1996], green algae [Duncan *et al.*, 2002], nematodes [de Chastonay *et al.*, 1992], arthropods, sea urchins and fishes [Goodwin and Poulter, 2001a]. Surprisingly, no DIRS1-like elements were reported so far from the transposon-rich genomes of plants and insects.

The description of DIRS1 retrotransposons highlights the remarkable ability of retroelements to acquire novel functional modules and adopt their activities for own needs.

1.1.4 Vertebrate retroviruses

There is a little doubt that vertebrate retroviruses are the most important group of LTR-containing elements. Nevertheless, their properties will be discussed here only briefly, as the genomic transposable elements are in the focus of this overview.

Due to their importance, the different aspects of retroviral biology were widely studied during the last decades. Our current knowledge on retroelements biology is based mainly on the results obtained in experiments with retroviruses.

Structurally the retroviruses resemble LTR retrotransposons. Again, their genomes are flanked by two identical LTRs, and they contain *gag* and *pol* genes. A crucial difference, however, is a presence of the *env* gene in the retroviral genomes. Its product mediates the entry of the virion into the host cell and provides in this way the infectious properties of the retrovirus. Many retroviruses have only these three genes. Yet, there are some important exceptions. First, genomes of some retroviruses acquired copies of cellular genes. The cellular counterparts of acquired genes are often involved in the key processes of signal transduction or cell cycle regulation. Thus, they may confer upon the virus the ability to transform normal into malignant cells and to cause tumors in the host animal. In this case, the acquired cellular genes are called oncogenes, and corresponding viruses – "transforming" viruses [Coffin, 1992; Vogt, 1997].

Further on, some retroviruses (so-called complex retroviruses) also contain accessory genes, which are usually involved in the coordination and regulation of the viral gene expression [Vogt, 1997]. Their examples are spumaviruses and lentiviruses.

In contrast to LTR retrotransposons, the retroviruses can produce extracellular virus particles – virions. They are enveloped by lipid membrane derived from the cell plasma membrane by budding. The products of *env* gene are associated with outer membrane and mediate the recognition of host cells. There are two protein layers inside the envelope composed mainly of different forms of Gag protein. The outer shell is generally called matrix, whereas the inner one is termed capsid. The retroviral enzymes together with dimeric genomic RNA and a variety of small RNAs (mainly tRNAs) of host origin are located within the capsid [Coffin, 1992; Vogt, 1997].

The life cycle of the most known retroviruses includes, in contrast to LTR retrotransposons, an extracellular phase. However, it seems that little occurs inside the virion during this phase. The process of reverse transcription takes place mainly after the entry of the viral core into the cytoplasm of the target cell, following by the integration of cDNA into the host genome and propagation of viral RNAs and proteins. It should be noted that certain retroviruses have persisted as stable genome-integrated forms (proviruses) for multiple generations [Boeke and

Stoye, 1997]. They have often undergone several adaptations allowing them to coexist with their host. So, the viruses induced from endogenous proviruses are relatively nonpathogenic, whereas many endogenous proviruses are transcriptionally silent or defective due to occurred mutations. Nevertheless, many of them are very successful in the colonization of host genomes. For instance, human endogenous retroviruses constitute for more than 8 % of human genome [Lander *et al.*, 2001]. Many their properties resemble those of LTR retrotransposons, although their RT sequences unambiguously place them among retroviruses.

The canonical retroviruses are restricted in their distribution to vertebrates, infecting fishes, amphibians and reptiles as well as birds and mammals, including human. Further on, retroviruses can cause a variety of diseases in the infected organisms, including a number of malignancies, immunodeficiencies, and neurological disorders [Rosenberg and Jolicoeur, 1997; Fauci and Desrosiers, 1997]. Some disorders caused by retroviruses are often fatal. The world-wide distribution of the HIV infection is one of the major public health problems faced by the mankind in the 21st century. Unfortunately, the evolutionary plasticity of retroviruses and their ability for rapid adaptations hamper the control of retrovirus infections.

1.1.5 Hepadnaviruses and caulimoviruses

Besides the retroviruses, two other viral families utilize the process of reverse transcription for their replication. There are hepadnaviruses and caulimoviruses, which are occasionally called animal and plant pararetroviruses, correspondingly [Boeke and Stoye, 1997]. The most obvious difference between canonical retroviruses and pararetroviruses is the type of nucleic acid found in the virion and the stage of the life cycle when the reverse transcription occurs. As described above, retroviral virions contain dimeric genomic RNA, reverse transcribed into cDNA after the entry of viral core into the cytoplasm of infected cell. Conversely, the reverse transcription process in the case of pararetroviruses occurs prior to virion release, and viral particles contain DNA rather than RNA molecules. The pararetroviruses have circular DNA genomes with interruptions either on plus strand (hepadnaviruses) or on both strands (caulimoviruses). It is another important difference from retroviruses characterized by linear DNA genomes. As in the case of retroviruses and LTR retrotransposons, a host tRNA molecule serves as a primer for reverse transcription of caulimoviruses, whereas reverse transcription of hepadnaviruses is primed by a reverse transcriptase protein itself rather than by a nucleic acid. Finally, unlike the retroviruses, the life cycle of the pararetroviruses does not involve integration of the viral genome into that of the host, all the replication being episomal [Boeke and Stoye, 1997].

Hepadnaviruses are known only from a small number of mammals and birds. Viral infection leads to hepatitis (so-called type B hepatitis) and is associated with a high rate of liver cancer in humans and animals [Jilbert and Mason, 2002]. Caulimoviruses are parasites of flowering plants, transmitted by aphids and other insect vectors. They induce mainly mottles and mosaics of host plants, and some of them cause considerable losses to tropical crops [Hull, 2001].

1.2 Phylogeny of retroelements

The known retroelements display a great diversity in their structure, encoded proteins and the mechanisms of replication (Table 1.1). What major events did lead to the emergence of such remarkable diversity? The conception of modular evolution, i.e. the stepwise acquisition of new information modules, was proposed to explain the main steps in the development of the variety of modern retroelements [Xiong and Eickbush, 1990; Boeke and Stoye, 1997; Malik and Eickbush, 1999; Malik *et al.*, 1999]. It is suggested that modern groups of retroelements have arisen from a certain ancestor by gradual acquisition of additional functional domains whose functions were adopted by retroelements.

Nowadays, the most primitive of the retroelements are found in the genomes of bacteria and eukaryotic organelles. This group includes retrons, retroplasmids and retrointrons. They are characterized by several archaic features and encode either only reverse transcriptase or, in the case of retrointrons, reverse transcriptase and endonuclease. Currently, they are known only from a few species of eukaryotes and they are generally present in host genome in a limited copy number. Thus, they are mainly regarded as “living fossils” among retroelements. However, the elements similar to modern retrointrons are believed to be the ancestors of present-day retrotransposons.

Indeed, the known retrointrons occupy the genomes of eukaryotic mitochondria and plastids, and there are multiple examples of mitochondrion-to-nucleus or chloroplast-to-nucleus gene transfer. This process may have contributed to the retrointrons migration to nuclear DNA. Furthermore, the nuclear genome of *A. thaliana* even contains an integrated fragment of presumably nonfunctional mitochondrial DNA that includes group II introns [Lin *et al.*, 1999]. One of the proposed scenarios suggests that retrointrons that had not inserted within genes would be under no selective pressure to retain splicing. That enables them to evolve into non-LTR retrotransposons [Lambowitz and Zimmerly, 2004].

The most primitive of modern non-LTR retrotransposons have a single ORF encoding a protein with RT and En activities [Malik *et al.*, 1999]. Such organization of coding sequences generally resembles the organization of group II introns. Another archaic feature of these elements is their

sequence-specific integration mechanism. It is proposed that in the course of evolution such primitive non-LTR retroelements have acquired AP-endonuclease domain from the DNA repair machinery of the cell. This acquisition resulted in the lost of target site specificity. It also coincides with the appearance of a second ORF in front of the major RT-encoding ORF [Malik *et al.*, 1999]. This additional ORF encodes for a protein with nucleic acid binding motifs, whose function is currently not completely understood. Further on, certain non-LTR retrotransposons acquired an RNaseH domain from some eukaryotic source.

Table 1.1 Comparison of known classes of retroelements

Retroelements	Encoded proteins	Terminal sequences	Distribution
Retrons	RT	-	Bacterial genomes
Retroplasmids	RT	Circular DNA molecules; RNA transcript has a tRNA-like structure at the 3´ end	Autonomous elements occurring in mitochondria of filamentous fungus *Neurospora*
Retrointrons	RT and En	-	Genomes of bacteria and eukaryotic organelles
Non-LTR retrotransposons	RT, RH*, En, nucleic acid-binding protein*	Poly(A) or related sequences at the 3´ end; no terminal repeats	Nuclear genomes of eukaryotes
LTR retrotransposons	RT, RH, IN (En), Gag, PR	Direct terminal repeats at both termini (LTRs)	Nuclear genomes of eukaryotes
DIRS1-like retrotransposons	RT, RNaseH, Y-transposase, Gag	Inverted terminal repeats or "split" direct repeats	Nuclear genomes of eukaryotes (have not been reported from plant genomes so far)
Vertebrate retroviruses	RT, RH, IN, Gag, PR, Env (other proteins of viral or host origin are sometimes also encoded)	Direct terminal repeats at both termini (LTRs)	Infect vertebrate animals; integration of viral DNA into host nuclear genomes occurs after infection
Hepadnaviruses	RT, RH, Gag, Env and X protein with unknown function	Circular DNA molecules with interruptions on plus strand	Infect mammals and birds; viral DNA is not integrated into host genomes
Caulimoviruses	RT, RH, coat protein, PR, movement protein and some proteins with unknown functions	Circular DNA molecules with interruptions on both strands	Infect plants; viral DNA is not integrated into host genomes

* - these proteins are encoded only by some non-LTR retrotransposons

The next class of retroelements, LTR retrotransposons, is characterized by significant complexity of their structure and life cycle compared to non-LTR elements. Several researches proposed that these elements derived from the fusion of a DNA-mediated transposon and a non-LTR retrotransposon [Capy *et al.*, 1998; Malik and Eickbush, 2001]. The proposed fusion event must have occurred at least twice in the evolution of retroelements, ones leading to the acquisition of DDE-integrase by LTR retrotransposons and, second, resulting in the appearance of DIRS1-like

elements with their tyrosine transposases [Curcio and Derbyshire, 2003]. On the other hand, non-LTR elements contributed the RT-RNaseH region and the domain encoding for nucleic acid-binding protein (Gag-like) to LTR retrotransposons. Also, LTR retrotransposons acquired the protease domain, which may have been derived from the host's pepsin gene family. Finally, the elements evolved long direct terminal repeats to overcome the problem of replicating the ends of their genomic DNA [Malik and Eickbush, 2001].

The most significant difference between LTR retrotransposons and simple retroviruses is the presence of an *env* gene, which confers infectious ability to retroviruses [Coffin, 1992; Vogt, 1997]. So, the most likely explanation for the origin of vertebrate retroviruses is the acquisition of the *env* gene from other viral source. This model is supported by numerous phylogenetic analyses and is currently widely accepted. Unfortunately, the rapid sequence divergence of the *env* genes has made it difficult to ascertain the origins of the *env* genes in retroviruses [Malik *et al.*, 2000]. It is also unclear whether *env* genes represent a single acquisition event or multiple events. Malik and Eickbush [2001] have proposed that vertebrate retroviruses have acquired additional RNaseH domain, probably from non-LTR retrotransposons. At the same time, the original domain lost its enzymatic activity but may still carry out some important structural functions, being represented by the so-called "tether" or "connection" domain between RT and RH regions.

An additional evidence for the acquisition of the *env* gene by vertebrate retroviruses is provided by numerous examples of invertebrate retrotransposons bearing *env*-like ORFs. For instance, it was shown that the *gypsy* element of *D. melanogaster* has acquired its envelope-like gene from a class of insect baculoviruses, whereas Cer elements from *Caenorhabditis elegans* bear *env*-like genes originating from phleboviruses [Malik *et al.*, 2000]. On the other hand, it was proposed that plant caulimoviruses represent a fusion of LTR element with a plant virus [Xiong and Eickbush, 1990].

The current model of the evolution of retroelements suggests a stepwise acquisition of new informational modules and their further adaptation for element's needs to play a key role in the appearance of their new classes. This evolutionary approach led to the development of a vast diversity of retroelements, which are currently by far the most successful players among eukaryotic mobile elements.

1.3 The life cycle of LTR retrotransposons and their interaction with host genomes

Our current knowledge of the retrotransposition mechanism of LTR retrotransposons are based mainly on the results of experiments performed with Ty1 and Ty3 elements of *S. cerevisiae* and

copia and *gypsy* retrotransposons of *D. melanogaster*. It reflects the basic principles of the life cycle of LTR retrotransposons, but the study of further elements has brought some examples showing significant deviations in the regulation of expression, in the priming of reverse transcription, in the approaches to achieve proper Gag:Pol ratios and in other life cycle details among LTR retrotransposons. Some of these examples will be discussed here as well.

Unlike the retroviruses, genomic transposable elements should regulate their activity to minimize possible deleterious impact on the host fitness [Boeke and Stoye, 1997]. Thus, despite general similarity, there are some key differences in the biology of LTR retrotransposons and retroviruses.

1.3.1 Expression of LTR retrotransposons

The life cycle of LTR retrotransposons begins with the transcription of the elements integrated into host genomes with RNA polymerase II. Sequences resembling eukaryotic Pol II promoters were reported from LTRs of many known elements. Terminally redundant transcripts are produced during the transcription [Boeke and Stoye, 1997; Wilhelm and Wilhelm, 2001]. They bear R and U5 regions of LTR on the 5′ end and U3 and R regions on 3′ end. The regions were named for their distribution, namely, R for "repeated", and U5 and U3 for "unique for 5′ end" and "unique for 3′ end", respectively. This redundancy is required for the correct replication of linear genome of LTR retrotransposons.

It could be expected that the transcription of transposable elements is tightly regulated to minimize the level of the retrotransposition to levels tolerable by the host. Surprisingly, some of the well-known elements are transcribed so efficiently that their transcripts are among the most abundant of the cellular mRNA, i.e. Ty1 RNA constitutes 5 – 10 % [Curcio *et al.*, 1990; Boeke and Sandmeyer, 1991] and *copia* RNA – about 3 % of the total polyadenylated RNA in the host cell [Flavell *et al.*, 1980]. In fact, the *copia* element is even named for the abundance of its RNA.

A variety of expression patterns was described for different LTR retrotransposons. For instance, the expression of some of them is regulated in a developmental or tissue-specific manner [Boeke and Stoye, 1997]; the transcription of *S. cerevisiae* Ty3 elements is increased in *MAT***a** cells treated with α-factor [Van Arsdell *et al.*, 1987] and severely reduced in diploid (*MAT***a**/*MAT*α) strains compared to haploid ones [Boeke and Sandmeyer, 1991]. A silencing of transposons by mechanisms involving RNA interference (RNAi) was reported recently [Schramke and Allshire, 2003]. Conversely, many studied LTR retrotransposons are activated by such stress factors as UV irradiation, heat shock, pathogen infection and other stress conditions [Rolfe *et al.*, 1986; Bradshaw and McEntee, 1989; Ziarczyk and Best-Belpomme, 1991; Ratner *et al.*, 1992; Mhiri *et*

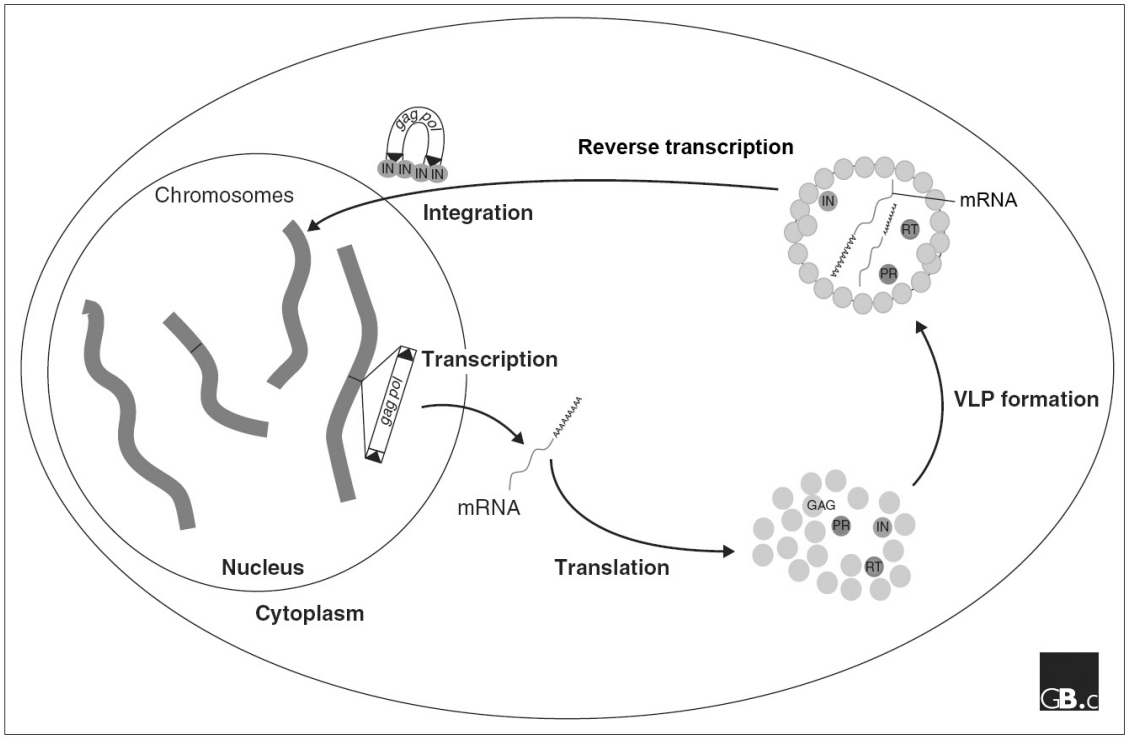

Figure 1.1 The life cycle of LTR retrotransposons. The scheme is taken from Havecker *et al.* (2004) with minor modifications

al., 1997; Ikeda *et al.*, 2001]. McClintock [1984] proposed that such stress activation of transposable elements could be regarded as an adaptive response of the genome and might lead to a new phenotype that can survive under severe conditions. Surprisingly, there are very few instances in which the increase of the retroelement transcription level has been directly correlated with the frequency of the transposition of the element [Boeke and Stoye, 1997]. Obviously, high transcription levels are not necessarily sufficient for frequent transposition.

An unexpected link between regulation of retrotransposons and host genes was recently observed for some organisms. Thus, LTRs of *Schizosaccharomyces pombe* elements were shown to be directly required for repression of nearby meiotically induced genes. This regulation mechanism depends on the components of RNAi pathway and silent chromatin [Schramke and Allshire, 2003]. Furthermore, LTR retrotransposons were reported to act as alternative promoters and first exons for a subset of host genes, regulating their expression in mouse oocytes and cleavage stage embryos [Peaston *et al.*, 2004]. These results indicated that the regulation of retrotransposon expression may be much more complex than it was suggested previously and that it is deeply implicated in the global regulation of cellular processes.

Translation of retroelement transcripts is performed on host ribosomes [Boeke and Stoye, 1997]. LTR retrotransposons are faced with the problem to produce the appropriate amounts of Gag and

Pol proteins during the translation. As noted above, the presence of molar excess of Gag protein is needed for their normal replication. Different approaches to reach the proper Gag:Pol ratio were evolved by various retrotransposons [Gao *et al.*, 2003]. One of the common strategies is the arrangement of *gag* and *pol* genes in different reading frames. In this case, either -1 or +1 translational frameshift is needed for the production of Pol protein. It is generally achieved by ribosome pausing caused by certain secondary structures in the template mRNA or by the use of rare codons [Farabaugh, 1996]. A number of frameshift-promoting sequences were described. A common structural motif for -1 frameshift site was identified as X-XXY-YYZ, where X, Y and Z are any nucleotides [Farabaugh, 1996]. On the other hand, sequences involved in +1 frameshift show little conservation. Another approach requires translational read-through of the stop codon between *gag* and *pol* genes for the synthesis of Pol protein, leading again to the production of a molar excess of Gag. An unusual mechanism was described for the *copia* element. In this case, two major transcripts are produced. The full-length transcript encodes the Gag-Pol fusion protein. The second transcript lacks most of the internal sequences, which are removed in frame by RNA splicing. This spliced RNA encodes a Gag-protease fusion protein [Brierley and Flavell, 1990; Yoshioka *et al.*, 1990].

Notably, some LTR retrotransposons with a single ORF rely on a post-transcriptional regulation of their Gag:Pol ratio. A good example of such regulation is provided by Tf1, a LTR retrotransposon from the genome of fission yeast *Sz. pombe*. Tf1 has a single reading frame, therefore both Gag and Pol proteins are initially produced in equimolar amounts. Gag:Pol ratio remains unchanged in log-phase cells, however, transition to the stationary phase is accompanied by rapid degradation of Pol proteins, resulting in the molar excess of Gag [Atwood *et al.*, 1996].

Finally, a proteolytic processing of Gag and Gag-Pol fusion is required for the maturation of individual proteins. It is generally carried out by a retroelement-encoded protease within the virus-like particles [Dunn *et al.*, 2002].

1.3.2 Assembly and maturation of virus-like particles

Although LTR retrotransposons lack an extracellular phase in their life cycle, they produce virus-like particles (VLPs), where the maturation of individual proteins and reverse transcription take place [Roth, 2000]. Details of VLPs assembly are currently less understood than other steps of the life cycle of LTR retrotransposons. It was studied primarily with Ty1 and Ty3 elements using overexpression strategy [Garfinkel *et al.*, 1985; Hansen *et al.*, 1992].

It is generally accepted that Gag and Gag-Pol fusion proteins assemble into VLPs. Moreover, element's RNA and host tRNA molecules are encapsulated within particles [Boeke and Stoye,

1997; Roth, 2000]. Some differences concerning the localization of the particles were observed between retroelements. In the case of yeast cells overexpressing either Ty1 or Ty3 element, VLPs are observed in the cytoplasm [Roth, 2000], whereas *copia* VLPs in *Drosophila* tissue culture cells were found in the nucleus [Miyake *et al.*, 1987].

The protease expressed as a part of the Pol polyprotein is able to release itself by specifically cutting peptide bonds on both ends of its sequence [Dunn *et al.*, 2002]. Then it processes the Gag protein to produce so-called capsid (CA) and nucleocapsid (NC) proteins. The CA protein is a main structural component of VLPs and plays a central role in particle assembly [Roth, 2000]. The significantly smaller NC protein encompasses the C-terminal part of the Gag protein including a nucleic acid-binding Zn-finger motif. It is presumably involved in the scaffolding of genomic RNA inside the VLPs.

The proteolytic processing of Gag-Pol fusion protein results in the release of CA and NC proteins, the protease, the integrase and the reverse transcriptase. The protease-cleavage sites were characterized in details for the Ty3 element [Kirchner and Sandmeyer, 1993]. Apart from the relatively hydrophobic character of the flanking residues, these sites show only limited conservation.

1.3.3 Reverse transcription

The life cycle of LTR retrotransposons does not include extracellular phase, so, unlike the retroviruses, the reverse transcription process can take place immediately after the assembly of VLPs [Boeke and Stoye, 1997]. The majority of known LTR retrotransposons use certain host tRNA molecules to prime reverse transcription [Wilhelm and Wilhelm, 2001]. In this case, the 3´ acceptor stem of the specific host tRNA anneals to the short region downstream from the 5´ LTR, so-called primer-binding site (PBS) (Fig. 1.2). The length of PBS varies between 8 and 23 nt for the majority of known LTR retrotransposons [Neuveglise *et al.*, 2002]. Further on, the existence of interactions between the primer tRNA and genomic RNA in regions other than PBS sequence was demonstrated for several retroelements [Friant *et al.*, 1998; Gabus *et al.*, 1998]. Since tRNA molecules possess highly ordered secondary and tertiary structure, it was proposed that some factors should be involved in their unwinding and following annealing to the template RNA. Indeed, it was shown that the C-terminal region of the Ty1 Gag protein contains a nucleic acid chaperone domain capable of promoting the annealing of primer $tRNA^{iMet}$ to the PBS and the initiation of reverse transcription [Cristofari *et al.*, 2000]. Similar functions are proposed for NC proteins encoded by other retrotransposons.

Interestingly, some remarkable exceptions from this rule have been described. Thus, the elements of Tf1/*sushi* group use an unusual self-priming mechanism. It includes the annealing of 5′ end of retroelement's transcript to primer-binding site (PBS) and the cleavage of transcript by the RNaseH. The cleavage releases the 5′ end of the transcript, which serves as a primer for the reverse transcription [Levin, 1995; Levin, 1996; Lin and Levin, 1997a; Butler *et al.*, 2001]. Further on, some elements like *copia* and *S. cerevisiae* Ty5 use an internal portion of tRNA molecule, which includes the anticodon stem-loop, to prime their reverse transcription [Kikuchi *et al.*, 1986; Ke *et al.*, 1999]. Another priming mechanism was proposed for two closely related retrotransposons Tca3 and Tcd3 from the genomes of *Candida albicans* and *Candida dubliniensis*. The putative primer-binding site of these elements contains an inverted repeat, which probably is recognized by an RNase and cleaved to generate a primer [Goodwin *et al.*, 2003]. Finally, the PBSs of the slime mould element *skipper* and some related retrotransposons show neither obvious complementarity to tRNAs nor the similarity to PBSs of the Tf1/*sushi* group, and their actual priming mechanism remains unknown [Leng *et al.*, 1998; Goodwin and Poulter, 2001b].

The formation of a primer-template complex allows RT to perform the synthesis of the so-called minus-strand strong-stop DNA (-sssDNA), which encompasses U5 and R regions of the 5′ LTR (Fig. 1.2). When RT reaches the 5′ end of the RNA template, -sssDNA is released and transferred to the 3′ end of the genomic RNA, where its R region anneals to the complementary R region of 3′ LTR. The strand transfer allows the minus-strand DNA synthesis to be continued. Again, the RT proceeds until 5′ end of the template RNA is reached. Notably, during the synthesis of the minus-strand DNA, RNA strand of the newly formed DNA-RNA hybrid is degraded by the RT-associated RNaseH activity. However, specific purine-rich region located upstream from 3′ LTR (so-called polypurine tract, PPT) is relatively resistant to RNaseH digestion. Thus, it remains attached to the newly produced minus-strand DNA and serves as a primer for plus-strand DNA synthesis. The next step after the generation of PPT primer includes the synthesis of the plus-strand strong-stop DNA (+sssDNA). RT uses the 5′ end of the newly synthesized minus-strand DNA as a template and terminates reaching the first modified base in the priming tRNA, so that the +sssDNA encompasses the complete LTR and PBS sequences. Primer tRNA is removed from the +sssDNA by RNaseH activity. Finally, +sssDNA should be transferred to the 3′ end of the minus-strand DNA where it can anneal to the sequence complementary to PBS. The details of this process are not completely understood as such transfer would imply the inheritance of the primer tRNA sequence during the replication of LTR retroelements. Indeed, this seems to be the case for retroviruses but not for Ty1 element of

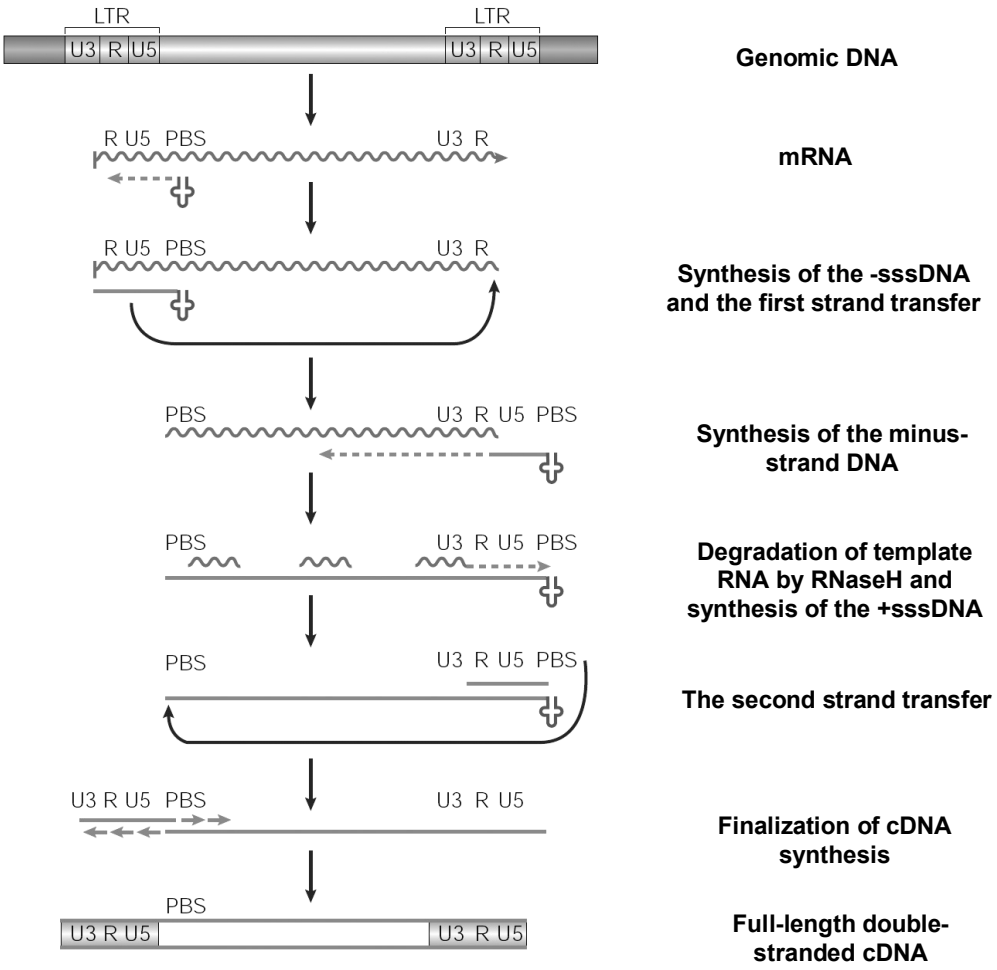

Figure 1.2 The reverse transcription of LTR retrotransposons. Abbreviations are as following: LTR, long terminal repeat; U3, R and U5 – corresponding fragments of the LTR; PBS, primer-binding site. LTR retrotransposon integrated in host genome is shown as shadowed box, its transcript - as wavy line, primer tRNA – as cloverleaf structure, newly synthesized single-strain DNA – as solid line, and double-stranded cDNA – as open box. See also text for more details. The scheme is taken from Curcio and Derbyshire (2003).

S. cerevisiae. Anyway, after the second strand transfer RT finalizes the synthesis of plus- and minus-strands of cDNA, and the whole process of reverse transcription results in a creation of linear genomic DNA of retroelement with characteristic terminal repeats (LTRs) (Fig. 1.2) [Wilhelm and Wilhelm, 2001].

1.3.4 Nuclear entry and integration of cDNA

The step of reverse transcription is followed in the life cycle of LTR retroelements by the integration of newly synthesized cDNA into the host genome. Most of them could enter the nucleus during mitosis when the nuclear envelope is broken down. However, the fungal retrotransposons are faced with an additional problem at this stage, since fungal cells undergo so-

called “closed” mitosis and their nuclear membrane remains intact throughout the cell cycle [Boeke and Stoye, 1997]. A similar hindrance arises in the life cycle of retrotransposons and retroviruses infecting non-dividing cells.

The details of the nuclear entry of fungal retrotransposons are largely unknown. It was shown that the integrase enzyme of Ty1 and Ty3 elements possesses a nuclear localization signal (NLS), which is required for the nuclear localization of the corresponding elements [Moore *et al.*, 1998; Lin *et al.*, 2001]. On the other hand, nuclear entry of Tf1 element from fission yeast depends on its Gag protein [Dang and Levin, 2000]. It is unlikely that VLPs may pass through the nuclear pore complex since their size exceeds the diameter of nuclear pore. So, several mechanisms were proposed to explain the nuclear import of yeast retrotransposons. First, it is assumed that VLPs dissociate before the nuclear entry and that their re-assembly occurs inside the nucleus. A second scheme suggests the formation of the preinitiation complex (PIC) consisting of integrase, cDNA and, probably, some other factors, which passes through the nuclear envelope, whereas other components of VLPs remain in cytoplasm.

The integration itself is performed by integrase. It introduces two staggered nicks in the target DNA and joins the 3′ ends of both strands of cDNA to the generated nicks [Curcio and Derbyshire, 2003]. It is thought that host DNA-repair enzymes fill in the single-stranded gaps at both junctions and produce in this way target site duplications in the host DNA.

The important prerequisite of the successful coexistence of transposable elements with their hosts is the minimization of the harmful effects of the retrotransposition. Therefore, many of the known LTR retrotransposons have developed the mechanisms allowing them to target their integration to the specific, presumably non-deleterious, regions of the host genome. Especially strongly expressed are the target preferences of the yeast retroelements, which have to adapt themselves to the compact genomes of their hosts lacking large intragenic or intergenic DNA regions. Moreover, yeasts spend a significant part of their life cycle in haploid phase, and almost every insertion in the coding sequence in this case would immediately affect host fitness [Boeke and Stoye, 1997].

Most of the known yeast retrotransposons avoid inserting into the protein-coding genes. For instance, four of five retrotransposon families (Ty1-Ty4) from the genome of budding yeast *S. cerevisiae* are strongly associated with genes transcribed by RNA polymerase III (Pol III), first of all, with tRNA genes [Kim *et al.*, 1998]. Integration into these sites is relatively harmless as the regions upstream tRNA genes are gene-poor and integration does not disrupt Pol III transcription. Notably, there are two different strategies of target site choice. Ty3 inserts almost exclusively within 1-2 bases of Pol III transcription start sites [Chalker and Sandmeyer, 1990;

Chalker and Sandmeyer, 1992; Kim *et al.*, 1998], whereas most of the Ty1, Ty2 and Ty4 insertions are found within a much wider 750 bp window upstream of target genes [Devine and Boeke, 1996; Kim *et al.*, 1998]. Further on, it was shown that the Ty1 element inserts quite readily into the silenced *HM* and rDNA loci [Boeke and Devine, 1998]. These apparently distinct insertion preferences suggest convergent evolution of targeting in different lineages of *S. cerevisiae* retrotransposons. Ty5 elements chose another "safe havens", namely, they integrate mainly near regions of silent chromatin at the telomeres and the *HM* mating loci [Zou *et al.*, 1995; Zou *et al.*, 1996]. Retrotransposons of fission yeast *Sz. pombe*, Tf1 and Tf2, have their own preferences. Their insertions are clustered in the intergenic regions containing RNA polymerase II promoters, 100 to 400 nucleotides upstream of protein-coding ORFs [Behrens *et al.*, 2000; Singleton and Levin, 2002; Bowen *et al.*, 2003].

In most cases the integration is not sequence-specific and the targeting is believed to be achieved mainly by interaction of integrase with certain host factors. Thus, the specificity of Ty5 integration is determined by interactions between the targeting domain of Ty5 integrase and the heterochromatin protein Sir4p, as was shown by experiments of Zhu *et al.* [2003]. Further on, it was demonstrated that components of RNA polymerase III complex TFIIIB and TFIIIC are involved in the targeting of Ty3 integration to Pol III-transcribed genes [Kirchner *et al.*, 1995; Yieh *et al.*, 2000].

The retrotransposons of multicellular organisms generally have no strict insertion preferences, what can be explained by the larger proportion of non-coding sequences in the genomes of higher eukaryotes. Nevertheless, even in this case a significant proportion of transposons is associated with heterochromatin regions.

Another issue important for the integrity of host genomes is the regulation of copy number of transposable elements. Again, the copy number control in the genomes of higher eukaryotes seems to be less stringent compared to compact genomes of yeast, which have evolved various mechanisms regulating the transposition. So, despite some variations in the copy number of Ty1 elements between *S. cerevisiae* strains, none of them has more than 40 copies of complete elements [Garfinkel *et al.*, 2003]. The mechanisms of such copy number control are far from being understood, but it was reported that they can be either involved in the regulation of transcription or can act at the post-transcriptional level [Garfinkel, 2003; Jiang, 2002].

Finally, it should be noted that transposable elements as well as their hosts have evolved a number of mechanisms allowing them to minimize deleterious impact of genomic parasites on host fitness.

1.4 Retroelements of yeast and filamentous fungi

Retrotransposons are widely distributed among eukaryotic organisms. A large variety of them was described from the genomes of yeast and filamentous fungi, too. Their study has made a significant progress in the last decade, especially due to the numerous sequencing projects deciphering complete fungal genomes and their fragments. Non-LTR retroelements as well as members of Ty1/*copia*, Ty3/*gypsy* and DIRS1 were found in fungal genomes. Although retroelements were reported from the genomes of ascomycetes, basidiomycetes and zygomycetes, the most of them were identified in ascomycetes. This bias is probably due to the high number of research groups working on various ascomycetes species and due to their wide use in biotechnology processes [Daboussi and Capy, 2003].

Non-LTR retrotransposons were reported from the genomes of several ascomycetous filamentous fungi: *Neurospora crassa* (Tad1 element) [Kinsey and Helber, 1989], *Ascobolus immersus* (*Mars*1) [Goyon *et al.*, 1996] and plant pathogenic species *Magnaporthe grisea* (also known as *Pyricularia oryzae*) (MGR583 and MGL) [Hamer *et al.*, 1989; Nishimura *et al.*, 2000] and *Colletotrichum gloeosporioides* (*Cg*T1) [He *et al.*, 1996]. They were also described from the genomes of some yeasts: two ascomycetous species, *C. albicans* (Zorro-1, Zorro-2 and Zorro-3) [Goodwin *et al.*, 2001] and *Yarrowia lipolytica* (Ylli) [Casaregola *et al.*, 2002], and medically important basidiomycetous yeast *Cryptococcus neoformans* (Cnl1) [Goodwin and Poulter, 2001b]. Further on, *marY2N* element was found in the genome of the economically important basidiomycetous mushroom *Tricholoma matsutake* [Murata *et al.*, 2001]. Notably, the genomes of *S. cerevisiae* and *Sz. pombe* lack the members of this group.

Most of the known fungal non-LTR retrotransposons are present in host genomes in high copy number. Similarly to LINE-element of other organisms, many of the copies are truncated at the 5′ end. Although full-length elements were reported in the most cases, the transpositional activity was currently demonstrated for Tad1 and MGL elements only [Daboussi and Capy, 2003].

Interestingly, all elements isolated from filamentous fungi (Tad1, *Mars*1, MGR583 and *Cg*T1) are closely related (Fig. 1.3). On the phylogenetic tree of non-LTR retrotransposons they form a separate branch, currently consisting exclusively from fungal elements [Malik *et al.*, 1999]. On the other hand, both Ylli and Zorro elements fall into the L1 group, characterized by broad distribution among eukaryotic kingdoms (Fig. 1.3) [Goodwin *et al.*, 2001; Casaregola *et al.*, 2002]. Thus, the yeast and fungal LINE-like elements are well distinct from each other. Surprisingly, the Cnl1 element of *C. neoformans* shows no similarity to other known fungal retrotransposons. The phylogenetic analysis placed it into the most ancient CRE group of non-

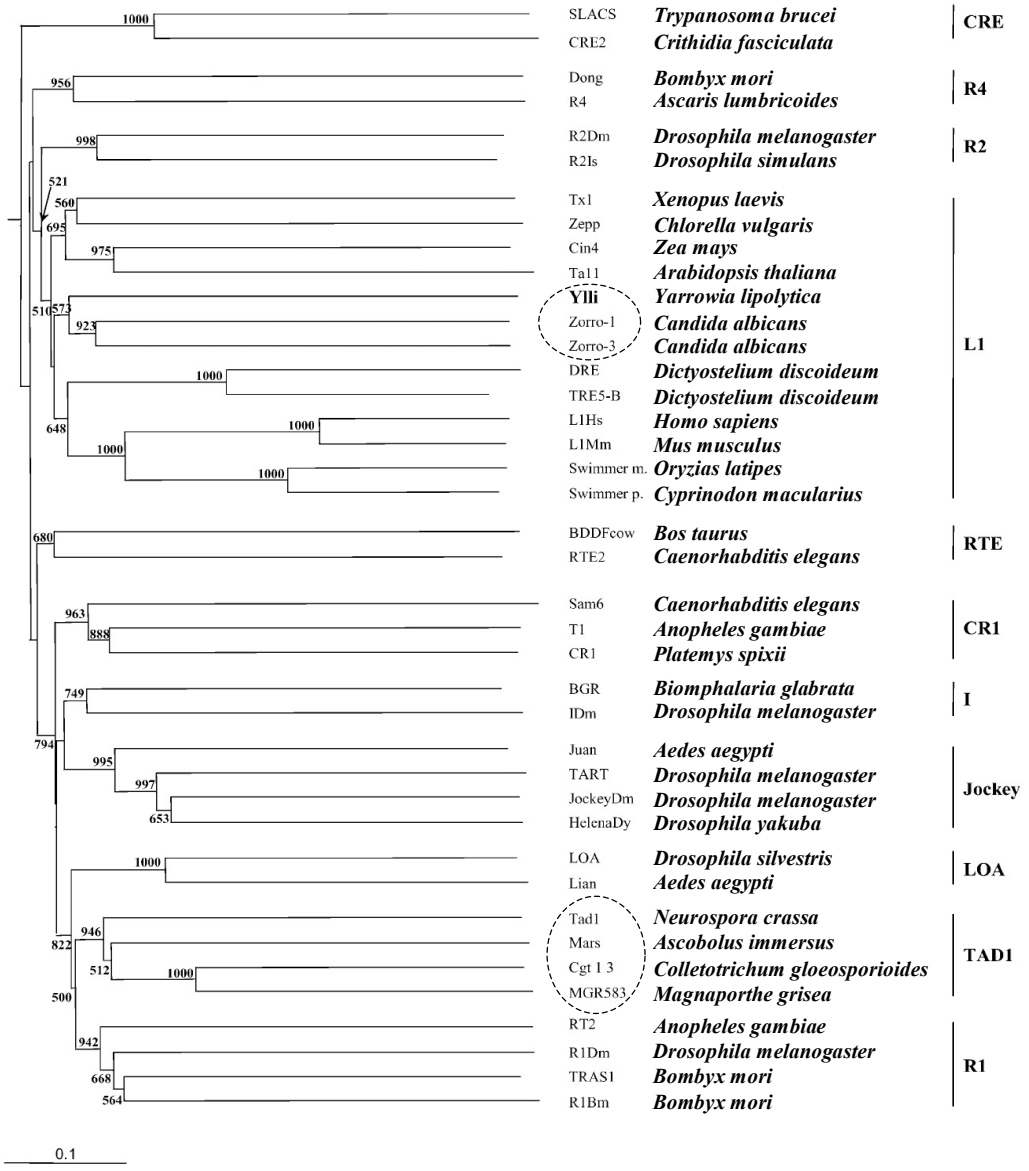

Figure 1.3 Phylogenetic relationships between non-LTR retrotransposons. The scheme includes all groups of non-LTR retrotransposons identified by Malik *et al.* (1999). Elements described from fungal genomes are indicated, and the retrotransposon Ylli from the genome of *Y. lipolytica* is shown in bold. This scheme does not include Cnl1 element of *Cryptococcus neoformans*, but analysis performed by Goodwin and Poulter (2001b) showed that Cnl1 is grouped together with the SLACS and CRE2 elements. The figure is taken from Casaregola *et al.* (2002).

LTR elements, which members were previously known only from the genomes of trypanosomes [Goodwin and Poulter, 2001b]. The position of *marY2N* element among non-LTR retrotransposons is currently unknown.

Sequences resembling SINE-like elements were described from several species of filamentous ascomycetes [Daboussi and Capy, 2003]. One of them, the *Foxy* element of *Fusarium oxysporum*, was shown to be transpositionally active, and its copy number may increase after

gamma irradiation [Mes *et al.*, 2000]. No corresponding LINE element mediating transposition of *Foxy* was described so far.

Although non-LTR retrotransposons are known from a number of fungi, their variety is limited compared to the diversity of LTR retrotransposons in fungal genomes. Thus, they are completely absent from the compact genomes of the majority of yeast species, where numerous LTR retrotransposons families still can be found. Probably, the members of this ancient group were eliminated from the compact genomes characterized by the rapid turnover of transposable elements. On the other hand, the genomes of numerous species of higher fungi (mushrooms) are currently poorly characterized and they may serve as a potential source of novel non-LTR elements.

Multiple families of LTR retrotransposons were identified in the genomes of yeast and filamentous fungi. Members of both Ty1/*copia* and Ty3/*gypsy* groups were reported from fungal genomes, although the first ones are found primarily in yeasts.

Numerous Ty1/*copia*-like elements occupy the genomes of *S. cerevisiae* and related hemiascomycetous yeasts [Boeke and Sandmeyer, 2001; Kim *et al.*, 1998; Neuveglise *et al.*, 2002]. Their sequences are mostly intact, and transpositional activity was demonstrated at least for some of them. At the same time, the members of this group found in the genomes of higher ascomycetous and basidiomycetous fungi often show the presence of frameshifts and/or stop codons indicating that they are degenerated [Daboussi and Capy, 2003].

Four groups of Ty1/*copia* retrotransposons were identified in hemiascomycetous yeasts [Neuveglise *et al.*, 2002]. Interestingly, all these groups are yeast-specific, i.e. currently known members of these groups were reported only from yeast genomes. Among them, Ty5-like elements are believed to be the most ancient ones [Neuveglise *et al.*, 2002]. All of them encode *gag* and *pol* genes in a single ORF. They were found in the genomes of *S. cerevisiae* (Ty5) [Voytas and Boeke, 1992], *S. paradoxus* (Ty5) [Zou *et al.*, 1995], *S. exiguus* (Tse5) [Neuveglise *et al.*, 2002], *C. albicans* (Tca5) [Plant *et al.*, 2000], *Pichia angusta* (Tpa5) [Neuveglise *et al.*, 2002] and *Debaryomyces hansenii* (Tdh5) [Neuveglise *et al.*, 2002]. All Ty5 elements of *S. cerevisiae* are inactive, so the active copy from *S. paradoxus* genome was used to study the properties of these elements [Zou *et al.*, 1995].

The next group, Tca2-like elements, is unique among yeast retroelements as its members have *gag* and *pol* genes separated by a stop codon [Neuveglise *et al.*, 2002]. Another common feature of these elements is the use of $tRNA^{Arg(UCU)}$ as a primer for reverse transcription. They were reported only from two yeast species, *C. albicans* (Tca2 and Tca4) [Matthews *et al.*, 1997; Goodwin and Poulter, 2000] and *D. hansenii* (Tdh2) [Neuveglise et al., 2002]. The Tca2 element

was shown to be transpositionally active [Holton *et al.*, 2001]. Interestingly, baker's yeast lacks this type of retroelements.

Two remaining groups, Ty1-like and Ty4-like elements, use +1 translational frameshift to express *gag-pol* polyprotein [Neuveglise *et al.*, 2002]. Ty1-like elements are widely distributed among various yeast species and were found, for example, in the genomes of *S. cerevisiae* (Ty1 and Ty2) [Cameron *et al.*, 1979; Clare and Farabaugh, 1985; Hauber *et al.*, 1985; Warmington *et al.*, 1985], *S. exiguus* (Tse1), *S. kluyveri* (Tsk1), *S. servazzii* (Tss1), *Kluyveromyces marxianus* (Tkm1) and *K. lactis* (Tkl1) [Neuveglise *et al.*, 2002]. At least Ty1 and Ty2 elements are transpositionally active. Conversely, Ty4-like elements were identified only in two species, *S. cerevisiae* (Ty4) [Stucka *et al.*, 1989] and *S. bayanus* (Tsu4) [Neuveglise *et al.*, 2002].

In contrast to the remarkable diversity of Ty1/*copia* retrotransposons in yeast genomes, only few mostly degenerated elements could be found in the filamentous ascomycetes [Daboussi and Capy, 2003]. Among them are *Mars*2 and *Mars*3 elements of *A. immersus* [Goyon *et al.*, 1996], Tcen of *N. crassa* [Cambareri *et al.*, 1998], Nht2 element of plant pathogen *Nectria haematococca* [Shiflett *et al.*, 2002] and *Elsa* from another phytopathogenic fungus *Stagonospora nodorum* [Rawson, 2000].

Further on, Ty1/*copia* retrotransposons were reported from three ectomycorrhizal basidiomycetes, *Laccaria bicolor* and two species of *Pisolithus* [Diez *et al.*, 2003]. Several elements were identified in the genome of basidiomycetous yeast *C. neoformans* (Tcn6-Tcn9) [Goodwin and Poulter, 2001b]; however, the entire sequence is available only for Tcn6.

The phylogenetic analysis of fungal Ty1/*copia* retrotransposons brought some surprising results. Whereas elements isolated from ascomycetous yeast and filamentous fungi form a separate group on the phylogenetic tree [Neuveglise *et al.*, 2002; Diez *et al.*, 2003], retrotransposons of basidiomycetes were placed closer to the transposons described from plants [Diez *et al.*, 2003]. These results suggest the independent origin of the members of Ty1/*copia* group in asco- and basidiomycetes, but the analysis of further elements from other fungal species is definitively required to clarify this issue.

A remarkable diversity of Ty3/*gypsy*-like elements was found in fungal genomes. First of all, the well-studied Ty3 element of *S. cerevisiae* itself should be noted [Hansen *et al.*, 1988; Hansen and Sandmeyer, 1990]. It is the only member of this group present in *S. cerevisiae* genome, and it is transpositionally active. Several partial sequences mostly referring to defective elements were reported from the genomes of other hemiascomycetous yeasts (*S. kluyveri*, *S. servazzii*, *D. hansenii*, *Candida tropicalis*, *Zygosaccharomyces rouxii* and *Y. lipolytica*) by Neuveglise *et al.* [2002]. Their phylogenetic positions were not analyzed. Further on, potentially active full-

length element Tse3 from *S. exiguus* was described in this study. It was shown to be related to the Ty3 element of *S. cerevisiae*. Two interesting Ty3/*gypsy* retrotransposons Tf1 and Tf2 occupy the genome of fission yeast *Sz. pombe* [Levin *et al.*, 1990]. They are members of the so-called Tf1/*sushi* group, which also includes fungal retrotransposons and some retroelements known from vertebrates [Butler *et al.*, 2001]. Both Ty3 and the members of the Tf1/*sushi* group belong to the larger group of chromoviruses, which are characterized by the presence of the chromodomain in the C-terminal part of an integrase [Marin and Llorens, 2000; Kordis, 2005]. Several fragments resembling Ty3/*gypsy*-like retrotransposons were found in the genome of *C. albicans* [Goodwin and Poulter, 2000; Goodwin *et al.*, 2003]. Phylogenetic analysis showed that at least two of them, Tca3 and Tca8, are not closely related to Ty3 and occupy an isolated position among known Ty3 elements. The only full-length Ty3/*gypsy* element currently known from *C. albicans* genome is Tca3. A closely related element Tcd3 was identified in the genome of *Candida dubliniensis* recently [Goodwin *et al.*, 2003]. Finally, the retrotransposon Ylt1 was described from dimorphic yeast *Y. lipolytica* [Schmid-Berger *et al.*, 1994].

Interestingly, the majority of known Ty3/*gypsy* elements from filamentous ascomycetes belong to the mentioned above Tf1/*sushi* group [Lin and Levin, 1997b; Butler *et al.*, 2001]. There are *MAGGY* and *Grasshopper* from *M. grisea* [Farman *et al.*, 1996; Dobinson *et al.*, 1993], CfT-1 from *Cladosporium fulvum* [McHale *et al.*, 1992], *Boty* from *Botrytis cinerea* [Diolez *et al.*, 1995], *Skippy* from *Fusarium oxysporum* [Anaya and Roncero, 1995], *REAL* from *Alternaria alternata* [Kaneko *et al.*, 2000], *Cg*ret from *Colletotrichum gloeosporioides* [Zhu and Oudemans, 2000] and *Afut*1 from *Aspergillus fumigatus* [Neuveglise *et al.*, 1996]. Intact full-length copies were described for most of these elements, but only *MAGGY* was shown to be transpositionally active [Nakayashiki *et al.*, 1999]. Further on, two retrotransposons of basidiomycetous fungi, Tcn1 from *C. neoformans* [Goodwin and Poulter, 2001b] and *marY1* from *Tricholoma matsutake* [Murata and Yamada, 2000], belong to the same group. Conversely, other Ty3/*gypsy*-like elements of *C. neoformans* (Tcn2-Tcn5) do not seem to be closely related with any of the previously known retrotransposons [Goodwin and Poulter, 2001b].

Thus, there are two main groups including the majority of known fungal Ty3/*gypsy*-like elements, Tf1/*sushi* and Ty3 *sensu stricto*. On the other hand, some yeast retrotransposons (like Tca3, Ylt1 or Tcn2-Tcn5) occupy an isolated position among known retroelements.

The only known fungal member of recently characterized DIRS1 group, Prt1, was described from the zygomycetous fungus *Phycomyces blakesleeanus* [Ruiz-Perez *et al.*, 1996]. Until now, the members of this unusual group were not found in the genomes of ascomycetes and basidiomycetes.

Finally, it should be note that a relatively small number of fungal species were investigated for the presence of retrotransposons, and further work in the field is required to get the whole picture of the evolution of fungal retroelements.

1.5 *Yarrowia lipolytica* as model organism

The ascomycetous dimorphic yeast *Y. lipolytica* (formerly also known as *Candida*, *Endomycopsis* or *Saccharomycopsis lipolytica*) is one of the most intensively studied "non-conventional" yeast species. It is the only known species of its genus, which is quite distantly related to the rest of the hemiascomycetous yeast genera and shares a number of common properties with filamentous fungi instead [Barth and Gaillardin, 1996; Barth and Gaillardin, 1997; Barth *et al.*, 2003].

All *Y. lipolytica* strains tested so far are heterothallic. The mating type is determined by the two alleles *MATA* and *MATB*. Both haploid and diploid cells are vegetatively stable. Notably, mating frequencies of natural isolates are always very low (1 % of viable zygotes/cells or even less). In currently used laboratory inbred strains it varies between 3 and 15 %. Other obstacles for the genetic analysis of *Y. lipolytica* are poor sporulation, frequent formation of incomplete tetrads and low ascospore viability in natural isolates. They were partly overcome by the creation of inbred strains with high proportion of complete tetrads and the spore germination reaching about 80 % [Barth and Gaillardin, 1996]. Low viability of spores resulting from the mating of natural isolates may be at least partially explained by the large genetic variation between individual isolates, observed in the polymorphism of chromosome length and genetic linkage groups.

Y. lipolytica is a natural dimorphic fungus, which forms yeast cells, pseudohyphae and septate hyphae. The proportion of the different cell types depends on the used strain and culture conditions.

1.5.1 Physiology and metabolism

The features that initially turned attention of researches to *Y. lipolytica* were its remarkable physiological characteristics. Strains of this species were commonly isolated from protein- or lipid-containing substrates like cheese and sausages. All currently known strains are obligate aerobes and can not survive under anaerobic conditions. *Y. lipolytica* uses few sugars (mainly glucose but not sucrose) as carbon source, but can utilize various polyalcohols, organic acids or normal paraffins instead. Ethanol and acetate also can serve as carbon source. They are tolerated in the concentrations up to 3 % and 1 % respectively. Further on, such unusual carbon sources as

1-alkenes, polymethylated or chlorinated alkanes can also be assimilated [Barth and Gaillardin, 1997].

One of the characteristic features of *Y. lipolytica* is its ability to secrete large amounts of certain metabolites, first of all, organic acids and extracellular proteins. Thus, wild type strains secrete a mixture of citric and isocitric acids when grown on *n*-paraffins as a carbon source. Thiamine limitation leads to the secretion of 2-ketoglutaric acid by cells grown on paraffins. Further on, processes for the production of 2-hydroxyglutaric acid and isopropylmalate using *Y. lipolytica* strains were developed [Barth and Gaillardin, 1997; Barth *et al.*, 2003].

Several proteins are secreted by *Y. lipolytica* strains, sometimes in amounts up to 1 - 2 g/l. Some of them including alkaline and acid extracellular proteases as well as extracellular RNase, lipases, esterase and phosphatases were isolated and characterized [Barth and Gaillardin, 1996]. The ability to secrete large amounts of proteins makes *Y. lipolytica* useful for the overproduction and secretion of heterologous proteins.

1.5.2 Genetics and molecular biology

A fair amount of data on molecular biology of *Y. lipolytica* was accumulated in the last decade. The recent completion of the genome sequencing project also significantly increases the potential of *Y. lipolytica* as a model organism.

Haploid strains of *Y. lipolytica* contain 6 chromosomes. The *Y. lipolytica* genome (20.5 Mb) is almost twice as large as the genome of *S. cerevisiae*. Nevertheless, the number of predicted protein-coding sequences is only 1.15-fold higher than in *S. cerevisiae* (6,703 vs. 5,807). In comparison with genomes of other hemiascomycetous yeasts, *Y. lipolytica* genome is characterized by the higher G+C content (49.0 % vs. 38.3 % in the case of *S. cerevisiae*), lower gene density, larger amount of non-coding sequences and relatively abundant introns, found in 13 % of genes [Casaregola *et al.*, 2000; Dujon *et al.*, 2004; http://cbi.labri.fr/Genolevures/elt/YALI].

Some other details emphasize the unique position of *Y. lipolytica* among well-studied hemiascomycetous yeasts. There are (1) the unusual organization of rDNA clusters, which do not include 5S rRNA genes; both rDNA clusters and 5S RNA genes are dispersed throughout the genome; (2) the absence of RNA polymerase I consensus sequence found in other yeasts and (3) larger size of snRNA and 7S RNA, resembling those of higher eukaryotes [Barth and Gaillardin, 1996; Casaregola *et al.*, 2000].

Another peculiar feature of *Y. lipolytica* is an inability of chromosomal origins of replication (*ORI*) alone to maintain the episomal replication of plasmids. A centromeric sequence (*CEN*) in

combination with *ORI* is required for the extrachromosomal plasmid maintenance instead. Therefore, autonomously replicating plasmids of *Y. lipolytica* are relatively stable and present in low copy numbers (usually 1-3 copies per cell). Interestingly, *ORI* and *CEN* sequences of *Y. lipolytica* show no homology to corresponding sequences of *S. cerevisiae* or *Kluyveromyces lactis* [Barth and Gaillardin, 1996].

Y. lipolytica is currently used as a model organism to investigate several cellular processes. Thus, significant progress was achieved in the study of peroxisome biogenesis and degradation, early steps of protein secretion, cellular dimorphism, utilization of hydrophobic substrates and dicarbon compounds and mitochondrial complex I biogenesis [Barth *et al.*, 2003].

1.5.3 Transposable elements of *Y. lipolytica*

The genomes of hemiascomycetous yeasts are generally characterized by the presence of various transposable elements. *Y. lipolytica* is not an exception, and the set of elements found in its genome shows remarkably diversity as it includes DNA transposons, LINE-like elements and LTR retrotransposons.

Ylt1 was the first transposable element identified in *Y. lipolytica* genome [Schmid-Berger *et al.*, 1994]. It was described as multi-copy element present in at least 35 full-length copies and more than 30 copies of solo LTR per haploid genome. However, it turned out later that copy number of Ylt1 varies between *Y. lipolytica* strains and that some strains of *Y. lipolytica* lack this element altogether (Fig. 1.4) [Juretzek *et al.*, 2001]. It is believed that the initial source of Ylt1 in laboratory strains of *Y. lipolytica* was the natural diploid isolate YB423. Its meiotic segregant YB423-12 was widely used in different inbreeding programs, and until now the presence of Ylt1 was observed only in YB423-12 itself or in strains having YB423-12 among their ancestors. Interestingly, no other natural isolate tested so far contains this element.

Ylt1 displays several interesting features. First of all, it is the largest fungal LTR retrotransposon described so far (9453 bp). The sequence analysis places it among Ty3/*gypsy* group, but it seems to be not closely related to other yeast retroelements. Ylt1 contains a single large ORF including *gag* and *pol* genes. Some properties of Ylt1 were characterized previously [Senam, 2004], but many details of its biology are still poorly understood.

For several years, Ylt1 was the only transposable element known from the *Y. lipolytica* genome. However, the genome sequencing project carried out by Genolevures consortium led to the identification of novel elements, those appear to be surprisingly diverse. Thus, non-LTR retrotransposon named Ylli was detected in *Y. lipolytica* genome [Casaregola *et al.*, 2000; Casaregola *et al.*, 2002]. It is only the second example of LINE-like elements found in the

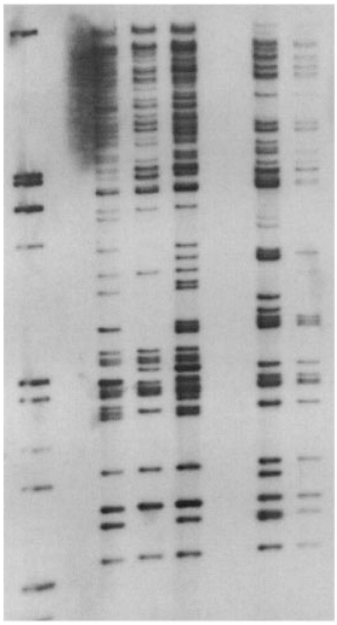

Figure 1.4 Presense of retrotransposon Ylt1 in different strains of *Y. lipolytica*. Southern blot of genomic DNA digested with *Eco*RI and probed with LTR fragment. Molecular weight marker λ/*Eco*RI-*Hind*III (lane 1); *Y. lipolytica* strains tested were H222 (lane 2), B512-3 (lane 3), B204-12A-213 (lane 4), B204-12C (lane 5), PO1d (lane 6), E150 (lane 7), E129 (lane 8). The illustration is taken from Juretzek *et al.*, 2001.

genomes of hemiascomycetous yeasts. Previously they were identified only in *C. albicans*. Interestingly, elements of *C. albicans* and *Y. lipolytica* both belong to the L1 group of non-LTR retrotransposons and seem to be closely related. Unlike Ylt1, Ylli was detected in all tested *Y. lipolytica* strains.

Fragments of a novel Ty3/*gypsy*-like element named Tyl3 was identified in the course of *Y. lipolytica* genome sequencing, too [Neuveglise *et al.*, 2002]. It appears to be related rather to Ty3 element of *S. cerevisiae* than to Ylt1. Unfortunately, no full-length element could be identified so far.

Finally, a DNA transposon belonging to the *Mutator* superfamily (so-called MULE, or *Mutator*-like element) was detected in *Y. lipolytica* [Casaregola *et al.*, 2003]. Members of this family were previously reported predominantly from plants and only recently a related element (*Hop*) was discovered in the genome of phytopathogenic fungus *Fusarium oxysporum* [Chalvet *et al.*, 2003]. No DNA transposons are currently known from the genomes of other hemiascomycetous yeasts.

The set of transposable elements identified in *Y. lipolytica* genome is unique for ascomycetous yeasts. It includes the only known yeast DNA transposon, LINE-like element similar to *C. albicans* one, and at least two different Ty3/*gypsy*-like LTR retrotransposons.

1.5.4 Structural properties of Ylt1 element

Ylt1 was a first retroelement described from *Y. lipolytica* genome. Different aspects of its biology were investigated in the last few years. Ylt1 was completely sequenced, and a comprehensive analysis of its nucleotide sequence was performed [Senam, 2004]. LTR sequences of Ylt1 were used for the construction of a set of integrative vectors for genetic transformation of *Y. lipolytica* [Pignede *et al.*, 2000; Juretzek *et al.*, 2001]. Further on, a system for insertional mutagenesis based on Ylt1 LTR sequences was developed and successfully used for the creation of a set of random *Y. lipolytica* mutants [Mauersberger *et al.*, 2001]. So, a large set of knowledge on the biology of Ylt1 was accumulated during the last decade.

Analysis of the structural properties of Ylt1 indicated that Ylt1 differs significantly from the majority of known Ty3/*gypsy* retrotransposons (Fig. 1.5). Its most prominent features are the following. First of all, Ylt1 is currently the largest known fungal retrotransposons [Senam *et al.*, 2004]. It is 1.5- to 2-fold larger then the well-known elements of *S. cerevisiae* (Ty1 and Ty3) or *Sz. pombe* (Tf1). Remarkably, Ylt1 encodes the same set of proteins as its smaller counterparts, and no additional coding capacities were detected in the course of its sequence analysis. The biological significance of the larger size of Ylt1 remains unclear.

Another uncommon feature of Ylt1 is an organization of its coding sequences. Both *gag* and *pol* genes of Ylt1 are arranged in a single reading frame [Senam, 2004] (Fig. 1.5). Therefore, a translation of Ylt1 mRNA should result in the production of equimolar amounts of Gag and Pol proteins. However, a molar excess of Gag protein is generally required for the correct assembly of virus-like particles and for the following replication of LTR retrotransposons. Ylt1 is thus faced with an additional problem of regulation of its Gag:Pol ratio. It was demonstrated that Tf1 element of *Sz. pombe* solves the same problem by regulated degradation of its Gag and Pol

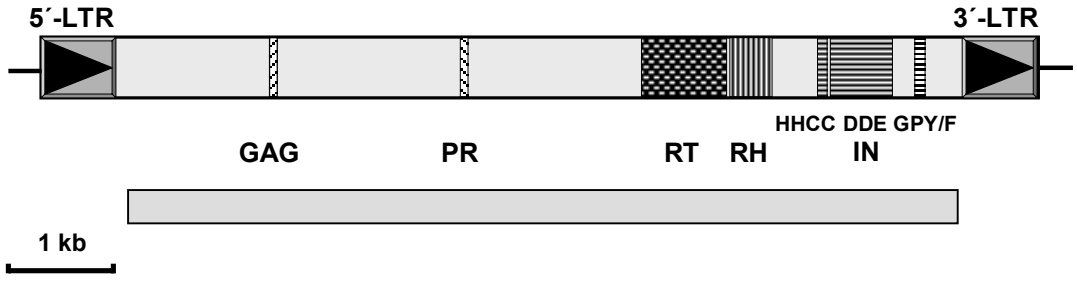

Figure 1.5 Structure of the retrotransposon Ylt1. LTR are shown as boxes with black triangles. Striped boxes indicate the localization of conserved motifs for GAG protein (Zn-fingers), protease (PR), reverse transcriptase (RT), RNaseH (RH) and three domains of integrase (IN): Zn-fingers (HHCC), core catalytic domain (DDE) and GPY/F domain. Rectangular represents schematically the single reading frame of Ylt1 [Senam, 2004 and this work].

proteins [Atwood *et al.*, 1996]. Its Gag and Pol proteins are also produced in equimolar amounts, and this ratio remains unchanged in log-phase cells. However, Pol protein is degraded preferentially upon transition to the stationary growth phase, resulting in the formation of the molar excess of Gag. The mechanism used by Ylt1 to regulate its Gag:Pol ratio has not been investigated so far, however, it is possible that this mechanism resembles the one used by Tf1.

Analysis of the primer-binding site (PBS) of Ylt1 suggests that $tRNA^{Ala}$ is used by Ylt1 as a primer for the synthesis of minus-strand cDNA [Senam, 2004]. Despite the diversity of primer tRNA molecules used by various LTR retrotransposons, primer-binding sites complementary to $tRNA^{Ala}$ are among the rarest ones. Tca13 from the genome of *C. albicans* is the only yeast retrotransposon, for which $tRNA^{Ala}$-complementary PBS was reported previously [Goodwin and Poulter, 2000]. This observation emphasizes an isolated position of Ylt1 among known yeast retrotransposons.

Phylogenetic analysis of Ylt1 relationships [Senam, 2004] again revealed connection between Ylt1 and *C. albicans* retrotransposons. This analysis placed Ylt1 together with *C. albicans* Tca3 element, and these both elements close to vertebrate retroviruses. However, some of the results obtained in the course of this analysis are quite controversial, so that further work in this field is necessary to clarify the phylogenetic position of Ylt1.

Due to its abundance, Ylt1 may have a significant impact on the genome structure and gene expression of its host. Indeed, it was identified initially in the *GPR1-1* mutant strain of *Y. lipolytica*. It turned out that the insertion of Ylt1 element in the 5′ upstream region of the *GPR1* gene causes a significant decrease in *GPR1* expression [Augstein *et al.*, 2003]. This examples shows that Ylt1 can indeed influence the expression of targeted genes. Further on, it was demonstrated that Ylt1 distribution in *Y. lipolytica* genome may change during the cell growth on two-carbon compounds [Schmid-Berger et al., 1994; Senam, 2004]. However, it is not clear whether these changes were caused by true retrotransposition of Ylt1 or by formation of solo LTRs as a result of homologous recombination within the retrotransposon.

Promoter activity of Ylt1 LTR was determined using β-galactosidase as a reporter enzyme [Senam, 2004]. These experiments demonstrated that Ylt1 LTR is a relatively weak promoter; however, its activity was increased during the incubation on acetate and ethanol.

Despite a large amount of data on Ylt1 biology accumulated since its discovery, many aspects of its life cycle remain enigmatic. Therefore, characterization of Ylt1 should have been continued in the course of this work.

1.6 Goals of the thesis

The aim of this work was a further characterization of the retrotransposon Ylt1. In the course of the work, the main interest was focused on the analysis of Ylt1 expression, on the detection of Ylt1-produced proteins and on the study of Ylt1 transposition *in vivo*. Further on, the Ylt1 distribution in the genome of *Y. lipolytica* was analyzed, and a search for Ylt1-related elements was performed. Therefore, following tasks for this work were set:

1. Tagging of proteins encoded by Ylt1 (Gag and IN) with the HA epitope and confirmation of its presence in *Y. lipolytica* cells by Western analysis.
2. Determination of sizes of individual proteins encoded by Ylt1.
3. Study of the Ylt1 expression and its regulation.
4. Development of an assay allowing the detection of Ylt1 retrotransposition events *in vivo*, and an analysis of the transpositional activity of Ylt1.
5. Investigation of the insertion specificity of Ylt1 and of the principles of its distribution in *Y. lipolytica* genome.
6. Search for Ylt1-related elements in *Y. lipolytica* genome and characterization of detected elements.

2 Materials and methods

2.1 Labor equipment

Centrifuges

Heraeus Biofuge fresco (microcentrifuge)

Heraeus Sepatech Biofuge 15R with rotor HFA 5.50

Sorvall RC 5C with rotors SS-34, SLA-1500 and SLA-3000

Incubators and shakers

Incubator Memmert BE500 (set at 37°C for the cultivation of *E. coli*)

Incubator Heraeus BK 600 (set at 28 °C for the cultivation of yeast strains)

Shaker Infors HT Novotron® AK82 (set at 37°C for the cultivation of *E. coli*)

Shaker Infors HT Multitron® (set at 28°C for the cultivation of yeast strains)

Sterilizers

Benchtop steam sterilizer KSG 112

Steam sterilizer H+P Varioklav 500E

Hot air sterilizer Memmert Model 500

Thermal cyclers

Biometra T1 Thermocycler

MWG-Biotech Primus 25

MWG-Biotech Primus 96 plus

Electroporators

Bio-Rad Gene Pulser II

Bio-Rad MicroPulser

DNA sequencers

Beckman Coulter CEQ™ 2000XL DNA Analysis System with CEQ Sequence Analysis 4.3.9 software

Power supplies

BioRad PowerPac 300 (used for DNA electrophoresis)

Biometra Standard Power Pack P25 (used for protein electrophoresis)

Pharmacia Biotech Electrophoresis Power Supply EPS 3500 (used for semidry protein transfer)

Electrophoresis systems

PeqLab Gel System Mini S (for horizontal DNA electrophoresis)

Hoefer HE 99X Max submarine unit (for horizontal DNA electrophoresis)

Phase PROT-RESOLV MINI-LC vertical gel electrophoresis system (for protein electrophoresis)

Electroblotters

PeqLab The Panther™ semi-dry electroblotter HEP-1

Spectrometers

Pharmacia Biotech Ultrospec 2000 (UV/visible spectrophotometer)

Kontron Instruments Uvikon 943

Balances

Sartorius MC1

Sartorius BP310P

Kern 770 (microbalance)

pH Meters

WTW pH 526

WTW pH 537

Further equipment

Reciprocal shaker Heidolph Promax 1020

Vibrating mixer Heidolph Reax 2000

Safety cabinet Heraeus HERAsafe HS9

Block-thermostat Kleinfeld Labortechnik BT100

MWG-Biotech UV transilluminator TEX-35M (312 nm) (with Biophotonics GelPrint 2000 software)

Savant Speed Vac® Concentrator DNA 110

Biometra OV5 hybridization oven

Biometra Vacu-Blot system

Techne Tempcold refrigerated thermostat

2.2 Chemicals and reagents

All chemicals and reagents used in this work are commercially available.

2.2.1 Enzymes, PCR reagents and ladders

Restriction endonucleases	Fermentas
Comby*Zyme* DNA polymerase mix	Invitek
OptiPerform buffer III	Invitek
Pwo DNA polymerase	Roche
dNTPs	Fermentas
Thermosensitive alkaline phosphatase	Gibco BRL
T4 DNA ligase	Promega
Ribonuclease A	USB
Zymolyase 20T	ICN Biomedicals
Glusulase	NEN Life Science Products
GeneRuler 1kb DNA ladder	Fermentas
BenchMark prestained protein ladder	Invitrogen/Gibco BRL

2.2.2 Kits and related products

JETSTAR Plasmid Midiprep Kit	Genomed
JETQUICK Plasmid Miniprep Spin Kit	Genomed
JETQUICK Gel Extraction Spin Kit	Genomed
JETQUICK PCR Product Purification Spin Kit	Genomed
JETSORB Gel Extraction Kit	Genomed
CEQ DTCS Quick Start Kit	Beckman Coulter
Gene Images random prime labelling module	Amersham Pharmacia Biotech
Gene Images CDP-*Star* detection module	Amersham Pharmacia Biotech
ECL Plus Western blotting detection reagents	Amersham Pharmacia Biotech
GB-003 blotting paper	Schleicher & Schüll
Immobilon-P PVDF Transfer Membrane	Millipore
Hybond-N nylon membrane	Amersham Pharmacia Biotech
Hyperfilm ECL X-ray film	Amersham Pharmacia Biotech

2.2.3 Antibodies

Antibodies used in this work were diluted according to supplier's instructions. Stock solutions were further diluted with block solution (see chapter 2.7.4) as described below before they were used for Western blots detection.

Primary monoclonal antibodies

Mouse anti-HA (Roche, clone 12CA5)	1:1000
Mouse anti-actin (ICN, clone C4)	1:1000

Secondary antibodies

Sheep anti-mouse Ig-HRP (Amersham Pharmacia Biotech)	1:5000

2.2.4 Nucleic acids

2.2.4.1 Acquired plasmids

Table 2.1 An overview of the plasmids used in this work

Plasmid	Features	Description	References
pUCBM21	*Amp*^{R}	Standard *E. coli* cloning vector	Roche Molecular Biochemicals, Mannheim
pINA237	*Amp*^{R} *Tet*^{R} *YlLEU2 ARS18*	Low copy *ARS*/*CEN Y. lipolytica*/*E.coli* shuttle vector	Fournier *et al.*, 1993
pINA302	*Amp*^{R} (*Yl*p*XPR2-ScSUC2*)	*Y. lipolytica* integrative vector containing *S. cerevisiae* fusion of *SUC2* gene with *Y. lipolytica XPR2* promoter	Barth and Gaillardin, 1996
pINA443	*Amp*^{R} *YlURA3 ARS68*	Low copy *ARS*/*CEN Y. lipolytica*/*E.coli* shuttle vector	Barth and Gaillardin, 1996
p67PT	*Amp*^{R} *YlURA3* (*Yl*p*ICL1-ICL1*t)	*Y. lipolytica* integrative vector containing regulatory regions (promoter and terminator) of *Y. lipolytica ICL1* gene	Juretzek, 1999
YEp351-3HA	*Amp*^{R} *ScLEU2* 2µ MCS/3xHA	*S. cerevisiae*/*E. coli* shuttle vector for the expression of HA-fusion proteins	R. J. Schweyen, Vienna
pUCETA	*Amp*^{R} *eta*	pUCBM21 vector containing internal part (*eta*) of Ylt1	Senam, 2004
pINAZA	*Amp*^{R} *YlLEU2* (LTRA-*lacZ*) *ARS18*	*Y. lipolytica* replicative vector allowing monitoring of LTR promoter activity	Senam, 2004

2.2.4.2 Constructed plasmids

Table 2.2 The plasmids constructed in the course of this work. Maps of the plasmids can be found in the Appendix.

Plasmid	Features	Description
pIET3	*Amp*^{R} *YlLEU2* *ARS18*	pINA237-based replicative *ARS*/*CEN* plasmid for the expression of 3xHA-tagged Gag protein of Ylt1 under the control of *Y. lipolytica ICL1* promoter. *ICL1* terminator is inserted downstream from *eta* region
pLEI3	*Amp*^{R} *YlLEU2* *ARS18*	pINA237-based replicative *ARS*/*CEN* plasmid for the expression of 3xHA-tagged Gag protein of Ylt1 under the control of native LTR promoter. *ICL1* terminator is inserted downstream from *eta* region
pUIN7	*Amp*^{R} *YlLEU2* *ARS18*	pINA237-based replicative *ARS*/*CEN* plasmid for the expression of 3xHA-tagged IN protein of Ylt1 under the control of native LTR promoter. *ICL1* terminator is inserted downstream from *eta* region
pUIN17	*Amp*^{R} *YlLEU2* *ARS18*	pINA237-based replicative *ARS*/*CEN* plasmid for the expression of 3xHA-tagged IN protein of Ylt1 under the control of *Y. lipolytica ICL1* promoter. *ICL1* terminator is inserted downstream from *eta* region
pYltS3	*Amp*^{R} *YlURA3* *ARS68*	pINA443-based replicative *ARS*/*CEN* plasmid for the expression of *ScSUC2*-marked Ylt1 element under the control of native LTR promoter.
pYltES3	*Amp*^{R} *YlURA3* *ARS68*	pINA443-based replicative *ARS*/*CEN* plasmid for the expression of *ScSUC2*-marked Ylt1 element under the control of *Y. lipolytica ICL1* promoter.

2.2.4.3 Synthetic oligonucleotides

Table 2.3 An alphabetical list of oligonucleotides used in this work. Recognition sites for restriction endonucleases are shown in bold. All oligonucleotides were purchased from MWG-Biotech.

No.	Name	Sequence (5´-3´)	Restriction endonuclease recognition site
1	GAGD2	CGT ACT AAC CCA GGC ATG TCG AAA GTA ACA AAA GAC G	-
2	GAGR1	ATA TA**C TTA AG**A CAG ATC TCG AGG CAC CCC	*Bsp*TI
3	HA-Nco1	ATA G**CC ATG G**TG AGG ACC CCT ACC CAT AC	*Nco*I
4	HA-Nco2	ATA T**CC ATG G**GT CTG CAG GGC AGC G	*Nco*I
5	ICPD3	ATA T**TC TAG A**TA GCT TGC TTC AAA CCA GAC G	*Xba*I
6	ICPR2	CGT CTT TTG TTA CTT TCG ACA TGC CTG GGT TAG TAC G	-
7	ICTD1	AAT T**GC TAG C**AT AT**C TTA AG**C TTT AAG CAG TTT GTT T	*Nhe*I, *Bsp*TI
8	ICTR1	TCC TGT **GGA TCC** GTA A	*Bam*HI
9	ICTR2	ATA T**GC GCG C**AT AT**A TCG AT**G ATC CGT AAA GTC ACG A	*Pau*I, *Bsu*15I
10	LED6	ATC ATA T**AT CGA T**TG TAA CAC TCG	*Bsu*15I
11	LEGD1	GCA CAA AGC ACT TTA TTT TCT CAC ATC TGG TGG ACG ACA CCT C	-
12	LER1	AAG AT**C TCG AG**G CAC C	*Xho*I
13	LER3	GAG GTG TCG TCC ACC AGA TGT GAG AAA ATA AAG TGC TTT GTG C	-

No.	Name	Sequence (5´-3´)	Restriction endonuclease recognition site
14	LTR-F1	ATA **GGA TCC** CCA GTA GAC TGG TCG TCC GGC TGT CCT	*Bam*HI
15	LTR-F2	CAT GCC ATG TAG GAC AGA G	-
16	pHA1	ATC AT**G TCG AC**G AGG ACC CCT ACC CAT ACG ATG TTC CTG	*Sal*I
17	pHA2	ATC AA**G TCG AC**G GTC TGC AGG GCA GCG TAA TCT GGA ACG	*Sal*I
18	pMUT1	TGC TTC GGA GTC ATC GCA GTT G	-
19	pMUT2	TGC GAT GAC TCC GAA GCA GTC C	-
20	PPT1	ACT C**AC GCG T**TA GAC CTG GGC AGG TGT AAC ACT CGC	*Mlu*I
21	pSL1	ATA T**GC GGC CGC GTC GAC** TCT AGA GGA TCT C	*Not*I, *Sal*I
22	pSL2	GTG TTA CAC CTG ACG ACA GTT AGA GCA GCA AC	-
23	pSL3	ACT GTC GTC AGG TGT AAC ACT CGC TCT GGA G	-
24	pSL4	ATT C**CT TAA G**TG TGA GAA AAT AAA GTG CTT TGT GC	*Bsp*TI
25	Yip1	ATG TCT **GCA TGC** GAG TAG TGG TGA TCT CCT G	*Pae*I
26	Yip2	ATG TCT **GCA TGC** GCA CCT GAA TCG TTA CGC	*Pae*I
27	Yt1	GCA G**TC TAG A**TG TAA TGA TTC GGA GAC ACT C	*Xba*I
28	Yt2	TCA T**CC ATG G**TG TAA CAG ATC GGG CGT AC	*Nco*I
29	Yt3	ACA CGA AGC GAT TGG CAA G	
30	Yt4	CGG A**TC TAG A**GA CTG GTT AAC GCA AGT GG	*Xba*I
31	Yt5	CGA CGT GAG ATC ATG AAG C	-
32	Yt6	TTG CTC TTC CTC CAG ATG C	-
33	Yt7	GGC ATA TGA GTC CAC GAA G	-
34	Yt8	CTG TAA CGT CTG CCA TAG C	-
Sequencing primers			
35	cms1	AGC TGG CCG CGT AAA TG	-
36	Cmsr1	CCT ACA AAA CCC GCT TCA A	-
37	has1	AGG GCC TCA ACA AGA TCA C	-
38	Hasr1	CAC CGA GAA AAG AGC GCA	-
39	pECHseq1	GAT ACT GCC TCG GAC AC	-
40	pECHseq2	ACA GAC CCG TGT CAC TG	-
41	pECHseq3	TAG CTC ACT CAT TAG GCA C	-
42	pECHseq4	AAC CTC AGC AAG CGT GTC	-
43	pECHseq5	CGC TTG CGA ACG TCT C	-
44	pECHseq6	AGC TGC ATG AAG GCT G	-
45	pETAseq1	GGC TTG GTT GAT TCA TCA G	-
46	pETAseq10	TGA TAA CCA CTA CGT GAG C	-
47	pETAseq11	GTC GAT GAC GTT GTT AGA C	-
48	pETAseq2	GAG ATC ATC CTC GTT ACG	-
49	pETAseq3	CTC CGT TTG CGA CTG TG	-
50	pETAseq4	GAC GGT ACG CTT GGA C	-
51	pETAseq5	AAG ACT CGT CGC CAG AG	-
52	pETAseq6	ACA GCT ATG ACT CTG ACT C	-
53	pETAseq7	TCA CGG ACA TCA CGA CG	-
54	pETAseq8	GGT CCA GCG ATT GCT C	-
55	pETAseq9	CAT TGT CGT AAC GAG GAT G	-
56	pIEseq1	TCG CTC GTG TTG ATT CTG	-
57	pIGAseq3	TAC ATG TTA GTG GAT GAG GTG	-
58	pIGAseq4	GAG GCT CTT CTG AGA TGG	-

No.	Name	Sequence (5´-3´)	Restriction endonuclease recognition site
59	pIGAseq5	TGC AAG GCG ATT AAG TTG G	-
60	pINAseq1	TGG ATG CTG TAG GCA TAG	-
61	pINAseq2	ATA GGC GCC AGC AAC C	-
62	pINAseq3	TGA TGT CGG CGA TAT AGG	-
63	RP	CAG GAA ACA GCT ATG AC	-
64	SP	GTA AAA CGA CGG CCA GT	-

2.3 Microorganisms

2.3.1 *Escherichia coli*

Escherichia coli strain DH5aC (*gyrA*96 *recA*1 *relA*1 *endA*1 *thi*-1 *hsdR*17 ($r_k^- m_k^+$) *supE*44 *deoR* Δ(*lacZYA-argF*)*U169*[φ80dΔ(*lacZ*)*M15*]) (Gibco BRL) was used for the amplification of plasmid DNA.

2.3.2 *Yarrowia lipolytica*

Y. lipolytica strains used in the work are summarized in the table 4.

Table 2.4 *Y. lipolytica* strains used in this work.

Strain	Genotype	References
B204-12C	*MATA met6-1 spo1-1*	Kujau *et al.*, 1992
B204-12C-20	*MATA met6-1 leu2-20 spo1-1*	Kujau *et al.*, 1992
E129	*MATA leu2-270 lys11-23 ura3-302 xpr2-322 SUC2*	Barth and Gaillardin, 1996
E150	*MATB leu2-270 his1 ura3-302 xpr2-322 SUC2*	Barth and Gaillardin, 1996
H222	*MATA*, wild type	Barth and Gaillardin, 1996
H222-41	*MATA ura3-41*	Mauersberger *et al.*, 2001
PO1d	*MATA leu2-270 ura3-302 xpr2-322 SUC2*	Le Dall *et al.*, 1994
YB423-12	*MATA*, wild type	Wickerham *et al.*, 1969

2.4 Media

Media were prepared as described below. Agar (20 g/l) was added to solid media.

2.4.1 LB medium (Luria-Bertani medium) (Sambrook *et al.*, 1989)

Pepton	1 % (w/v)
Yeast extract	0.5 % (w/v)
NaCl	1 % (w/v)

2.4.2 SOC medium (Sambrook *et al.*, 1989)

Pancreatic pepton from casein	2 % (w/v)
Yeast extract	0.5 % (w/v)
NaCl	10 mM
$MgCl_2$	10 mM
KCl	2.5 mM
Glucose	20 mM

2.4.3 YPD medium

Yeast extract	1 % (w/v)
Pepton	2 % (w/v)
Glucose	2 % (w/v)

2.4.4 Minimal medium for *Y. lipolytica* (Reader medium)

Mineral salts	*Final concentration (g/l)*
$(NH_4)_2SO_4$	3.0
KH_2PO_4	1.0
$K_2HPO_4 \cdot 3H_2O$	0.16
$MgSO_4 \cdot 7H_2O$	0.70
NaCl	0.50
$Ca(NO_3)_2 \cdot 4H_2O$	0.40

Microelements	*Final concentration (mg/l)*
H_3BO_3	0.600
$CuSO_4 \cdot 5H_2O$	0.048
KI	0.120
$MnSO_4 \cdot 4H_2O$	0.480
$Na_2MoO_4 \cdot 2H_2O$	0.240
$ZnSO_4 \cdot 7H_2O$	0.480
$FeCl_3 \cdot 6H2O$	2.000

Vitamins	*Final concentration (mg/l)*
Thiamin hydrochloride	0.3

Supplements	*Final concentration (mg/l)*
Uracil (Ura)	20
Leucine (Leu)	60
Methionine (Met)	50

Carbon sources	*Final concentration*
Glucose (G)	1 % (w/v)
Sucrose (S)	1 % (w/v)
Ethanol (E)	1 % (v/v)
Glycerol (G)	1 % (v/v)
Sodium acetate (A)	0.4 % (w/v)

Minimal media were named according to the contained carbon source and supplements, for instance, MG-Ura is a minimal medium with glucose and uracil.
Phosphate buffer (pH 6.8) and pepton were added to the minimal medium with sucrose to induce *XPR2* promoter controlling invertase expression. Their final concentrations were 50 mM and 0.05 % (w/v), correspondingly.

2.5 Cultivation of microorganisms

2.5.1 Cultivation of *E. coli*

E. coli cells were grown either in liquid or on solid medium at 37°C. Liquid cultures were agitated continuously at 220 rpm. Ampicillin was added to media for the selection of bacterial transformants to the final concentration 100 µg/ml.

2.5.2 Cultivation of *Y. lipolytica*

Y. lipolytica strains were grown in complete or minimal media at 28°C. Liquid cultures were incubated on an orbital shaker with continuous agitation at 220 rpm. Transformant selection was performed on the corresponding minimal media. To obtain higher cell density, cells were pregrown either in YPD or in minimal medium with glucose (MG) for 24 h, harvested and washed once with minimal medium without carbon source. Finally, the main culture was inoculated with the pregrown cells to the initial OD_{600} 0.5-1.0.

2.6 Recombinant DNA techniques

Basic recombinant DNA manipulations were performed as described in Sambrook *et al.* (1989) and Ausubel *et al.* (1997)

2.6.1 Agarose gel electrophoresis of DNA

Gels containing 0.6-2.0 % (w/v) of agarose (Biozym) in 1x TAE buffer were used for the separation of DNA fragments. Gel percentage was determined depending on the size range of the fragments. Ethidium bromide was added to gels (except for the ones further used for Southern blotting) to the final concentration 0.5 µg/ml. Before the loading, DNA samples were mixed with 1/5 volume of 6x gel-loading buffer. GeneRuler 1kb DNA ladder (Fermentas) was used as a molecular weight standard. The electrophoresis was performed in 1x TAE buffer at 8-10 V/cm. Finally, ethidium bromide-DNA complexes were visualized under UV transilluminator (312 nm) due to their bright red-orange fluorescence.

50x TAE buffer:	242 g Tris base 57.1 ml glacial acetic acid 100 ml 0.5 M EDTA (pH 8.0)
Ethidium bromide stock solution:	10 mg/ml
6x gel-loading buffer:	10 mM Tris-HCl (pH 7.6) 60 % (v/v) glycerol 60 mM EDTA (pH 8.0) 0.03 % bromophenol blue

2.6.2 Digestion of DNA with restriction endonucleases

Analytical digestion of DNA samples was generally performed in a volume of 20 µl as follows: approximately 500 ng of DNA were mixed with 2U of an enzyme in a corresponding buffer and incubated at 37°C for 1.5-2 h. Digestion of preparative amounts of DNA (5-10 µg) was performed in a volume of 30-40 µl at 37°C overnight with 10-20 U of an enzyme. In the case of double digestion with two different endonucleases, a buffer providing a maximal activity for both enzymes was selected. If no single buffer satisfied the requirements of both enzymes, digestion was performed sequentially. Briefly, after the digestion of DNA with an enzyme requiring low-salt buffer, salt concentration in the reaction mix was increased and the second digestion was carried out. In all cases, HPLC-grade water (Roth) was used in the reactions. The reactions were stopped by adding of 1/5 volume of the gel-loading buffer or 0.5 M EDTA to a final concentration 20 mM.

2.6.3 Treatment of DNA fragments with alkaline phosphatase

Linearized vector DNA molecules were treated with thermosensitive bacterial alkaline phosphatase (TsAP; Gibco BRL) to remove their 5´ phosphate groups and, thus, to reduce their self-ligation rate. The reaction was performed according to manufacture's instructions. Namely, 1U of the enzyme was added directly to the endonuclease digest, and the reaction mix was incubated at 65°C for 30 min. Then the reaction was terminated by adding EDTA (pH 8.0) to the final concentration 20 mM. Finally, fragments of interest were purified through an agarose gel using Genomed JETQUICK Gel Extraction Spin Kit.

2.6.4 Amplification of DNA fragments with PCR

Amplification of DNA fragments was performed either with Combi*Zyme* DNA Polymerase Mix (Invitek) or with Pwo DNA polymerase (Roche). In both cases, 3´-5´ exonuclease (proofreading) activity was provided by used enzymes, minimizing mutation frequency during PCR. Reaction was carried out in the volume of 100 µl using following components:

- template DNA (15 -50 ng)
- 50 pmol of both downstream and upstream primers
- dNTPs (dATP, dCTP, dGTP and dTTP) in the final concentration 200 μM
- $MgCl_2$ in the final concentration 250 μM
- 10 μl 10x reaction buffer
- 2U of the enzyme

The following cycling conditions were generally used:

Initial denaturation		5 min at 94°C
5 cycles	Denaturation	94°C for 1 min
	Annealing	T_m-6°C for 1 min
	Elongation	72°C, 1 min for every kb
25 cycles	Denaturation	94°C for 1 min
	Annealing	T_m-3°C for 1 min
	Elongation	72°C, 1 min for every kb
Final elongation		7 min at 72°C

Melting temperature of the primers was determined from the following equation:

$$Tm = 69.3 + 0.42(G + C) - \frac{650}{n},$$

where T_m is a melting temperature (°C), $G+C$ – G+C content of the primer and n – primer length. The composition of reaction mix and cycling conditions were slightly modified depending on the expected product length and enzyme and template used.

Overlap extension PCR (Pogulis *et al.*, 1996) was used to produce fusions of fragments of interest and to introduce point mutations. In this case, two subfragments sharing a common short terminal sequence (16-19 bp) were first amplified and purified. Then both fragments and corresponding terminal primers were used in a PCR reaction resulting in the amplification of a fusion molecule of interest.

PCR products were purified using JETQUICK PCR Product Purification Spin Kit (Genomed) according to manufacture's instructions.

2.6.5 DNA extraction from agarose gel

DNA extraction from agarose gel was done using JETQUICK Gel Extraction Spin Kit or JETSORB Gel Extraction Kit (Genomed). Before extraction, DNA fragments were separated in agarose gel as described in the chapter 2.6.1, and the gel slices containing a fragment of interest

were cut off and further processed according to manufacture's manual. JETSORB Gel Extraction Kit was preferentially used for the purification of larger fragments (> 5 kb).

2.6.6 Ligation of DNA fragments with T4 DNA ligase

For the ligation reaction, vector and insert DNA were generally taken in the molar ratio 1:4 to 1:6. Amount of the used vector DNA varied from 50 to 75 ng. Ligation was carried out in 20 µl 1x ligation buffer at 16°C overnight or at 20°C for 4 h. Polyethylene glycol (PEG 4000) was added to the final concentration of 5 % when blunt-ended fragments were ligated. The ligase was inactivated at 70°C for 10 min before bacterial cells were transformed with ligation mixture.

2.6.7 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from *E. coli* cells grown overnight in the selective medium (LB with ampicillin). Routine plasmid isolation was carried out by alkaline lysis method as described by Sambrook *et al.* (1989). DNA samples obtained in this way were further used for analytical digestion and cloning procedures. DNA for sequencing experiments was isolated either with JETQUICK Plasmid Miniprep Spin Kit or JETSTAR Plasmid Midiprep Kit (Genomed). The last one was taken for large-scale DNA preparations (more than 100 µg of plasmid DNA).

2.6.8 Isolation of DNA from *Y. lipolytica*

2.6.8.1 Rapid small-scale DNA isolation with glass beads

Isolation of yeast DNA was performed according to the procedure described by Hoffman and Winston (1987) with minor modifications. Together with the preparation of yeast chromosomal DNA, this method was also used for the isolation of plasmid DNA from yeast transformants and following retransformation of *E. coli* cells.

Yeast cells were grown in 10 ml of medium until stationary phase. Then they were harvested and washed once with water. Cell pellet was resuspended in 200 µl of TEST buffer, then 200 µl of phenol-chloroform and 0.3 g of glass beads were added to cells, and they were subjected to vortex mixing for 3-4 min. Next, 200 µl of TE buffer were added to suspension, it was centrifuged for 5 min at 13000 rpm and aqueous phase was transferred to a new tube. The samples were once again treated with 400 µl of phenol-chloroform and afterwards phenol traces were extracted with 400 µl of chloroform. Finally, DNA was precipitated by 2 volumes of ethanol, dried in SpeedVac concentrator and resuspended in 40 µl of TE buffer containing 100 µg/ml RNase A.

TEST buffer	10 mM Tris-HCl, pH 8.0 1 mM EDTA 100 mM NaCl 2 % Triton X-100 1 % SDS

2.6.8.2 Isolation of yeast DNA by spheroplasts lysis.

This method was preferentially used for the preparation of larger amounts of high molecular weight yeast genomic DNA.

Yeast cells were pregrown in 10 ml of YPD medium for 24 h. Then 100 ml of fresh YPD medium were inoculated with 1 ml of this culture and incubated for the next 8-12 h. Cells were harvested, washed once with water and resuspended in 10 ml SPβ buffer. 250 µl of zymolyase 20T and 200 µl of glusulase were added to the cell suspension, and it was incubated at 37°C until at least 90 % of cells lost their cell wall (60-90 min). Afterwards, spheroplasts were spun down by centrifugation, washed once with 10 ml of SP buffer and resuspended in 10 ml of Tris-EDTA buffer. 1 ml of 10 % SDS was added to the suspension, and the lysate was incubated at 70°C for 30 min. Next, proteins were precipitated by the addition of 3 ml of 5M potassium acetate and following incubation on ice for 1 h. The probes were centrifuged for 10 min at maximal speed to remove the protein pellet, and the supernatant was transferred to a new tube. Nucleic acids were precipitated with 0.7 volume of isopropanol for 10 min at room temperature, centrifuged at maximal speed for 10 min, washed once with 70 % ice-cold ethanol and dissolved in 5 ml of TE buffer containing 100 µg/ml of RNase A. Afterwards, probes were incubated at 37°C for 30 min to degrade RNA molecules, and remaining proteins were extracted twice with phenol, once with phenol-chloroform-isoamyl alcohol and, finally, once with chloroform-isoamyl alcohol. Next, DNA was precipitated from aqueous phase by adding 0.5 ml of 3M sodium acetate and 5 ml of isopropanol. The probes were incubated at room temperature for 10 min and centrifuged at maximum speed for the following 10 min. DNA pellet was washed once with ice-cold 70 % ethanol, air-dried and dissolved in 500 µl of TE buffer.

SP buffer	1.2 M sorbitol 0.1 M potassium phosphate buffer, pH 6.5
SPβ buffer	20 mM β-mercaptoethanol in SP buffer
Zymolyase 20T	10 mg/ml in SP buffer
Tris-EDTA buffer	50 mM Tris-HCl pH 7.4 20 mM EDTA

2.6.9 Transformation of *E. coli* by electroporation

Competent *E. coli* cells were prepared as described by Dower *et al.* (1998). They were portioned in 40 µl aliquots and stored by -80°C.

Prior the transformation, aliquots were thawed on ice, mixed with 5 ng of plasmid DNA or 1.5 ml of ligation mix and transferred to the prechilled electroporation cuvette (0.1 cm, Peqlab). Then electric pulse of 1.8 kV was applied to cell suspension, and 960 µl of SOC medium were added to cells immediately. Finally, cells were incubated at 37°C for 1 h with continuous agitation, plated out on selective plates (LB with ampicillin) and incubated at 37°C overnight.

2.6.10 Transformation of *Y. lipolytica* by electroporation

The protocol proposed by Dower *et al.* (1988) with minor modifications was used for the transfromation of *Y. lipolytica* with autonomously replicating plasmids.

100 ml of YPD medium were inoculated with 1 ml of overnight culture of *Y. lipolytica* cells. The cells were further grown to the OD_{600} 1.3-1.5, chilled on ice for 10 min, harvested and washed once with 1 volume of ice-cold water, once with 0.5 volume of water, and once again with 0.1 volume of ice-cold 1 M sorbitol. All centrifugation steps were performed at 4°C in prechilled centrifuge tubes. Finally, cells were resuspended in 1 M sorbitol (final volume 500 µl) and portioned in 40 µl aliquots.

For the transformation, 40 µl of cell suspension were mixed with approximately 500 ng of plasmid DNA, placed on ice for 5 min and then transferred to the prechilled electroporation cuvette (0.2 cm, Peqlab). Cells were subjected to electric pulse (1.8 kV), resuspended in 1 ml of ice-cold 1 M sorbitol and plated out on corresponding selective medium. Transformant colonies appeared after 2-3 days of incubation at 28°C.

2.6.11 Southern blotting and hybridization

For Southern analysis, approximately 1 µg of genomic DNA was digested with restriction endonucleases overnight (for details see chapter 2.6.2) and resulting digest was separated in 0.8 % agarose gel without ethidium bromide as described in the chapter 2.6.1. Transfer of DNA from gel to Hybond-N nylon membrane (Amersham Pharmacia Biotech) was done using Vacu-Blot system (Biometra) following the manufacture's recommendations. After the transfer, the membrane was rinsed in 5x SSC and air-dried. Finally, DNA was fixed to the membrane by UV cross-linking under UV transilluminator TEX-35M (MWG Biotech; wave length 312 nm) for 3 min.

Probes for hybridization were prepared using *Gene Images* random prime labelling module (Amersham Pharmacia Biotech) exactly as described in the manufacture's manual. The next steps (pre-hybridization, hybridization and stringency washes) were done according to the recommendations of the same manual. Finally, *Gene Images* CDP-*Star* detection module (Amersham Pharmacia Biotech) was used for the membrane detection.

2.6.12 Construction of the plasmids

2.6.12.1 Construction of the plasmids for the expression of the HA-tagged Gag protein

Two sets of plasmids expressing HA-tagged Gag protein either under the control of the native LTR promoter or the strong inducible *ICL1* promoter of *Y. lipolytica* were created. In both cases *ICL1* terminator was used to prevent homologous recombination between two identical LTRs.

As only few suitable unique cloning sites are present in the large Ylt1 element, numerous subcloning steps were often required for the construction of the plasmids of interest.

The plasmid pUCETA was constructed by Senam (2004) and it contains internal part (*eta*) of Ylt1 element. First, *ICL1* terminator was amplified from p67PT plasmid using primers ICTD1 and ICTR2. Amplification product was treated with *Bsp*TI and *Pau*I endonucleases and inserted into pUCETA plasmid digested with *Bsp*TI and *Mlu*I to create **pIE11**. Further on, in-frame fusion of *ICL1* promoter and *gag* gene of Ylt1 was constructed by overlap PCR. Part of the *ICL1* promoter was amplified from p67PT plasmid with the primers ICPD3 and ICPR2, whereas 5´ region of *gag* gene was produced from pUCETA plasmid using the primers GAGD2 and GAGR1. The both fragments were joined in a final PCR reaction where terminal primers (ICPD3 and GAGD2) were used once again. The product was cloned into pUCBM21 vector digested with *Xba*I and *Bsp*TI and sequenced. Next, it was cut out with *Pae*I and *Xho*I endonucleases and cloned into pIE11 digested with the same enzymes giving **pIE21**.

It was decided to place the construct encoding for HA-tagged Gag protein on a plasmid replicating autonomously in *Y. lipolytica* cells. The shuttle vector pINA237 was chosen with this aim. Both *ICL1* promoter and *ICL1* terminator were cut from p67PT with *Nhe*I enzyme and inserted into *Nhe*I-digested pINA237 resulting in **pICL16** plasmid. This one and pIE21 were digested with *Bsu*15I and *Bgl*II (in the case of pIE21 partial digestion was performed) and corresponding fragments were joined to place *gag* gene of Ylt1 under the control of *ICL1* promoter (**pEICL3**).

As mentioned above, plasmid pUCETA contains solely internal part of Ylt1 element. To restore original LTR upstream from *gag* gene, following overlap PCR was carried out. LTR was

amplified from pINAZA plasmid [Senam, 2004] with primers LED6 and LER3 whereas 5′ terminal part of *eta* region was produced with primers LEGD1 and LER1 using pUCETA plasmid as a template. Both fragments were joined in a PCR reaction with the terminal primers LED6 and LER1. Resulted LTR-*eta* fusion product was treated with *Xho*I endonuclease and ligated with *Eco*32I- and *Xho*I-digested pUCETA plasmid giving **pECH3** construct.

To insert HA-tag, a *Hind*III fragment of pECH3 bearing 5′ terminal part of *gag* gene and characterized by the presence of unique *Sal*I recognition site was first subcloned in pUCBM22 vector (the corresponding construct was named **pUC-gag1**). **pUCBM22** is a direct derivative of pUCBM21 with partially deleted multiple cloning site, which was obtained by the self-ligation of the *Eco*32I and *Sma*I-digested vector. The sequence encoding for three tandemly repeated HA epitopes was amplified from YEP351-3HA plasmid with primers pHA1 and pHA2. Next, it was treated with *Sal*I and inserted in *Sal*I-cut pUC-gag1 giving **pUC-gag3**. Further on, 5′ terminal part of *gag* gene was cut out from pEICL3 with *Kpn*I and *Mun*I and cloned into *Kpn*I and *Eco*RI-digested pUCBM21 (**pICG1**), where its *Bsp*68I-*Xho*I fragment was replaced with the corresponding fragment from pUC-gag3 plasmid (**pICG3**). The obtained construct bears the fusion of *ICL1* promoter with the 5′ part of *gag* gene encoding for HA-tagged Gag protein. To restore the complete coding sequence of Ylt1, *Pae*I-*Xho*I fragment of pICG3 was used to replace the corresponding fragment in pUCETA, resulting in the creation of **pICE3**. Finally, the Ylt1 element expressing HA-tagged Gag protein under the control of *ICL1* promoter was placed on an autonomously replicating plasmid by replacing of *Kpn*I-*Bsp*TI fragment of pEICL3 plasmid with the *Kpn*I-*Bsp*TI fragment derived from pICE3. The obtained plasmid was named **pIET3**.

The plasmid expressing HA-tagged Gag protein under the control of native LTR promoter was constructed in the next way. First, *ICL1* terminator was amplified from p67PT with primers ICTD1 and ICTR1, treated with *Nhe*I and *Bam*HI and inserted in the replicative pINA237 vector also cut with *Nhe*I and *Bam*HI (**pINA-ter**). Then *eta* region of Ylt1 flanked by LTR and *ICL1*t was cut out from pECH3 with *Bsu*15I and *Bsp*TI and placed on pINA-ter digested with *Bsu*15I and *Bsp*TI as well. This construct was named **pEICL11**. At the same time, **pLE3** plasmid was created by joining of the portion of *gag* gene bearing HA-encoding sequence with the remaining part of the *eta* region of Ylt1. This was done by the cloning of *Bsu*15I-*Xho*I fragment of pUC-gag3 plasmid into the pECH3 digested with *Bsu*15I and *Xho*I. Finally, *Bsu*15I-*Bsp*TI fragment of pEICL11 was replaced with the corresponding fragment of pLE3 resulting in the creation of **pLEI3**. The last one is an autonomously replicating *Y. lipolytica* plasmid containing Ylt1 element expressing HA-tagged Gag protein under the control of the native LTR promoter.

2.6.12.2 Construction of the plasmids for the expression of the HA-tagged integrase

The next set of plasmids used in this work encodes for HA-tagged integrase (IN) of Ylt1. Again, it was expressed under the control of the native LTR promoter and inducible *ICL1* promoter.

First of all, the derivative of pUCBM21 lacking *Nco*I recognition site was constructed. With this aim, pUCBM21 was cleaved with *Nco*I, recessed 3´ termini ends of the produced fragments were filled in with the Klenow fragment of *E. coli* DNA polymerase I, and linear DNA molecules created in this way were recircularized with T4 DNA ligase. Obtained plasmids were named **pUCBM23**. Further on, *Bsp*68I-*Bsp*TI fragment from pUCETA plasmid was cloned in the pUCBM23 digested with *Eco*32I and *Bsp*TI giving **pUIN1**. The sequence encoding for the three tandemly repeated HA epitopes was amplified from pUC-gag3 using primers HA-Nco1 and HA-Nco2. Obtained fragment was treated with *Nco*I and inserted in the *Nco*I site of pUIN1 plasmid producing **pUIN3**. To restore the complete coding sequence of the Ylt1, *Bsp*120I-*Bsp*TI fragment of pECH3 was replaced with the corresponding sequence from pUIN3, resulting in the creation of **pUIN5**. Finally, to place the whole Ylt1 element expressing tagged integrase on a replicative plasmid, *Bsu*15I-*Bsp*TI fragment of pUIN5 was joined with pEICL11, also cleaved with *Bsu*15I and *Bsp*TI. The final construct received a name **pUIN7**. It expresses HA-tagged integrase under the control of native LTR promoter.

On the other hand, a plasmid bearing HA-tagged integrase expressed under the control of *ICL1* promoter was created. First, *ICL1* promoter was cut out from pICG1 with *Xba*I and *Xho*I and produced fragment was inserted in the pUIN5 digested in the same way. The product named **pUIN15** was further cleaved with *Xba*I and *Bsp*TI. The obtained fragment was inserted in the *Nhe*I- and *Bsp*TI-cleaved pINA-ter. The produced final construct was named **pUIN17**.

2.6.12.3 Construction of the plasmids bearing *ScSUC2*-marked Ylt1 element

A set of plasmids containing Ylt1 element marked with *SUC2* gene of *S. cerevisiae* were constructed in the course of this work. These plasmids were created to monitor the transpositional activity of Ylt1 retrotransposon *in vivo*.

The *S. cerevisiae SUC2* gene expressed under the control of *Y. lipolytica XPR2* promoter [Nicaud *et al.*, 1989] was amplified from pINA302 plasmid [Barth and Gaillardin, 1996] with primers pSL3 and pSL4. Next, LTR of Ylt1 was produced from pECH3 plasmid with primers pSL1 and pSL2, and both fragments (*XPR2*p-*SUC2* and LTR) were joined together through the overlap PCR reaction with primers pSL1 and pSL4. Created fusion product was cloned in pUCBM21 vector digested with *Bsp*TI and *Not*I giving **pSUC-LTR** plasmid. The functionality of the amplified *SUC2* gene was tested directly in *Y. lipolytica* cells. With this aim, *Eco*RI-*Pae*I

fragment of pSUC-LTR was cloned in pINA443 vector. The obtained plasmid **pSUC21** was introduced in the *Y. lipolytica* strain H222-41 (Suc$^-$) and the ability of transformants to grow on glucose as a sole carbon source was confirmed.

At the same time, Ylt1 element was placed on an autonomously replicating plasmid with *URA3* marker gene. Namely, pECH3 was partially digested with *Bsu*15I and *Bam*HI and obtained fragment was inserted in the *Bsu*15I- and *Bam*HI-digested pINA443 vector. Produced plasmid was named **pLE11**. Then *SUC2* gene together with LTR was cut out from pSUC-LTR with *Mlu*I and *Not*I enzymes and inserted in the pLE11 cleaved with the same endonucleases resulting in the construction of **pYltS1**. Finally, polypurine tract (PPT) of Ylt1 was amplified from pSUC21 plasmid with primers PPT1 and pIGAseq5, treated with *Mlu*I and inserted in *Mlu*I-digested pYltS1 plasmid giving **pYltS3** construct. The last one is a plasmid replicating autonomously in *Y. lipolytica* cells and bearing completeYlt1 element marked with *S. cerevisiae SUC2* gene.

Further on, a similar construct with marked Ylt1 expressed under the control of *ICL1* promoter was created. First, regulatory sequences of *ICL1* gene (promoter and terminator) were cut out from p67PT with enzymes *Bam*HI and *Nhe*I and introduced in the pINA237 cleaved with the same endonucleases producing in this way plasmid **pICL3**. Next, transcribed portion of LTR was amplified from pUC-gag1 with primers LTR-F1 and LTR-F2, treated with *Bam*HI and *Bsp*68I enzymes and then inserted in the correspondingly cleaved pICL3 plasmid so that it replaced the transcribed part of *ICL1*gene. The obtained *ICL1*-LTR fusion was cut out from the plasmid **pILTR1** with *Bsu*15I and *Bsp*68I. Then it was used to replace the corresponding fragment of the pUC-gag1 producing **pILTR3**. Next, this fusion was joined with the full-length Ylt1 element by replacing of *Bsu*15I-*Xho*I fragment of pECH3 with the *Bsu*15I-*Xho*I fragment of pILTR3 giving **pILTR5**. Finally, this *ICL1*-regulated Ylt1 element was cut out from pILTR5 with *Bsu15*I and *Not*I and then cloned in the pYltS3 plasmid to be connected with *SUC2*. The final construct was named **pYltES3**.

2.6.13 DNA sequencing

All DNA fragments obtained by PCR were sequenced to ensure an absence of PCR-introduced mutations. Furthermore, certain regions of Ylt1 were resequenced in the course of this work to correct existed sequence errors or uncertainties. DNA sequencing was performed by the modified dideoxy-mediated chain termination method [Sanger *et al.*, 1977]. CEQ DTCS Quick Start Kit (Beckman Coulter) was used in sequencing reactions according to manufacture's instructions. Sequencing itself was done with CEQ™ 2000XL DNA Analysis System (Beckman Coulter).

2.7 Protein techniques

2.7.1 Preparation of cell-free extracts by cell disruption using glass beads

Cell-free extracts were prepared from yeast cultures at OD_{600} between 1 and 5. Cells were harvested by centrifugation at 3,000 rpm for 5 min, washed once with 1 volume of the ice-cold potassium phosphate buffer (pH 7.5) and resuspended in 300 µl of the ice-cold phosphate buffer with Complete Protease Inhibitor Cocktail (Roche). All subsequent steps were performed at 4°C. The equal volume of acid-washed glass beads (0.25-0.5 mm, Roth) was added to the cells, and they were vortexed 4-5 times at maximum speed for 1 min. The tubes were chilled on ice for 1 min in between the vortexing steps. Then the supernatant was collected, and the beads were washed once with phosphate buffer containing Complete Protease Inhibitor Cocktail. Again, the supernatant was collected. Both supernatants were pooled together and centrifuged at 3,000 x g for 10 min, and the supernatant (the crude extract) was used in further experiments.

2.7.2 Determination of protein concentration by the Lowry method

This method was first proposed by Lowry *et al.* [1951].

To determine the protein concentration, the samples were diluted 50-fold and 100-fold with 0.1 M NaOH. 1 ml of solution D was added to 0.2 ml of the diluted probe; the tube was mixed and allowed to stand for 10 min at room temperature. Then 0.1 ml of solution E was added, the probe was immediately vigorously mixed and allowed to stand for the next 30 min at room temperature. The probe extinction was measured at λ=720 nm against empty probe (0.1 M NaOH instead of protein sample). Finally, the protein concentration was determined by the comparison of the extinction values with the ones of the standard curve obtained using BSA (Merck) dissolved in 0.1 M NaOH (0.01-0.4 mg/ml).

Solution A	2 % (w/v) Na_2CO_3 in 0.1 M NaOH
Solution B	1 % (w/v) $CuSO_4 \cdot 5H_2O$
Solution C	2 % (w/v) Na-K-tartrate
Solution D	49 ml of solution A mixed with 0.5 ml of solution B and 0.5 ml of solution C
Solution E	1N Folin-Ciocalteu reagent

2.7.3 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Electrophoretic separation of proteins was carried out under denaturing conditions (i.e., in the presence of SDS) in discontinuous gel system [Laemmli, 1970]. Electrophoresis was performed in PROT-RESOLV MINI-LC vertical gel electrophoresis system (Phase). The concentration of separating gels varied from 7 to 10 % depending on the size of proteins of interest, whereas

5 % stacking gel was used in all experiments. The recipes for commonly used 8 % separating gel and 5 % stacking gel are given in the Table 5.

Table 2.5 Composition of polyacrylamide separating and stacking gels.

	8 % separating gel (10 ml)	5 % stacking gel (5 ml)
H_2O	4.6 ml	3.40 ml
Rotiphorese gel 30 (Roth)	2.7 ml	0.83 ml
1.5 M Tris-HCl pH 8.8	2.5 ml	-
0.5 M Tris-HCl pH 6.8	-	0.63 ml
10 % SDS	100 µl	50 µl
TEMED	5 µl	5 µl
10 % APS	100 µl	50 µl

First, the separation gel was poured and layered with distilled water. It was allowed to polymerize for 30 min at room temperature. Then water was removed, the stacking gel was poured, and a comb was inserted into the layer of stacking gel solution. Again, it was left for 30 min to polymerize. Afterwards, the gel was placed in the electrophoresis apparatus, it was filled with 1x electrophoresis buffer, and the comb was removed carefully.

Crude extracts prepared as described in the chapter 2.7.1 were mixed with 1/5 volume of 6x sample buffer and heated at 95°C for 5 min prior the electrophoresis. The probes' volumes containing 25-50 µg of total protein were loaded on the gel. The electrophoresis was performed at the current of 25-30 mA for 2-3 h.

Electrophoresis buffer pH 8.3 (10x)
0.25 M Tris
1.92 M glycine
1 % SDS

Sample buffer (6x)
0.35 M Tris-HCl pH 6.8
30 % (v/v) glycerol
10 % (w/v) SDS
0.6 M DTT
0.012 % bromphenol blue

2.7.4 Western blotting

Proteins separated in SDS-polyacrylamide gel were transferred to membrane to enable the following immunodetection. PVDF membrane Immobilon P (Millipore) was used in this work. The protein transfer was done using so-called semidry blotting. Prior the blotting, membrane was immersed in methanol for 1 min, rinsed with distilled water and then equilibrated in transfer buffer for 15 min. Stacking gel was removed from separating gel and discarded. Separating gel

was also equilibrated in transfer buffer for 15 min. Afterwards the gel and membrane were placed between buffer-saturated blotting paper in the blotting apparatus. The Panther™ semidry electroblotter HEP-1 (PeqLab) was used in this work. The transfer was performed at 0.8 mA/cm^2 for 1-2 h. Then the membrane was stained with 1x Ponceau S solution for 5 min and destained in water for the next 2 min to visualize protein bands. Finally, the membrane was left in blocking buffer (see chapter 2.7.5) overnight.

Transfer buffer	25 mM Tris 192 mM glycine 5 % methanol 0.001 % SDS
10x Ponceau S solution	2 % (w/v) Ponceau S 30 % (w/v) trichloracetic acid 30 % (w/v) 5-sulfosalicylic acid

2.7.5 Immunodetection of blotted proteins

Following the transfer of proteins from gel to membrane, the membrane was incubated in blocking buffer at 4°C overnight. Then it was washed briefly with washing buffer and incubated for 1 h with primary antibodies diluted in blocking buffer as indicated in the chapter 2.2.3. Next, the membrane was washed three times by agitating with washing buffer and incubated for the next 1 h with secondary antibodies diluted in blocking buffer. The membrane was again washed three times with washing buffer. Finally, the detection of membrane was performed using ECL Plus Western blotting detection reagents (Amersham Pharmacia Biotech) according to the manufacture's instructions.

Washing buffer	20 mM Tris-HCl pH 7.6 137 mM NaCl 0.1 % Tween 20
Blocking buffer	5 % nonfat dry milk in washing buffer

2.7.6 Gel staining with Coomassie blue

SDS-polyacrylamide gels were stained with Coomassie staining solution for 30 min. The destaining was done with destaining solution until blue bands and a clear background were obtained.

Coomassie staining solution	42 % (v/v) methanol 17 % (v/v) acetic acid 0.05 % Coomassie Brilliant Blue G-250
Destaining solution	5 % (v/v) methanol 5 % (v/v) acetic acid

2.8 Bioinformatics analysis

2.8.1 Sequence analysis

Functional protein domains in amino acid sequences were identified with ProfileScan program [Falquet *et al.*, 2002] and NCBI Conserved Domain (CD) search [Marchler-Bauer and Bryant, 2004]. Putative regulatory elements were predicted with Hamming Clustering method for signals prediction [Milanesi *et al.*, 1996] and AUG_EVALUATOR tool [Rogozin *et al.*, 2001] located at the WebGene site (http://www.itba.mi.cnr.it/webgene/) as well as with BCM search launcher tools for promoter prediction [Smith *et al.*, 1996; http://searchlauncher.bcm.tmc.edu]. Sequence alignments were obtained using ClustalX 1.81 [Thompson *et al.*, 1997] and then further adjusted with GeneDoc [Nicholas *et al.*, 1997]. Nucleotide-nucleotide (blastn), protein-protein (blastp), and translated (blastx and tblastn) BLAST searches [Altschul *et al.*, 1997] against the non-redundant database of the National Center for Biotechnology Information were performed to find the homologous sequences.

2.8.2 Phylogenetic analysis

For the analysis of phylogenetic relationship of *Y. lipolytica* retroelements with LTR retrotransposons, the most conserved residues of reverse transcriptase domain were selected [Xiong and Eickbush, 1990]. They were first aligned with ClustalX [Thompson *et al.*, 1997], and then phylogenetic trees were generated using the programs SEQBOOT, PROTDIST, NEIGHBOR and CONSENSE from program package PHYLIP3.57c [Felsenstein, 1993]. Obtained trees were visualized with TreeView, Version 1.5.2 [Page, 1996]. Accession numbers of the sequences used in the phylogenetic analysis are present in the table 6.

Table 2.6 The retroelements included in the phylogenetic analysis.

Retroelement	Organism	NCBI accession number	Reference
17.6	*Drosophila melanogaster*	X01472	Saigo *et al.*, 1984
297	*D. melanogaster*	X03431	Inouye *et al.*, 1986
412	*D. melanogaster*	X04132	Yuki *et al.*, 1986
Athila	*Arabidopsis thaliana*	X81801	Pelissier *et al.*, 1995
Blastopia	*D. melanogaster*	Z27119	Frommer *et al.*, 1994
BLV	Bovine leukemia virus	K02120	Sagata *et al.*, 1985
Cer1	*Caenorhabditis elegans*	U15406	Wilson *et al.*, 1994
CfT-1	*Cladosporium fulvum*	Z11866	McHale *et al.*, 1992
copia	*D. melanogaster*	X04456	Mount and Rubin, 1985
Cyclops	*Pisum sativum*	AJ000640	Chavanne *et al.*, 1998
dea1	*Ananas comosus*	Y12432	Thomson *et al.*, 1998
del1	*Lilium henryi*	1510387A	Smyth *et al.*, 1989
FeLV	Feline leukemia virus	AF052723	Chen *et al*., 1998
Gimli	*A. thaliana*	AB007727	Marin and Llorens, 2000
Grasshopper	*Magnaporthe grisea*	M77661	Dobinson *et al.*, 1993
Gypsy	*D. melanogaster*	M12927	Marlor *et al.*, 1986
HFV	Human foamy virus	Y07725	Schmidt *et al.*, unpubl
HIV-1	Human immunodeficiency virus 1	NC_001802	Martoglio *et al.*, 1997
Mag	*Bombyx mori*	X17219	Michaille *et al.*, 1990
Maggy	*M. grisea*	L35053	Farman *et al.*, 1996
mdg1	*D. melanogaster*	X59545	Avedisov *et al.*, 1990
mdg3	*D. melanogaster*	X95908	Dzhumgaliev *et al.*, 1986
Micropia	*D. melanogaster*	X14037	Lankenau *et al.*, 1988
Osvaldo	*Drosophila buzzatii*	AJ133521	Pantazidis *et al.*, 1999
Petra	*Hordeum vulgare*	Y14573	Panstruga *et al.*, 1998
Reina	*Zea mays*	U69258	Avramova *et al.*, 1996
RSV	Rous sarcoma virus	NC_001407	Schwarz *et al.*, 1983
Skipper	*Dictyostelium discoideum*	AF017040	Leng *et al.*, 1998
Skippy	*Fusarium oxysporum*	L34658	Anaya and Roncero, 1995
suchi-ichi	*Takifugu rubripes*	AF030881	Poulter and Butler, 1998
SURL	*Tripneustes gratilla*	M75723	Springer *et al.*, 1991
Tca3	*Candida albicans*	AF510498	Goodwin *et al.*, 2003
Tcd3	*Candida dubliniensis*	AF499464	Goodwin *et al.*, 2003
Tcn1	*Cryptococcus neoformans*	Retrobase[1]	Goodwin and Poulter, 2001
Tcn2	*C. neoformans*	Retrobase[1]	Goodwin and Poulter, 2001
Tcn3	*C. neoformans*	Retrobase[1]	Goodwin and Poulter, 2001
Tcn4	*C. neoformans*	Retrobase[1]	Goodwin and Poulter, 2001
Tf2	*Schizosaccharomyces pombe*	L10324	Weaver *et al.*, 1993
Tma1-1	*A. thaliana*	AC002534	Wright and Voytas, 1998
Tom	*Drosophila ananassae*	Z24451	Tanda *et al.*, 1994
Tse3	*Saccharomyces exiguus*	AJ439555	Neuveglise *et al.*, 2002
Ty3	*Saccharomyces cerevisiae*	M23367	Hansen *et al.*, 1988
Ty3-1p	*Saccharomyces paradoxus*	AY198186	Fingerman *et al.*, 2003
Ty4	*S. cerevisiae*	X67284	Janetzky and Lehle, 1992
Ulysses	*Drosophila virilis*	X56645	Evgen'ev *et al.*, 1992
Woot	*Tribolium castaneum*	U09586	Beeman *et al.*, 1996
yoyo	*Ceratitis capitata*	U60529	Zhou and Haymer, 1997

[1] The sequences of *Cryptococcus neoformans* retrotransposons are deposited in the Retrobase maintained by the University of Otago (http://biocadmin.otago.ac.nz/retrobase/home.htm)

3 Results

3.1 Detection of the Gag protein encoded by the retrotransposon Ylt1

The retrotransposon Ylt1 was described from the *Y. lipolytica* genome [Schmid-Berger *et al.*, 1994] more than ten years ago. The following sequence analysis [Senam, 2004] revealed several unusual properties of Ylt1. On the other hand, little has been known about the life cycle and the activity of Ylt1 in the host genome. The aim of this work was to clarify some of the unknown aspects of the biology of Ylt1.

The first part of the work was devoted to the study of proteins encoded by Ylt1. Performed sequence analysis showed that protein-coding sequences of Ylt1 are arranged in a single ORF and, thus, the corresponding proteins should be expressed as a single polyprotein and then processed by retroelement's protease. This conclusion raised several question. First, the recognition sequences of proteases of retroelements are characterized by a low degree of conservation. So, the size of the encoded proteins could not be estimated directly from the translated nucleotide sequence of Ylt1. Further on, Ylt1 is the largest of the known fungal retrotransposons, and it was interesting whether its unusual length is reflected in the size of encoded proteins. Finally, a molar excess of Gag protein is required for the normal replication of LTR retrotransposons, and it is unclear how the proper ratio is achieved in the case of Ylt1, which is proposed to express all its proteins as a single polyprotein, i.e., in equimolar amounts.

Structural Gag protein is known to be the most abundant among proteins expressed by LTR retrotransposons. So, it was decided to tag the Gag protein of Ylt1 with three tandemly repeated HA epitopes ($(HA)_3$). The fragment encoding for these epitopes was amplified from YEP351-3HA plasmid. Then it was introduced in the *Sal*I recognition site at the position 1069 (numeration in all cases is started from the 1st base of 5′ LTR of the full-length Ylt1) (Fig. 3.1). This position corresponds to the N-terminal region of the predicted Gag protein. The proteins of the most studied retroelements are only poorly expressed under normal conditions. Thus, the sequence encoding for the tagged Gag protein was placed under the control of the native LTR promoter (plasmid pLEI3) and the strong inducible *ICL1* promoter of *Y. lipolytica* (plasmid pIET3) (see also the chapter 2.6.11.1 for details) (Fig. 3.1). It was reported previously that some *Y. lipolytica* strains lack Ylt1 elements altogether. So, it was interesting to find out whether some differences in the Ylt1 expression exist between Ylt1-containing and Ylt1-lacking strains. Therefore, two *Y. lipolytica* strains, B204-12C-20 (bearing Ylt1) and PO1d (Ylt1-free), were transformed with the constructed plasmids.

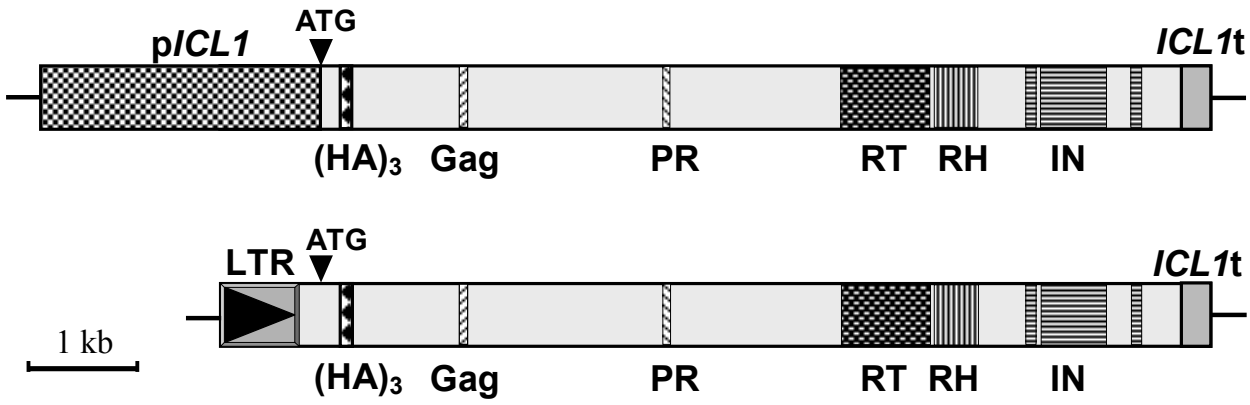

Figure 3.1 Tagging of the Gag protein of Ylt1 with the HA epitope. The constructs expressed under the control of *ICL1* promoter and of LTR promoter are shown. Positions of the start codon (ATG) and of the sequence encoding for the three tandemly repeated HA epitopes $(HA)_3$) are indicated.

Further on, it was examined whether the tagged Gag protein could be detected in *Y. lipolytica* transformants. First, the constructs bearing Gag-encoding sequence under the control of *ICL1* promoter were tested. *ICL1* promoter of *Y. lipolytica* is known to be induced among others by acetate and ethanol, reaching the maximum of its activity between 9 and 12 h after the cell transfer to the medium with corresponding carbon source. So, both B204-12C-20 and PO1d transformants containing autonomously replicating pIET3 plasmid (strains BI32 and PI31, correspondingly) were pre-cultured in minimal medium with glucose and then transferred to minimal medium with acetate as sole carbon source. It was reported that the activity of *YlICL1* promoter reaches its maximum in acetate-grown cells after 10 - 12 h of induction [Juretzek *et al.*, 2001]. Therefore, the samples were taken just before cell transfer and after 9 h and 12 h of growth in acetate medium. Crude cell-free extracts were prepared from the samples. Tagged Gag proteins were detected in the extracts by Western analysis using anti-HA antibodies.

Essentially the same results were obtained with Ylt1-containing (B204-12C-20) and Ylt1-lacking (PO1d) strains. In both cases the strong signals corresponding to the proteins with molecular weight of approximately 80 and 82 kDa were observed (Fig. 3.2). These values correspond well with the predicted size of the Gag protein, 83 kDa. Furthermore, additional high-molecular bands (about 140 and 170 kDa) were present in the samples taken before the cell transfer to the acetate medium. Notably, the second one (170 kDa) was less intensive in the B204-12C-20 strain. These data indicate that the HA-tagged Gag protein of Ylt1 can be successfully detected in *Y. lipolytica* cells. Several explanations of the multiple bands observed on the blots may be proposed. First, it is known that Gag proteins of retroviruses and LTR retrotransposons are processed by their own

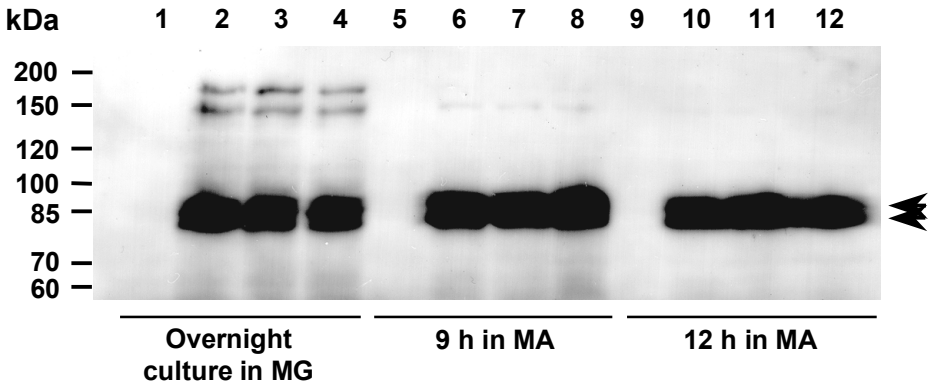

Figure 3.2 Expression of HA-tagged Gag protein of Ylt1 under the control of *ICL1* promoter. *Y. lipolytica* PO1d strain was transformed with the autonomously replicating plasmid pIET3. Three randomly selected transformants were pre-cultured in minimal medium with glucose (lanes 2-4) and then transferred to minimal medium with acetate. The probes were taken after 9h (lanes 6-8) and 12h (lanes 10-12) of acetate growth. The PO1d strain transformed with the vector pINA237 was used as a negative control (lanes 1, 5 and 9). Two main bands with apparent molecular weight of 80 and 82 kDa are indicated by arrows.

protease. Generally, CA (capsid) and NC (nucleocapsid) proteins arose in the course of Gag processing. Interestingly, in the case of the Ty1 element of *S. cerevisiae* the resulting C-terminal peptide has a size of just 4 kDa. It has never been observed directly and is currently believed to have no functional role. So, the presence of two forms of Gag protein may be explained either by proteolytic processing or by other post-translational modifications. Similarly, high-molecular bands of 140 and 170 kDa may correspond to the intermediate products of proteolytic processing of the whole Ylt1 polyprotein. The size of these peptides suggests that they consist of Gag protein and protease. Interestingly, these bands were not observed in cell extracts prepared after 9 h and 12 h of acetate growth. Their disappearance may be explained by the increase of the processing rate due to the accumulation of protease in the course of incubation.

Alternatively, the appearance of a number of additional bands on the blots may be due to the overexpression of Ylt1 proteins. The increased production of Ylt1 polyprotein may affect its processing resulting in the formation of the products of abnormal size. So, the next stage of the work was devoted to the study of Gag expression under the control of native LTR promoter.

Transformants obtained after the transformation of B204-12C-20 and PO1d strains with pLEI3 plasmid (BL31/BL32 and PL31/PL32, respectively) were analyzed. It was reported previously [Schmid-Berger *et al.*, 1994] that Ylt1 mobility can be observed in cells grown in acetate-containing medium. So, both strains were pre-cultured in minimal medium with glucose and then transferred to minimal medium with acetate. The samples were taken prior to cell transfer and after the 9 h and 12 h of acetate growth. The tagged Gag protein was detected in prepared cell-free extracts by Western analysis using anti-HA antibodies (Fig. 3.3).

The apparent size of the tagged Gag protein was the same as in the previous experiment. It was again estimated to be approximately 82 kDa. Some weaker bands were observed at 80 and 140 kDa as well (Fig. 3.3). Despite the general similarity, there were some important differences between the patterns of Gag protein expressed from *ICL1* and LTR promoters. First, the band with an apparent size of 80 kDa was as intensive as the 82 kDa band in the case of p*ICL1*-driven Gag expression, but it had much lower intensity and disappeared completely after 12 h of growth in acetate medium when LTR promoter was used. Next, the band corresponding to the putative Gag precursor with an apparent molecular weight of 170 kDa was weakly expressed comparing with the one at 140 kDa. Notably, both precursors completely disappeared after 9 h of growth in acetate medium. Further on, clear differences in the amount of Gag protein between Ylt1-containing and Ylt1-lacking strains were noted. Gag protein was more abundant in both tested transformants of B204-12C-20 strain than in the transformants of PO1d strain. It is unclear whether these differences are due to the different expression rate of the Gag protein or its different stability in tested strains. Regardless of these strain differences, an obvious increase of Gag protein amount in both strains was observed after the cells were transferred from glucose to

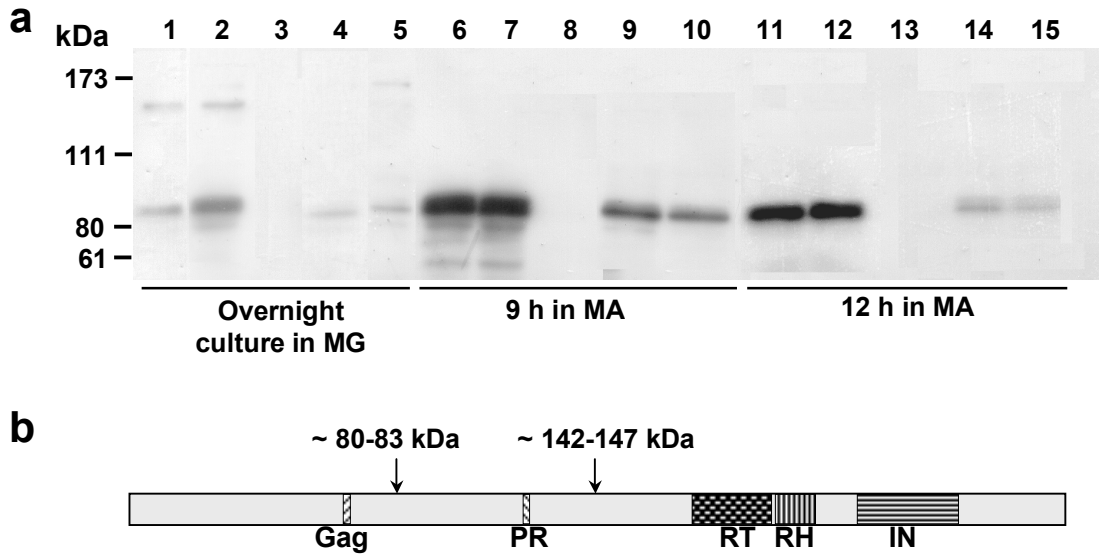

Figure 3.3 LTR-driven expression of HA-tagged Gag protein of Ylt1.

(a) *Y. lipolytica* strains B204-12C-20 (Ylt1-containing) and PO1d (Ylt1-free) were transformed with the autonomously replicating plasmid pLEI3. Two randomly selected transformants of each strain were pre-grown in minimal medium with glucose and then transferred to minimal medium with acetate. *Y. lipolytica* PO1d strain transformed with the vector pINA237 was used as a negative control. Lanes 1-2, 6-7 and 11-12: transformants B204-12C-20/pLEI3; lanes 3, 8 and 13: PO1d/pINA237; lanes 4-5, 9-10 and 14-15: PO1d/pLEI3.

(b) Scheme showing the estimated sizes of the proteins detected with anti-HA antibodies in the comparison with the single ORF of Ylt1. Conserved motifs for Gag protein, protease (PR), reverse transcriptase (RT), RNaseH (RH) and integrase (IN) are indicated.

acetate medium. These data demonstrate that LTR-driven expression occurs in cells grown in acetate medium.

The results of both experiments show that HA-tagged Gag protein can be detected in *Y. lipolytica* cells. Expression from both *ICL1* promoter and native LTR promoter results in the production of the main protein of approximately 82 kDa. Some putative precursors of Gag protein were observed as well. The HA-tagged Gag protein expressed under the control of the LTR promoter can be further used as a suitable tool to monitor its activity. On the other hand, an overexpression driven by the *ICL1* promoter may be useful for the production of larger amounts of Gag protein.

3.2 LTR-driven expression of Gag protein on different carbon sources

Y. lipolytica strains transformed with the autonomously replicating plasmid pLEI3 express HA-tagged Gag protein under the control of LTR promoter. The presence of the Gag protein during the growth in acetate medium was confirmed in the previous experiments. The next point of interest was the regulation of the LTR-driven expression of Gag protein in these transformants. Therefore, the presence of Gag protein during the growth on different carbon sources was studied by Western blot analyses. The experiments were performed as follows. Cells were pre-cultured in minimal medium with glucose, washed once with minimal medium without carbon source and then transferred to the medium with a carbon source of interest. The samples were taken just prior to the cell transfer and then after 3, 6, 9 and 24 h of incubation. Then crude cell-free extracts were prepared, and HA-tagged Gag protein was detected in Western blots with anti-HA antibodies. After the detection, the blots were stripped and processed with anti-actin antibodies to control the amount of loaded protein.

Notably, Gag protein was expressed during cell growth on all tested carbon sources, namely, glucose, glycerol, acetate and ethanol (Fig. 3.4). Corresponding bands were observed in all samples taken between 0 and 9 h of incubation. Interestingly, these bands disappeared from all samples except ethanol-grown cells after 24 h of incubation. The amount of produced Gag protein only slightly depends on the provided carbon source. So, some excess of Gag protein in glucose- and glycerol-growth cells was observed after 3 h and 6 h. These differences, however, were noticeably reduced after 9 h of growth. Next, the amount of Gag protein gradually increased in all cultures reaching its maximum after 9 h. The following decrease resulted in its disappearance after 24 h of growth. As mentioned before, the only exceptions were ethanol-grown cultures. In this case, the comparable amounts of Gag protein were observed after 9 and 24 h of incubation.

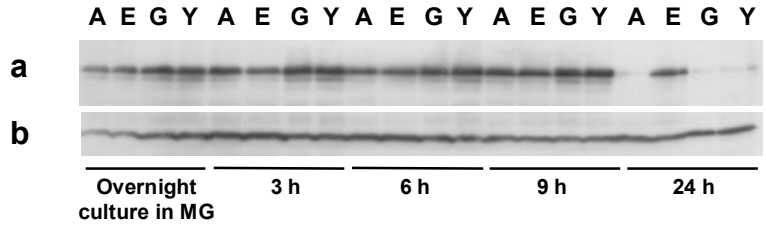

Figure 3.4 LTR-driven expression of Gag protein of Ylt1 on different carbon sources. Transformants of *Y. lipolytica* strain B204-12C-20 bearing pLEI3 plasmid were pre-grown in minimal medium with glucose and then transferred to the minimal media with acetate (A), ethanol (E), glucose (G) or glycerol (Y). The samples were taken after 3, 6, 9 and 24 h of growth.
(a) Detection of the HA-tagged Gag protein with anti-HA antibodies
(b) Detection of the actin with anti/actin antibodies (loading standard)

Taking into account the obtained results, it is suggested that expression of Ylt1 driven by its native LTR promoter takes place during cell growth on a variety of carbon sources. The kind of provided carbon source has only a limited effect on the amount of produced Gag protein. Further on, its level is proposed to be dependent rather on the growth phase of the culture than on a certain carbon source. Hence, an accumulation of Gag protein was observed in the first 9 h of growth, i.e. during the lag phase and logarithmic phase. Conversely, Gag protein was detected only in ethanol-grown cultures after 24 h of growth. Previous experiments showed that cells of this strain reach the stationary phase in glucose and glycerol minimal media after 18 h of growth, in acetate medium – after 20-21 h of growth, and in ethanol medium – after 27-30 h of growth. Therefore, it is suggested that the observed degradation of the produced Gag protein is caused by transition to stationary phase.

The effect of heavy metal ions on Gag expression was studied as well. The experiment was performed as follows. The cells were pre-cultured in minimal medium with glucose and transferred to the minimal medium with glucose containing 0.1 mM $CuSO_4$. The samples were taken after 4 h, 8 h and 10 h of incubation and analyzed as described before. Notably, a poor expression of the tagged Gag protein was observed compared to cultures grown in glucose medium without copper ions (Fig. 3.5). It was clearly detectable only after 10 h of growth.

Therefore, it is suggested that LTR promoter of Ylt1 is primarily active during the active cell growth, namely, in the logarithmic growth phase. It is also proposed that the availability of nutrients may play a certain role in its activation. The transition to stationary phase was accompanied by the decrease in the LTR promoter activity. Furthermore, presence of heavy metal ions had a negative effect on the LTR-driven expression.

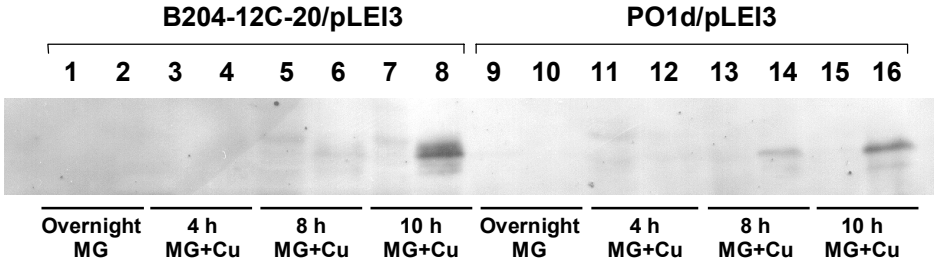

Figure 3.5 Effect of copper (II) ions on LTR-driven expression of HA-tagged Gag protein of Ylt1. Transformants of *Y. lipolytica* strains B204-12C-20 (Ylt1-containing) and PO1d (Ylt1-free) bearing the autonomously replicating plasmid pLEI3 were tested in this experiment. They were pre-grown in minimal medium with glucose and then transferred to the minimal glucose medium containing 0.1mM $CuSO_4$. Lanes 1-8: transformants B204-12C-20/pLEI3; lanes 9-16: PO1d/pLEI3.

3.3 Detection of the integrase encoded by the retrotransposon Ylt1

All coding sequences of the retrotransposon Ylt1 are arranged in a single ORF. Thus, the translation of the predicted mRNA should result in the production of Gag protein, protease, reverse transcriptase and integrase in equimolar amounts. At the same time, the molar excess of the structural Gag protein is required for normal replication of LTR retroelements. So, it was decided to investigate the expression of the integrase encoded by Ylt1.

The sequence coding for three tandemly repeated HA tags was introduced into the *NcoI* recognition site located at the position 7407 of Ylt1 (see chapter 2.6.11.2 for further details). The resulted construct was placed under the control of LTR and *ICL1* promoters. Plasmids containing corresponding constructions were named pUIN7 and pUIN17, respectively. One Ylt1-containing (B204-12C-20) and one Ylt1-lacking (PO1d) strain were transformed with these plasmids. The obtained transformants were used for the expression of the HA-tagged integrase.

First, the p*ICL1*-driven integrase expression was investigated. The transformants bearing pUIN17 plasmid were pre-grown in minimal medium with glucose and then transferred to the minimal medium with acetate to induce the *ICL1* promoter. The samples were taken after 9 h of growth. The detection of Western blots with anti-HA antibodies revealed the presence of a specific band with an apparent molecular weight about 140 kDa (Fig. 3.6). It is suggested that this band correspond to the polyprotein containing reverse transcripatse, RNaseH and integrase of Ylt1 with a predicted size of 136 kDa. Surprisingly, no specific bands with a size of integrase alone (55-58 kDa) were observed on the blots.

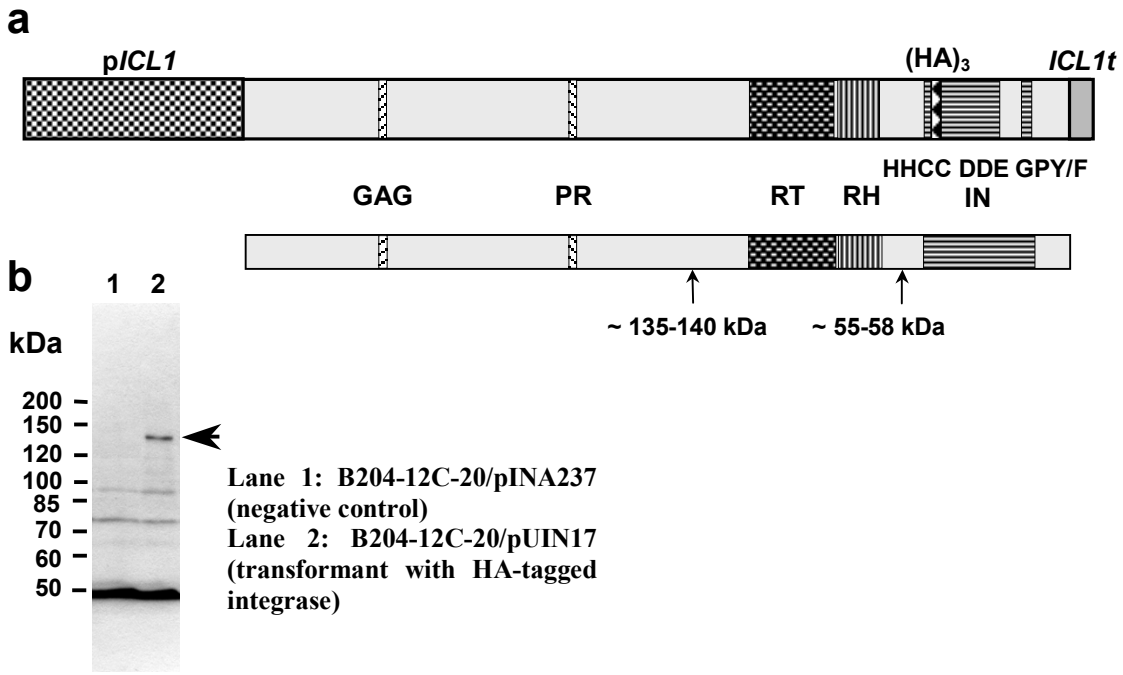

Figure 3.6 Detection of the HA-tagged integrase protein of Ylt1. *Y. lipolytica* B204-12C-20 strain transformed with the autonomously replicating plasmid pUIN17 was used in this experiment. The plasmid pUIN17 contains Ylt1 element with HA-tagged integrase expressed under the control of *ICL1* promoter (**a**). The B204-12C-20 strain transformed with the vector pINA237 was used as a negative control. Both strains were pre-cultured in minimal medium with glucose and then transferred to minimal medium with acetate to induce the *ICL1* promoter. The samples were taken after 9 h of cultivation and analyzed by Western blotting (**b**). Position of specific band is indicated by an arrow.

The transformants bearing pUIN7 plasmid were tested for the presence of tagged integrase, too. They were incubated in acetate and glycerol minimal media, and the expression of tagged integrase was again analyzed on Western blots with anti-HA antibodies. No specific bands were detected in these experiments.

So, the HA-tagged integrase was observed only in the cells expressing it under the control of strong *ICL1* promoter. No specific signals were detected in the case of LTR-driven expression. Both Gag- and integrase-encoding sequences are arranged in a single ORF, so their expression in equimolar amounts would be expected. The following explanations for the differences in Gag and integrase detection are proposed. First, the structural features of integrase may be responsible for the poor HA-epitope accessibility. The impact of this effect, however, should not be so significant because the proteins were separated in denaturing conditions. Second, as mentioned already, the molar excess of Gag protein is required for the normal assembly of virus-like particles. Although the mechanism of the regulation of the Gag:Pol ratio in the case of Ylt1 is currently unknown, it may involve either the poor expression or rapid degradation of the proteins located downstream from Gag. The existence of such mechanisms would explain the observed low content of Ylt1 integrase in *Y. lipolytica* cells.

3.4 Transposition of the marked Ylt1 element

Data suggesting the transpositional activity of Ylt1 in *Y. lipolytica* genome were collected previously [Schmid-Berger *et al.*, 1994; Senam, 2004]. A novel *in vivo* retrotransposition assay that should allow direct observation of Ylt1 transposition was developed in the course of this work. Its main components were the plasmids pYltS3 and pYltES3, whose construction is described in the chapter 2.6.11.3. These two plasmids bear an Ylt1 element marked with the *SUC2* gene of *S. cerevisiae*. It should be noted here that natural isolates of *Y. lipolytica* do not produce invertase and that they can not grow on sucrose as a sole carbon source. Thus, the *ScSUC2* gene can be used as a selective marker in *Y. lipolytica* strains. *ScSUC2* gene expressed under the control of *Y. lipolytica XPR2* promoter [Nicaud *et al.*, 1989] was used in this work. The expression of the marked Ylt1 element is driven by the LTR promoter in the case of pYltS3 and by the *Y. lipolytica ICL1* promoter in the case of pYltES3. So, the plasmid pYltES3 was used as a positive control in our experiments. Both plasmids contain ARS and centromere sequences of *Y. lipolytica* and they replicate autonomously in *Y. lipolytica* cells. Another important feature of both plasmids is the presence of *Y. lipolytica URA3* gene, which allows selection of yeast transformants of an *ura3*-deficient recipient strain.

The selection for clones with a newly transposed marked Ylt1 element was performed as follows. First, the constructed plasmids pYltS3 and pYltES3 were introduced in the *ura3*-mutated *Y. lipolytica* strain H222-41. Obtained transformants were named HY31 and HE31, respectively. Then the retrotransposition of the marked Ylt1 elements was induced by cultivation of both strains in minimal medium with acetate (MA) and minimal medium with glycerol (MY) at 24°C or at 28°C for 96 h. After the induction, the cells were either directly plated out on a selective medium or they were transferred to YPD medium to induce plasmid loss first. However, it turned out during the work that plasmid loss occured efficiently on selective medium with FOA, so no additional steps including an incubation in YPD medium were required.

A selective medium MSPPUF with sucrose as a sole carbon source was used in these experiments. The medium contained 50 mM phosphate buffer (pH 6.8) and 0.05 % peptone to induce *XPR2*-driven expression of *ScSUC2* gene. Further on, uracil and 5-fluoroorotic acid (FOA) were added to the medium at the concentration 15 mg/l and 1.5 g/l, respectively. Approximately 10^6 cells were plated on each plate, because higher cell densities on media with FOA result in high background growth [Barth and Gaillardin, 1996].

The MSPPUF medium was used to select for the clones with newly transposed *ScSUC2*-marked Ylt1 elements. Only cells with a Suc^+FOA^r phenotype should grow on this medium. It is known that the *ura3* mutation confers resistance to the FOA, so the *ura3* clones containing *ScSUC2*

gene are expected to grow on the MSPPUF medium. The initial transformants HY31 and HE31 care wild type copy of *URA3* gene on the pYltS3 or pYltES3 plasmid, respectively, so their growth on MSPPUF medium is inhibited by the fluoroorotic acid. The plasmid loss by these strains would result in an inability of cells to utilize sucrose. On the other hand, the retrotransposition of the marked Ylt1 element would lead to the transfer of *ScSUC2* gene to the nuclear genome. In this case, the plasmid loss would lead to the appearance of invertase-producing *ura3* clones. These clones, which exhibit Suc^+Ura^- phenotype, would survive on the MSPPUF medium.

There are some alternative mechanisms which would allow the cells to grow on the selective medium. For instance, spontaneous FOA-resistant clones would grow on this medium, too. So, the colonies grown on MSPPUF medium were proved for their Ura^- phenotype. They were picked up and transferred on plates with MSPPU medium (without FOA). Then they were replica-plated on MSPPU and MSPP (without uracil) plates. Only clones which were unable to grow on medium without uracil were chosen for further work.

The results of the selection for Suc^+Ura^- clones are shown in the table 7.

Table 3.1 The results of the selection for the clones with newly transposed marked Ylt1 element using *in vivo* retrotransposition assay.

Strain	Carbon source	Induction temperature	Number of Suc^+FOA^r clones	Number of Suc^+Ura^- clones	Frequency of Suc^+Ura^- clones
HY31	Acetate	24°C	55	42	$1.4 \cdot 10^{-5}$
		28°C	8	0	$< 3.3 \cdot 10^{-7}$
	Glycerol	24°C	20	6	$2.0 \cdot 10^{-6}$
		28°C	18	0	$< 3.3 \cdot 10^{-7}$
HE31	Acetate	24°C	95	81	$2.7 \cdot 10^{-5}$
		28°C	15	2	$6.7 \cdot 10^{-7}$
	Glycerol	24°C	12	4	$1.3 \cdot 10^{-6}$
		28°C	11	0	$< 3.3 \cdot 10^{-7}$

Striking differences were observed between cultures grown at 24°C and at 28°C during the induction phase. In contrast to 24°C, where 133 clones with the Suc^+Ura^- phenotype were obtained, only two clones exhibiting this phenotype were found when cells were cultivated at 28°C. Both derived from HE31 strain, where the expression of the marked Ylt1 element was

regulated by the strong *ICL1* promoter. Obtained data suggest that the transposition of Ylt1 is a temperature-dependent process and that the transposition rate dramatically increases at higher temperatures.

In the performed experiments, Suc^+Ura^- clones were obtained both from HY31 and HE31 strains. However, the number of HE31-derived clones is significantly higher than the number of HY31-derived ones. The strong *ICL1* promoter drives the expression of the marked Ylt1 element in HE31 strain. Thus, the higher number of HE31-derived Suc^+Ura^- clones may be explained by the higher level of the expression of marked Ylt1 element in HE31 strain. The dependence of the number of Suc^+Ura^- clones on the expression of the marked Ylt1 element is confirmed by the fact that the number of HE31-derived Suc^+Ura^- clones obtained after growth on acetate and on glycerol differs more than 20-fold. The observed difference reflects the induction of the *ICL1* promoter during the growth on acetate, whereas its activity on glycerol is repressed to a basal level.

The majority of the HY31-derived Suc^+Ura^- clones was obtained after the induction of transposition on acetate, too. These results agree with the previous observation suggesting the activation of Ylt1 during the growth on acetate [Schmid-Berger *et al.*, 1994; Senam, 2004].

Obtained Suc^+Ura^- clones were further analyzed to confirm that the acquisition of the Suc^+Ura^- phenotype was caused by the transposition of the marked Ylt1 element. A *Bsp*TI-*Pst*I fragment of the pSUC-LTR plasmid containing a part of the LTR of Ylt1 was used as a probe for the Southern analysis. In this analysis, the initial strain H222-41 was shown to be free from LTRs of Ylt1. At the same time, the presence of LTRs of Ylt1 in the genome of the strains HE31 and HY31 and the selected Suc^+Ura^- strains was confirmed in this experiment. Differences between the initial transformants (HY31 and HE31) and analyzed Suc^+Ura^- clones in the sizes of the bands corresponding to the LTR fragments suggest the different localization of the marked Ylt1 element, i.e., that it was transposed from the autonomous pYltS3/pYltES3 plasmids into the nuclear genome (Fig. 3.7).

Two Suc^+Ura^- clones obtained from HE31 strain after the incubation at 28°C and designed as E1-52 and E1-72 were analyzed in more details (Fig. 3.8). First of all, the Southern analysis with an *URA3*-containing *Sal*I fragment of the pINA443 plasmid as a probe confirmed that both strains have lost the plasmid pYltES3. Further on, the presence of Ylt1 element in the genome of the strains was confirmed by the hybridization with the fragments of LTR and *gag* region of Ylt1. A *Bsp*TI-*Pst*I fragment from the plasmid pSUC-LTR was used as a LTR probe, and a *Hinc*II fragment from the plasmid pUC-gag1 was taken as a *gag* probe. Surprisingly, the Southern analysis showed the presence of two *gag* fragments and two LTR fragments in the genome of

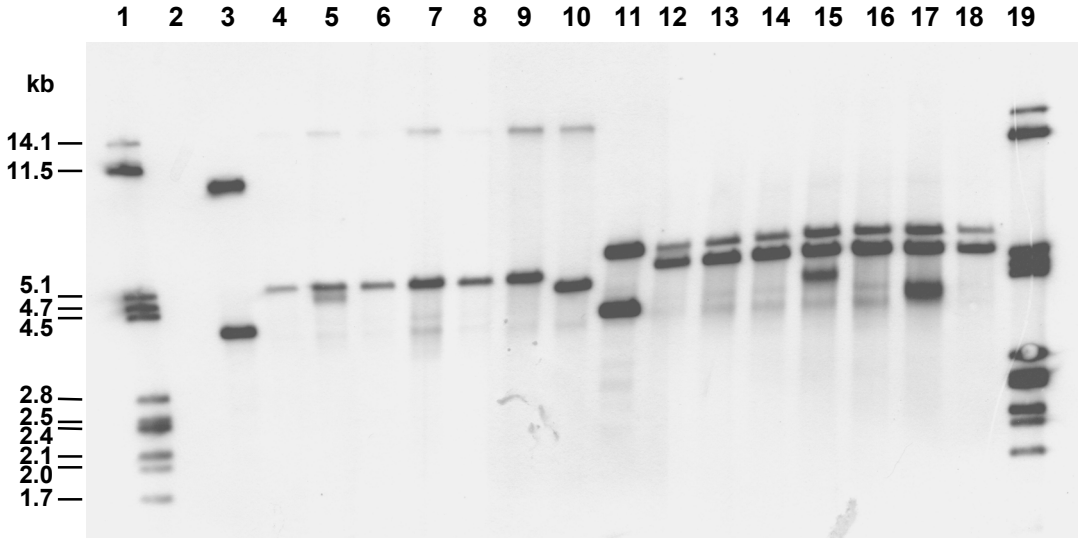

Figure 3.7 Southern analysis of $Suc^{+}Ura^{-}$ clones obtained in the *in vivo* retrotransposition assay. Genomic DNA was isolated from *Y. lipolytica* strain H222-41 (lane 2), its transformants HY31 (lane 3) and HE31 (lane 11) and from 14 randomly selected $Suc^{+}Ura^{-}$ clones. Seven of these clones (lanes 4-10) were obtained from the strain HY31 and remaining ones (lanes 12-18) – from the strain HE31. DNA samples were digested with the endonuclease *Nco*I. A *Bsp*TI-*Pst*I fragment of the plasmid pSUC-LTR was used as a LTR probe in a following Southern analysis. DNA of phage λ digested with the endonuclease *Pst*I was used as a molecular weight marker (lanes 1 and 19):

both strains, whereas the 2:1 ratio of LTR and *gag* fragments was expected. Therefore, it was decided to amplify the insertion site of the marked Ylt1 element by the "reverse PCR" technique. The genomic DNA isolated from both strains was digested with the restriction endonuclease *Pau*I, which has no recognition sites inside the marked Ylt1 element. The digest was treated with T4 DNA ligase to circularize the fragments of genomic DNA. Then both the untreated digest and the digest treated with DNA ligase were used as a template in a PCR reaction. The divergent primers (Yip1 and Yip2) were used in the PCR reaction, so the amplification of a specific product only after the circularization of the genomic DNA was expected. However, the products of the same size were amplified from both the untreated and ligase-treated digests. The sequencing of these products showed that the insertions of marked Ylt1 element in both strains were arranged in tandem repeats separated by a single LTR. It is suggested that this abnormal integration was caused either by the increased induction temperature or by the overexpression of the marked Ylt1 element from the strong *ICL1* promoter.

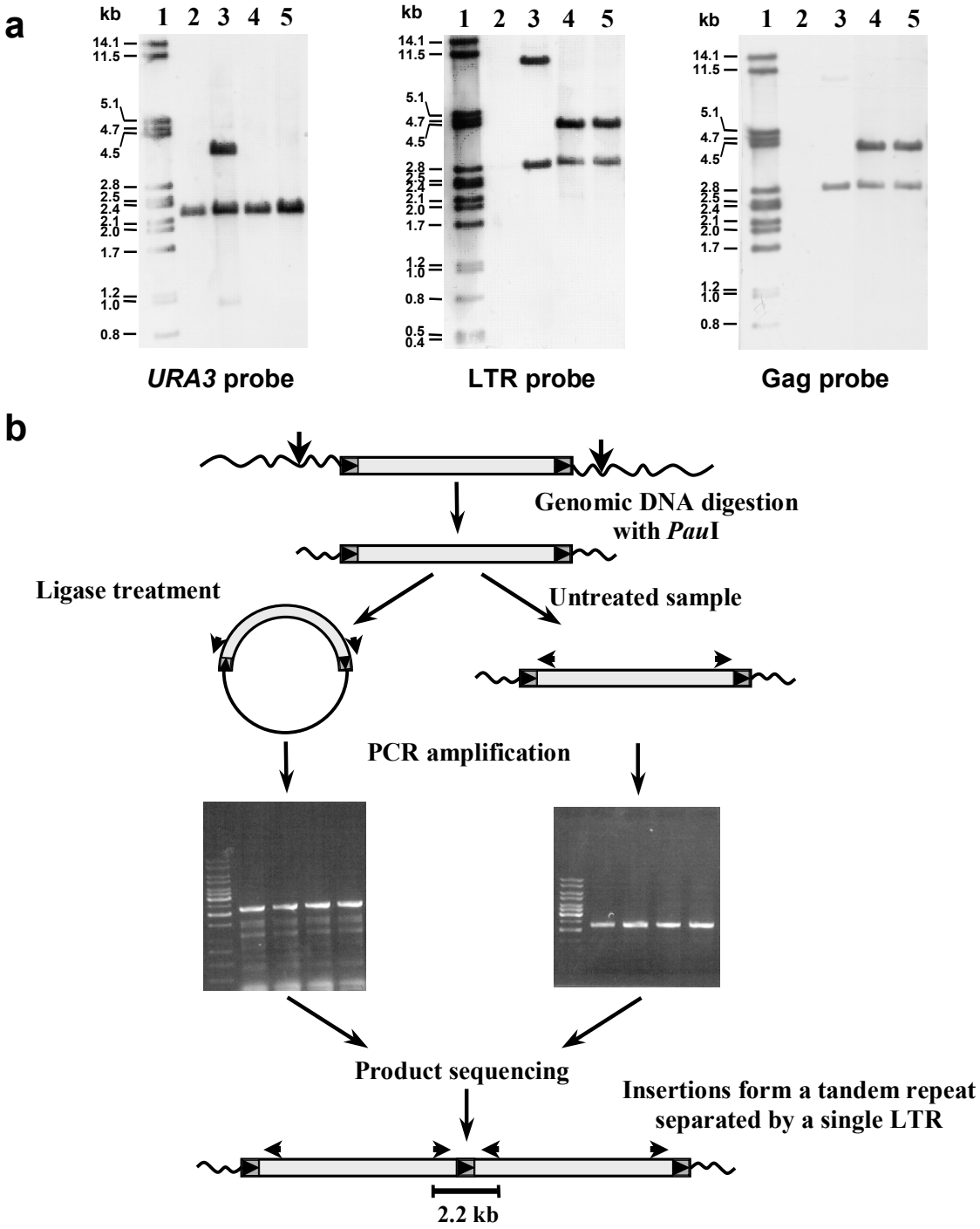

Figure 3.8 Analysis of Suc$^+$Ura$^-$ clones E1-52 and E1-72 obtained from HE31 strain in the *in vivo* retrotransposition assay.

(a) Southern analysis. *Y. lipolytica* strains H222-41 (lane 2), HE31 (lane 3), E1-52 (lane 4) and E1-72 (lane 5) were included in the analysis. Genomic DNA samples were digested with the endonucleases *Sal*I (the first blot) or *Hind*III (the second and the third blots). The blots were hybridized with *URA3*-, LTR- or Gag-specific probes. DNA of phage λ digested with the endonuclease *Pst*I was used as a molecular weight marker (lane 1).

(b) A scheme illustrating the "reverse PCR"-based analysis of the clones of interest. Divergent primers Yip1 and Yip2 used to amplify the insertion site are shown as small black arrows, and LTRs – as black triangles. The product obtained in both PCR reactions is indicated. GeneRuler 1 kb DNA ladder (Fermentas) was used as a molecular weight standard in DNA electrophoresis.

3.5 Genome wide analysis of the integration preferences of the retrotransposon Ylt1

The complete sequencing of the genome of the *Y. lipolytica* strain E150 carried out by the Genolevures consortium was finished in 2004. Obtained data allowed the analysis of the Ylt1 distribution in the *Y. lipolytica* genome.

Most LTR retrotransposons generate so-called target site duplication (TSD) upon their integration. These short duplications (4-5 bp) can be used to derive a target site consensus sequence of retroelements. The genome sequence of the *Y. lipolytica* E150 strain contains 11 full-length Ylt1 elements and 18 solo LTR, all of them are flanked by perfect 4-bp target site duplications. Thus, 29 TSDs were analyzed. Obtained results are shown in the tables 3.2 and 3.3.

Table 3.2 Target site duplications produced by Ylt1 in the genome of *Y. lipolytica* E150

TSD	Number	Frequency, %
TATG	8	28
CATG	7	25
CATA	6	21
TATA	3	10
GATG	3	10
AATG	1	3
TATC	1	3
Total	29	-

These data demonstrate that Ylt1 indeed has some integration preferences. The first position of its target site is occupied predominantly by pyrimidine residues (T in 12 cases and C in 13 cases out of 29). At the same, purine residues (G and A) were found in all but one case in the fourth position. Finally, A and T were invariantly present in the second and in the third position, respectively.

Table 3.3 Target site consensus sequence of Ylt1

	Position (%)			
	1	**2**	**3**	**4**
A	3	100	0	31
T	42	0	100	0
G	10	0	0	66
C	45	0	0	3
Consensus	C/T	A	T	G/A

Next, the integration preferences of Ylt1 in the genomic context were investigated. All but one from the 29 analyzed insertions were not associated with tRNA genes, i.e., no tRNA genes were found within 750 bp window upstream and downstream of insertion sites. Further on, the single insertion of Ylt1 associated with $tRNA^{Gly}$ gene was found downstream from tRNA gene, whereas insertions of the *S. cerevisiae* retrotransposons Ty1 and Ty3 are located mostly upstream of tRNA genes.

The majority of the analyzed Ylt1 insertions (28 out of 29) were associated with the intergenic regions instead. Only five of them were found between convergent genes, suggesting some bias for the regions upstream of RNA polymerase II (Pol II) promoters. Majority of the remaining 23 insertions (16) were located within regions 500-3000 bp upstream of the predicted ORFs. Further on, five insertions were found at the distances of <500 bp to the next ORF, and in three cases the distances between insertion point and the next ORF exceeded 3000 bp.

Obtained results suggest that Ylt1 has no strict insertion preferences. The whole-length elements and solo LTRs of Ylt1 are located mainly in the intergenic regions of Pol II transcribed genes. The distances between insertion points and neighboring ORFs vary widely, from 22 bp to more than 3000 bp. At the same time, insertions of Ylt1 are characterized by significant sequence specificity, as more than 82 % of the analyzed target size duplications fit the consensus sequence C/T-A-T-G/A.

3.6 A novel Ty3/*gypsy* retrotransposon Tyl6 from the genome of *Y. lipolytica*

The results obtained in the course of the study of the retrotransposon Ylt1 indicate that it occupies an isolated position among known retroelements. The combination of properties demonstrated by Ylt1 is quite unusual and suggests its ancient origin. So, the attempts to find related elements in the complete genome of the *Y. lipolytica* strain E150 were done. First, BLAST searches against the genome database of *Y. lipolytica* maintained by the Genolevures consortium [Dujon *et al.*, 2004; http://cbi.labri.fr/Genolevures/] were performed. Predicted amino acid sequences of the reverse transcriptases and integrases of Ylt1, Ty3 and Tyl3 retrotransposons were used as a query in these searches. No Ylt1-related elements were found in the genome of *Y. lipolytica* E150. However, the BLAST searches revealed a novel retroelement belonging to the Ty3/*gypsy* group. Further on, the complete sequence of the novel retrotransposon was identified.

Table 3.4 Analysis of Ylt1 insertions in the genome of *Y. lipolytica* E150

Position	Insertion	Orientation	Upstream locus				Downstream locus			
			Locus tag	**Similarity**	**Orientation**	**Distance to the Ylt1 insertion (bp)**	**Locus tag**	**Similarity**	**Orientation**	**Distance to the Ylt1 insertion (bp)**
Chromosome A										
213,927-214,640	Solo LTR	←	YALI0A 01606g	Pseudogene; weakly similar to *S. cerevisae* Rhc18 protein	→	3435	YALI0A 01650g	No similarity	←	1250
870,426-879,878	Ylt1	←	YALI0A 08888g	No similarity	→	380	YALI0A 08976r	$tRNA^{Gly}$	←	55
915,377-924,829	Ylt1	→	YALI0A 09284g	No similarity	←	21	YALI0A 09361	No similarity	←	640
1,058,220-1,067,672	Ylt1	→	YALI0A 10362g	No similarity	→	205	YALI0A 10439g	No similarity	→	1873
1,203,289-1,204,002	Solo LTR	←	YALI0A 11561g	No similarity	←	2373	YALI0A 11605g	Weakly similar to *Sz. pombe* putative MSF transporter	→	1845
1,324,357-1,325,070	Solo LTR	←	YALI0A 12859r	$tRNA^{Tyr}$	→	2153	YALI0A 12925g	Weakly similar to *S. cerevisiae* Zn-finger protein Rme1	→	3256
1,745,634-1,746,347	Solo LTR	→	YALI0A 16863g	Similar to *S. cerevisiae* Flx1 protein	←	1216	YALI0A 16907r	$tRNA^{Arg}$	→	885
2,268,388-2,269,101	Solo LTR	←	YALI0A 21197g	Similar to *S. cerevisiae* YGR145w	→	699	YALI0A 21241g	No similarity	→	834
2,302,413-2,303,126	Solo LTR	→	YALI0A 21461g	Similar to *S. cerevisiae* YKR018c	←	2226	-	-	-	-
Chromosome B										
18,607-19,320	Solo LTR	→	YALI0B 00220g	Similar to *S.cerevisiae* Npt1 protein	←	1876	YALI0B 00264g	Similar to *S. cerevisiae* iron homeostasis protein Nfu1	→	761
841,364-850,816	Ylt1	→	YALI0B 06285g	DNA transposon Mutyl	→	1469	YALI0B 06347g	No similarity	→	2549
863,908-873,360	Ylt1	→	YALI0B 06413g	No similarity	→	126	YALI0B 06490g	No similarity	←	631
Chromosome C										
1,031,320-1,032,033	Solo LTR	→	YALI0C 07777g	No similarity	←	1853	YALI0C 07821g	Weakly similar to *N. crassa* transcriptional regulator amyR	→	2094
1,090,601-1,091,314	Solo LTR	←	YALI0C 08140g	No similarity	→	1348	YALI0C 08184g	No similarity	←	5125
1,780,526-1,781,239	Solo LTR	→	YALI0C 12881g	Similar to *S.cerevisiae* YPR143w	→	867	YALI0C 12918g	No similarity	→	58
2,235,842-2,236,555	Solo LTR	←	YALI0C 15741g	No similarity	→	2583	YALI0C 15785g	No similarity	→	5198
Chromosome D										
2,505,196-2,505,909	Solo LTR	←	YALI0D 19778g	Weakly similar to *N. crassa* Q96U77	←	678	YALI0D 19822g	No similarity	←	50
2,579,603-2,580,316	Solo LTR	→	YALI0D 20350g	Weakly similar to *S. cerevisiae* Taf1 protein	←	2127	YALI0D 20394g	Weakly similar to *S. cerevisiae* Uga3 protein	→	509
2,664,336-2,665,049	Solo LTR	←	YALI0D 21032g	Similar to *N. crassa* Cax4 protein	←	913	YALI0D 21076g	Similar to *C. albicans* IPF7514 protein	←	271

Position	Insertion	Orientation	Upstream locus				Downstream locus			
			Locus tag	Similarity	Orientation	Distance to the Ylt1 insertion (bp)	Locus tag	Similarity	Orientation	Distance to the Ylt1 insertion (bp)
Chromosome D										
2,686,388-2,695,839	Ylt1	←	YALI0D 21230g	Some similarities with *Emericella nidulans* FlbD protein	←	1354	YALI0D 21318g	Similar to *S. cerevisiae* YOR131c	←	6211
2,768,739-2,778,191	Ylt1	→	YALI0D 21813g	No similarity	←	739	YALI0D 21890g	Similar to *S.cerevisiae* Ams1 protein	→	1878
3,298,564-3,299,277	Solo LTR	←	YALI0D 24772g	Weakly similar to *S. cerevisiae* hypothetical protein YJL163c	←	1323	YALI0D 24816g	No similarity	←	338
Chromosome E										
1,117,573-1,127,025	Ylt1	←	YALI0E 09306g	Weakly similar to *Kluyveromyces lactis* KLLA0E08899g	←	1239	YALI0E 09372g	No similarity	←	0
1,326,997-1,327,710	Solo LTR	←	YALI0E 10681g	No similarity	←	767	YALI0E 10725g	Weakly similar to *Sz. pombe* RNase H	←	1017
1,608,248-1,617,700	Ylt1	→	YALI0E 13343g	Weakly similar to *N. crassa* hypothetical protein NCU03670.1	←	196	YALI0E 13420g	Similar to *Pseudomonas aeruginosa* electron transfer flavoprotein Q9HZP7	→	1212
4,004,253-4,004,966	Solo LTR	→	YALI0E 33715g	No similarity	←	62	YALI0E 33737g	Weakly similar to *C. albicans* IFF2 protein	←	0
4,116,680-4,126,132	Ylt1	→	YALI0E 34573g	No similarity	→	2700	YALI0E 34650g	Weakly similar to *Sz. pombe* putative PHD-type Zn-finger protein	→	1089
4,173,487-4,182,939	Ylt1	→	YALI0E 34947g	Similar to *C. albicans* IPF20013 protein	←	389	YALI0E 35024g	No similarity	←	3129
Chromosome F										
2,198,957-2,199,670	Solo LTR	→	YALI0F 16423g	Similar to *Debaryomyces hansenii* DEHA0C13233g	←	3356	YALI0E 16467g	Some similarities with *N. crassa* hypothetical protein NCU02421.1	←	1336

The novel element was named Tyl6 (for the Transposon of *Yarrowia lipolytica* 6). This name follows the systematic nomenclature proposed by Neuveglise *et al*. [2002]. The nomenclature uses the first letters of the generic and species names of the host organism in the combination with the T letter (for Transposon). Further on, it takes into account the similarity of novel retroelements with the reference retrotransposons of *S. cerevisiae* and *C. albicans* (Ty1, Tca2, Ty3, Ty4 and Ty5). Each of the reference elements defines a distinct family of yeast LTR retrotransposons. Thus, the newly identified *Y. lipolytica* retrotransposon should be named Tyl3 as it belongs to the Ty3/*gypsy* family of retroelements. However, the retrotransposon Tyl3 has

been described from the genome of the *Y. lipolytica* W29 strain before [Neuveglise *et al.*, 2002], and it is clearly different from the novel element. Thus, the name Tyl6 was chosen.

3.6.1 Analysis of the nucleotide sequence of Tyl6

3.6.1.1 Regulatory elements

The entire element Tyl6 is 5108 bp long. It is flanked by two identical LTRs each of 276 bp (Fig. 3.9). Tyl6 was detected as a single copy element located at the chromosome E. This copy is flanked by 5 bp target site duplications TCTTA...TCTTA. The LTRs are bordered by 5 bp long inverted repeats TGTAA...TTACA, which are characteristic for retrotransposons and retroviral genomes. Searches for transcription-regulating elements in the LTR sequence were carried out. A putative TATA box was identified at the position 134-138 (numeration in all cases is started from the 1st bp of 5´ LTR). A polyadenylation signal is not clearly defined in the LTR of Tyl6, but a sequence, that resembles it, is located at the position 156-161.

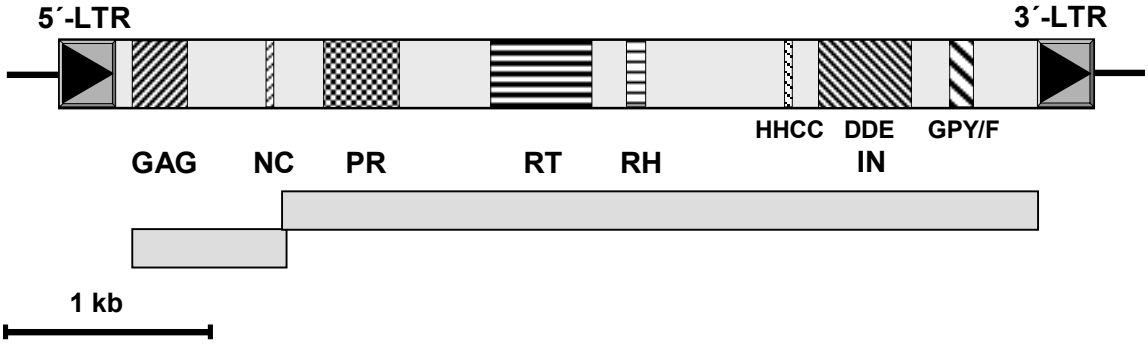

Figure 3.9 The structural organization of the retrotransposon Tyl6. LTRs are presented as boxes with black triangles. Striped boxes indicate the localization of conserved motifs for Gag protein (Zn-finger domain of Gag protein is shown as nucleocapsid motif (NC)), protease (PR), reverse transcriptase (RT), RNaseH (RH) and three domains of integrase (IN) (see text for details). Rectangles represent schematically the organization of coding sequences and a translational –1 frame-shift.

A putative primer-binding site (PBS) of Tyl6 is located 2 bp downstream from the 5´ LTR (Fig. 3.10). The identified PBS sequence is characteristic for retrotransposons whose reverse transcription is primed by host-encoded tRNAs as it contains the trinucleotide TGG, which is complementary to the universal CCA of the acceptor stem of tRNA. First 8 nucleotides of the proposed PBS of Tyl6 are identical to the PBS of the *S. cerevisiae* retrotransposon Ty3. The Ty3 element is known to use the initiator $tRNA^{iMet}$ as a primer for its reverse transcription. Further on, it turned out that the identified copy of Tyl6 is integrated 15 bp upstream of a putative $tRNA^{iMet}$ gene of *Y. lipolytica* (for more details, see below), and that first 12 nt of the PBS are

Figure 3.10 Scheme showing the complementarity of a putative primer-binding site of Tyl6 to the cytoplasmic $tRNA^{iMet}$ from *Y. lipolytica* and *S. cerevisiae*.

complementary to the 3´ acceptor stem of a tRNA encoded by the predicted gene (Fig. 3.9). Notably, the closely related Tse3 element from the genome of *Saccharomyces exiguus* also has a PBS complementary to $tRNA^{iMet}$ and of the same length [Neuveglise *et al.*, 2002]. These observations strongly suggest that Tyl6 has adopted $tRNA^{iMet}$ as a primer for its reverse transcription.

Another region playing an important role in the life cycle of retroelements – the polypurine tract – is well defined in the sequence of Tyl6. This relatively small purine-rich sequence is relatively resistant to RNaseH degradation, thus, it serves as a primer for the (+)-strand cDNA synthesis. It is located immediately upstream of the 3´ LTR. The nucleotide sequence of the polypurine tract of Tyl6 is AGG GGG GGA GAG.

3.6.1.2 Coding sequences

The organization of coding sequences of Tyl6 is typical for LTR retrotransposons. There are two ORFs separated by a –1 translational frame-shift. A putative frame-shift promoting sequence (slippery heptamer) G-GGA-AAT was found at the positions 1093-1099. It agrees well with the common model for such sequences, which can be defined as X-XXY-YYZ [Farabaugh, 1996]. Further on, the ribosomal frame-shifting is known to be stimulated by the secondary structures of mRNA (pseudoknot or even a simple hairpin loop), commonly occurring 6 nt downstream of the slippery heptamer. The distance between the heptamer and the pseudoknot is critical and its changing by as little as 2 nt eliminates stimulation [Farabaugh, 1996]. Indeed, a putative stem-loop structure between residues 1107 to 1136, i.e. exactly 6 nt downstream of the heptamer G-GGA-AAT, and a deduced pseudoknot interaction between bases 1120-1124 and 1150-1154 were detected during the sequence analysis (Fig. 3.11).

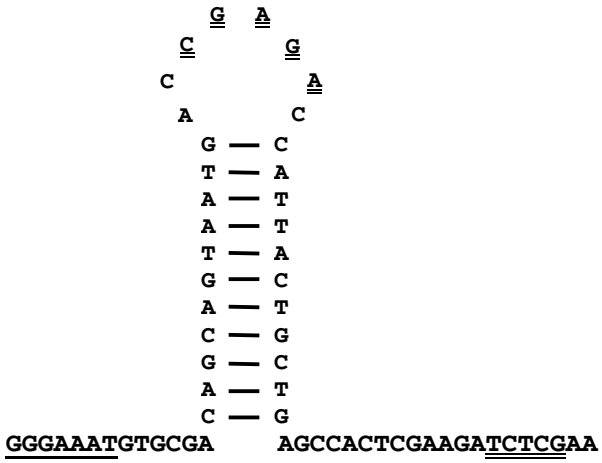

Figure 3.11 Stem-loop structure and a pseudoknot located downstream from the frame-shift promoting heptamer G-GGA-AAT. The slippery heptamer is underlined, and the nucleotides involved in the formation of the putative pseudoknot are double underlined.

The first ORF of Tyl6 encodes a deduced Gag protein consisting of 252 amino acid residues. The predicted product of the second ORF includes regions of homology to protease, reverse transcriptase and integrase of known retrotransposons. The observed arrangement Gag – PR – RT – IN is characteristic for most described members of the Ty3/*gypsy* group. The length of a predicted fusion polyprotein (Gag+Pol) resulting from a –1 frame-shift during translation is 1488 aa. Interestingly, Tyl6 is the first example of a retrotransposon from ascomycetous yeasts utilizing -1 translational frame-shifting, although this type of the translational recoding commonly occurs in retroelements of filamentous ascomycetes [Gao *et al.*, 03]. Notably, the retrotransposon Tcn1 from the genome of the basidiomycetous yeast species *Cryptococcus neoformans* [Goodwin and Poulter, 2001b] has two ORFs separated by -1 frame-shift.

Gag proteins of various retroelements often show only a limited level of sequence similarity. Nevertheless, two characteristic regions were identified in a deduced sequence of a Gag protein of Tyl6. The first one is encoded by the nucleotides 357–633 of Tyl6. It displays homology with the corresponding regions of some previously described Gag proteins. The second conserved domain present in virtually all Gag proteins is the so-called Zn-finger motif $CX_2CX_4HX_4C$ (also named nucleocapsid (NC) region) [Covey, 1986]. It is responsible for the binding of Gag protein to genomic RNA of a LTR retrotransposon. In the retrotransposon Tyl6 it was identified at the positions 1014–1056. It has the amino acid sequence CYTCGKIGHLARDC (conserved residues are underlined) (Fig. 3.12a). Other parts of the putative Gag protein encoded by Tyl6 show no significant homology with known proteins.

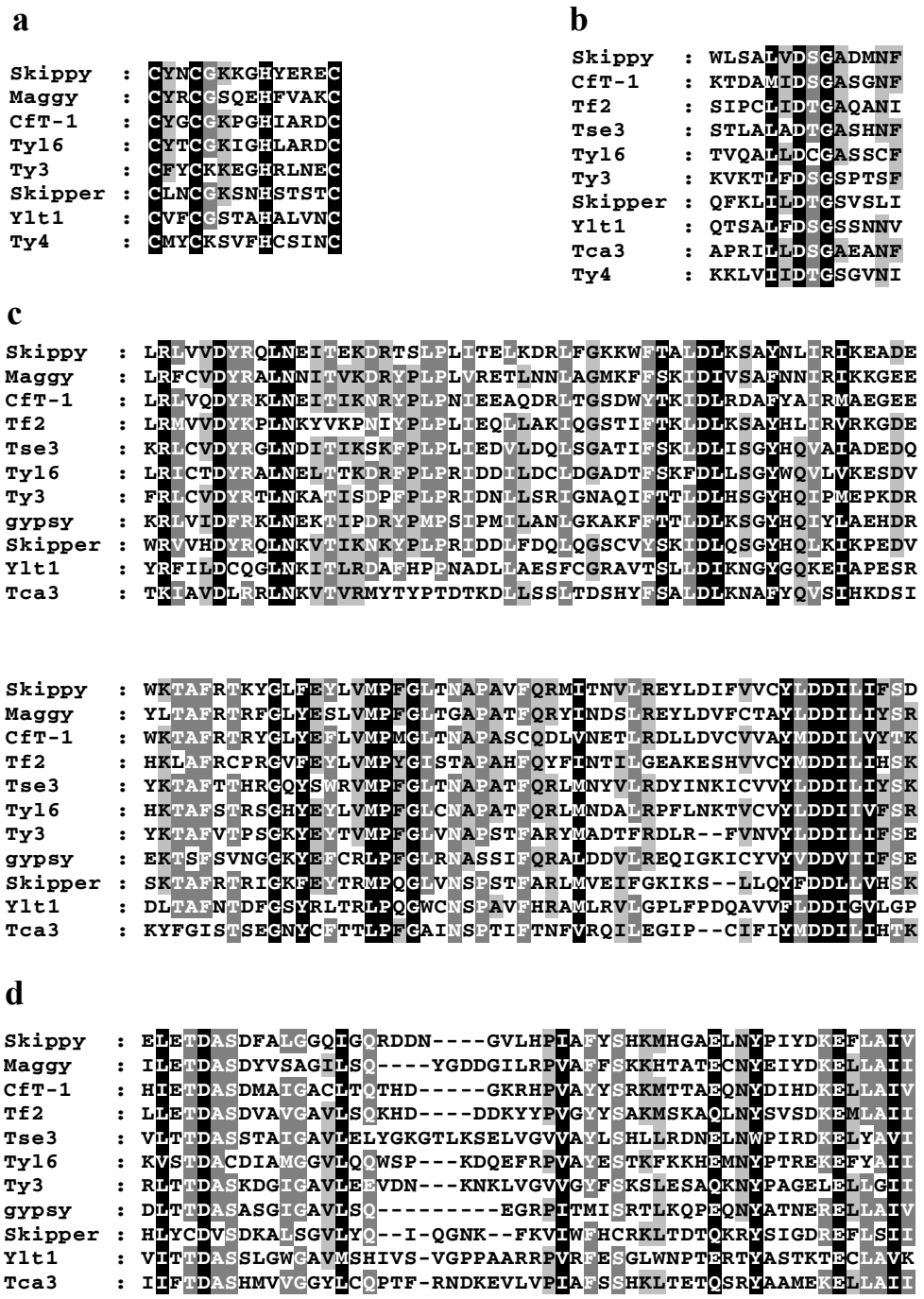

Figure 3.12 Multiple sequence alignments of conserved sequence motifs from various LTR-retrotransposons. The alignments of the conserved parts of Gag protein (Zn-finger) (**a**), protease (**b**), reverse transcriptase (**c**) and RNaseH (**d**) are presented. Conserved residues are shown in white on a black background, residues with moderate similarity (≥ 80 %) - on a dark-grey background, and with low similarity level (≥ 60 %) – on a grey background. The following retroelements were included in the alignments: Ty3 and Ty4 from *S. cerevisiae*, Tse3 from *S. exiguus*, *Skippy* from *Fusarium oxysporum*, *Maggy* from *M. grisea*, CfT-1 from *Cladosporium fulvum*, *Skipper* from *Dictyostelium discoideum*, Tca3 from *C. albicans*, Tf2 from *Sz. pombe*, Tyl6 and Ylt1 from *Y. lipolytica* and *gypsy* from *D. melanogaster*. The sources of the sequences are indicated in *Materials and Methods*.

A region with a similarity to aspartic protease was identified at the 5´ end of the second ORF of Tyl6. It encompasses nucleotides 1295-1696. A characteristic protease motif LLDcGASS was found in the deduced protein (Fig. 3.12b). It is deviated from the consensus protease motif (hydrophobic residue)$_2$-D-S/T-G-A/S, which is found in the large majority of retroelements [Dunn *et al.*, 2002]. Namely, the active center of the putative Tyl6 aspartic protease has a cysteine residue (shown in low case) instead of conserved serine/threonine. The next domain in the deduced polyprotein possesses a strong homology to reverse transcriptase, which is known to be the most conserved domain among different retroelements [Xiong and Eickbush, 1990]. The conserved motif of reverse transcriptases, the so-called YXDD box, is encoded by the sequence located at the positions 2468-2479 (VCVYLDDIIV) (Fig. 3.12c). The TDAS motif of RNaseH was found in a slightly deviated form (TDAC) downstream from the reverse transcriptase domain, namely at the positions 2915-2926 of Tyl6 (Fig. 3.12d). The characteristic features of integrase – the N-terminal Zn-finger domain with HHCC motif involved in the binding to LTR sequences, the core domain containing the catalytic D,D$_{35}$-E motif and the C-terminal GPY/F module [Khan *et al.*, 1991; Malik and Eickbush, 1999] – are identified in the deduced polyprotein at the positions 3526-3664, 3751-4051 and 4369-4480, respectively.

The amino acid sequence of reverse transcriptase and the presence of the GPY/F module in the integrase sequence indicate that Tyl6 clearly belongs to the Ty3/*gypsy* group of retroelements.

3.6.2 Distribution of Tyl6 among *Y. lipolytica* strains

Y. lipolytica strains of different origin can vary significantly in their genetic structure and chromosome lengths. Differences in the genome size between various isolates were also reported [Barth and Gaillardin, 1996]. Further on, the retrotransposon Ylt1 was shown to be unequally distributed among *Y. lipolytica* strains [Juretzek *et al.*, 2001; Senam, 2004]. Therefore, it was decided to study the distribution of the newly detected Tyl6 element as well. The genomic DNA isolated from selected strains was analyzed by Southern blot hybridization with several Tyl6-specific probes (Fig. 3.12). The following strains were taken for the study. First, two wild type isolates of independent origin were chosen, YB423-12 [Wickerham *et al.*, 1969] and H222 (isolated from German soil). The strain YB423-12 is one of the ancestors of the majority of commonly used laboratory strains [Barth and Gaillardin, 1996]. A third strain PO1d is derived directly from W29, a wild type isolate from French sewage. Representatives of two inbreeding lines widely used in a laboratory practice were taken as well. They were B204-12C (German inbreeding line) [Barth and Gaillardin, 1996], and E129 and E150 (both derived from French inbreeding line) [Barth and Gaillardin, 1996]. It should be noted here that the last strain (E150)

was used in the genome sequencing project performed by the Genolevures consortium [Dujon *et al.*, 2004].

The genomic DNA samples were digested with a restriction endonuclease *Sal*I. The enzyme *Sal*I has a single recognition site in Tyl6 at the position 2060 (Fig. 3.13a). The following fragments of Tyl6 were used as probes for hybridization (Fig. 3.13): the complete LTR sequence, a fragment of the *gag* region located between unique *Pst*I and *Bgl*II recognition sites (669-1147 bp) and an *Eco*RI-*Eco*RV fragment of the *pol* region (3047-3903 bp).

Hybridization of these fragments of Tyl6 with genomic DNA isolated from different strains of *Y. lipolytica* brought following results (Fig. 3.13). First of all, it turned out that Tyl6 is not equally distributed among *Y. lipolytica* strains. It is present in the genome of the wild type isolate YB423-12 and in the genome of strains from both French and German inbreeding lines (Fig. 3.13b). Since YB423-12 was used in both of these inbreeding programs as one of the ancestor strains, it is suggested to be an original host of Tyl6. This is also supported by the fact that strains of independent origin (H222 and PO1d) were shown to be Tyl6-free.

Some differences were detected in the hybridization patterns obtained with different probes. Hybridization with the *gag* probe gave no non-specific signals (Fig. 3.13c), confirming the unequal distribution of Tyl6 and the presence of its single copy in Tyl6-harboring strains. At the same time, additional weak signals were observed when the LTR or the *pol* probe was used (Fig. 3.12b and 3.12d). The nature of these signals was further investigated. A BLAST search against *Y. lipolytica* genome database was performed using the nucleotide sequences of both the LTR and the *pol* probes as a query.

Two sequences of interest located on chromosomes C and E were found when the sequence of the LTR probe was used as a query. Both identified sequences have a common region (73 bp, 87% identity to Tyl6 LTR), which could be recognized by the LTR probe. The expected sizes of bands on Southern blots, which should be obtained if the LTR probe indeed recognizes both identified fragments, are 4.76 kb and 4.46 kb, respectively. First value agrees well with the size of the smaller unspecific band observed on blots. However, the size of second band (>5.1 kb) differs significantly from the predicted value (4.46 kb). These differences may be partially explained by the errors in the available sequence of *Y. lipolytica* genome, especially taking into account a presence of numerous undetermined nucleotides in corresponding region of the chromosome E (positions 2388970-2389080 in the sequence deposited to GenBank). Further on, a distribution of this second weak band (about 5.1 kb) among *Y. lipolytica* strains provides additional support for the suggestion that it results from the hybridization of LTR probe with the identified region on the chromosome E. This band was observed only in Tyl6-containing strains,

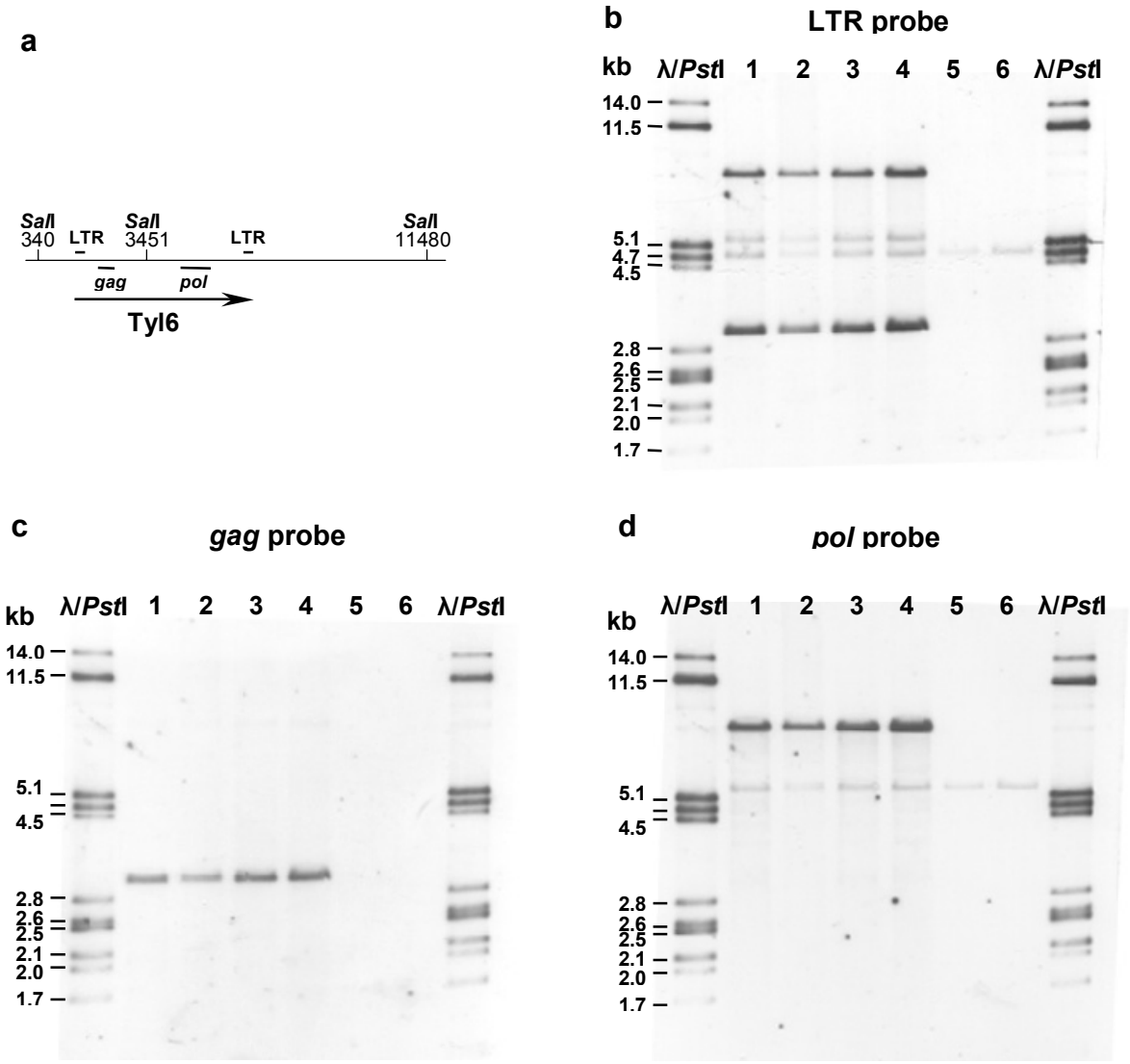

Figure 3.13 Detection of Tyl6 by Southern blot hybridization in several strains of *Y. lipolytica* of different origin. Strains B204-12C (1), E129 (2), E150 (3), YB423-12 (4), H222 (5) and PO1d (6) were analyzed. Genomic DNA samples were digested with *Sal*I endonuclease. DNA of phage λ digested with the endonuclease *Pst*I was used as molecular weight marker.

(a) Localization of recognition sites for *Sal*I endonuclease in the Tyl6-containing fragment of chromosome V of *Y. lipolytica* E150 strain, taken from the Genolevures database. Positions of probes used for Southern hybridization (LTR, *gag* and *pol*) are indicated.

(b) Hybridization of genomic DNA of *Y. lipolytica* strains with the LTR probe. Two bands (8.0 and 3.1 kb) corresponding to both 5′ and 3′ LTRs were observed in Tyl6-containing strains. Two weak unspecific bands (5.3 and 4.8 kb) were also detected. Their nature is discussed in the text.

(c) Hybridization of genomic DNA of *Y. lipolytica* strains with the *gag* probe. A single band of predicted size of 3.1 kb was detected in the same strains.

(d) Hybridization of genomic DNA of *Y. lipolytica* strains with a *pol* probe. A band of 8.0 kb is observed in Tyl6-harboring strains. A weak unspecific band (5.3 kb) was present in all tested strains. Its nature is discussed in the text.

suggesting that both loci are linked genetically, i.e., that both Tyl6 and the sequence non-specifically recognized by LTR probe are located on the chromosome E.

Further on, our analysis showed that the *pol* probe could recognize a short fragment on chromosome F (87 bp, 93% identity). The sizes of corresponding band on Southern blots in this case should be 5.33 kb, what agrees well with the obtained results.

The fragment interacting with the LTR probe showed no significant similarity to any known retroelement. Further on, no significant homology with other protein-coding sequences could be detected. On the other hand, the fragment hybridizing with the *pol* probe displays very limited homology to the RT/RH region of some retrotransposons, first of all, of Tyl6. It is unclear, whether these sequences represent highly degenerated retrotransposons, arised as a result of a recombination event or whether the observed homology is just a result of coincidence.

3.6.3 Integration specificity of Tyl6

All analyzed Tyl6-harboring strains bear only a single copy of this retrotransposon. Moreover, the position of the element in the analyzed strains seems to be conserved, as the banding pattern was essentially the same in all cases (Fig. 3.12). It is not surprising since the Tyl6 was found only in the strain YB423-12 and its derivatives. The last ones should acquire their copy of Tyl6 from the genome of YB423-12.

The sequences surrounding the insertion site of Tyl6 in the genome of strain E150 were extracted from the *Y. lipolytica* genome database. The analysis of these sequences revealed that the Tyl6 element in the *Y. lipolytica* strain E150 is integrated 15 bp upstream of a putative $tRNA^{iMet}$ gene. This integration pattern resembles those of Tse3 and Ty3 elements, which integrate preferentially within a narrow "window" 13-19 bp upstream of tRNA genes (Fig. 3.14) [Chalker and Sandmeyer, 1992; Neuveglise *et al.*, 2002]. Although the presence of just a single copy of Tyl6 does not allow to make definitive conclusions about its target site specificity, the revealed location and observed similarity with Ty3 and Tse3 elements strongly suggest that the preferable target site of Tyl6 integration lays immediately upstream of tRNA genes.

3.6.4 Phylogenetic relationships of Tyl6 with other Ty3/*gypsy* retrotransposons

Sequence similarity unambiguously placed Tyl6 into the Ty3/*gypsy* group of retrotransposons. A detailed phylogenetic analysis was carried out in order to determine its relationships with other members of this group and with the known *Y. lipolytica* retrotransposons. The analysis was based on multiple alignments of the amino acid sequences of reverse transcriptase, which is the most conserved protein among various groups of retroelements. Seven previously

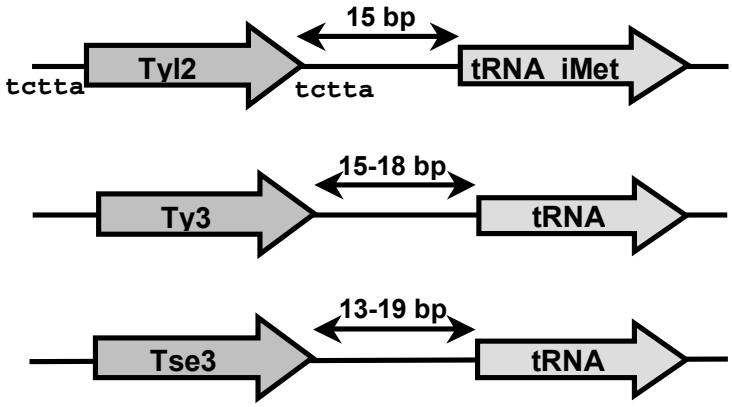

Figure 3.14 Integration preferences of the retrotransposons Tyl6, Ty3 and Tse3. Their insertions are found upstream of PolIII transcribed genes, within 1-2 bp of the transcription start. Target site duplication flanking the single known copy of Tyl6 is indicated.

identified amino acid domains including 178 residues [Xiong and Eickbush, 1990], which are common for all reverse transcriptases, were used for this study.

Malik and Eickbush [1999] have defined eight groups of Ty3/*gypsy* retrotransposons, named Ty3, mdg1, gypsy, Osvaldo, mag, Athila, Cer1 and mdg1. The members of all these lineages were included in the analysis. The phylogenetic tree was rooted with *copia*, a member of Ty1/*copia* class of retroelements known from the genome *D. melanogaster*.

The performed analysis grouped Tyl6 together with the Tse3 element of *S. exiguus* (Fig. 3.15). Further on, both elements appeared within the Ty3 group on the obtained phylogenetic tree.

The analysis also demonstrates that the two complete Ty3/*gypsy* retrotransposons found in *Y. lipolytica* genome (Ylt1 and Tyl6) are not closely related. The Tyl6 element clearly belongs to the Ty3-like elements, whereas Ylt1 was placed in the basal part of the phylogenetic tree of Ty3/*gypsy*-like retrotransposons. It is grouped together with Tca3 and Tcd3 retrotranposons of *C. albicans* and *C. dubliniensis*. These results provide a new insight on the phylogenetic position of the Ylt1 element, as well. The isolated position of Ylt1 in the basal part of the phylogenetic tree of Ty3/*gypsy* elements was shown by Goodwin and Poulter [2002]. Next, the relationship between Ylt1 and Tca3 was mentioned [Senam, 2004]. However, Ylt1 and Tca3 were not well resolved from the vertebrate retroviruses in the last work. The current analysis places Ylt1/Tca3 elements among the Ty3/*gypsy* group and clearly separates them from retroviruses. Thus, it defines a new basal clade of fungal Ty3/*gypsy* retrotransposons, whose properties suggest their ancient origin.

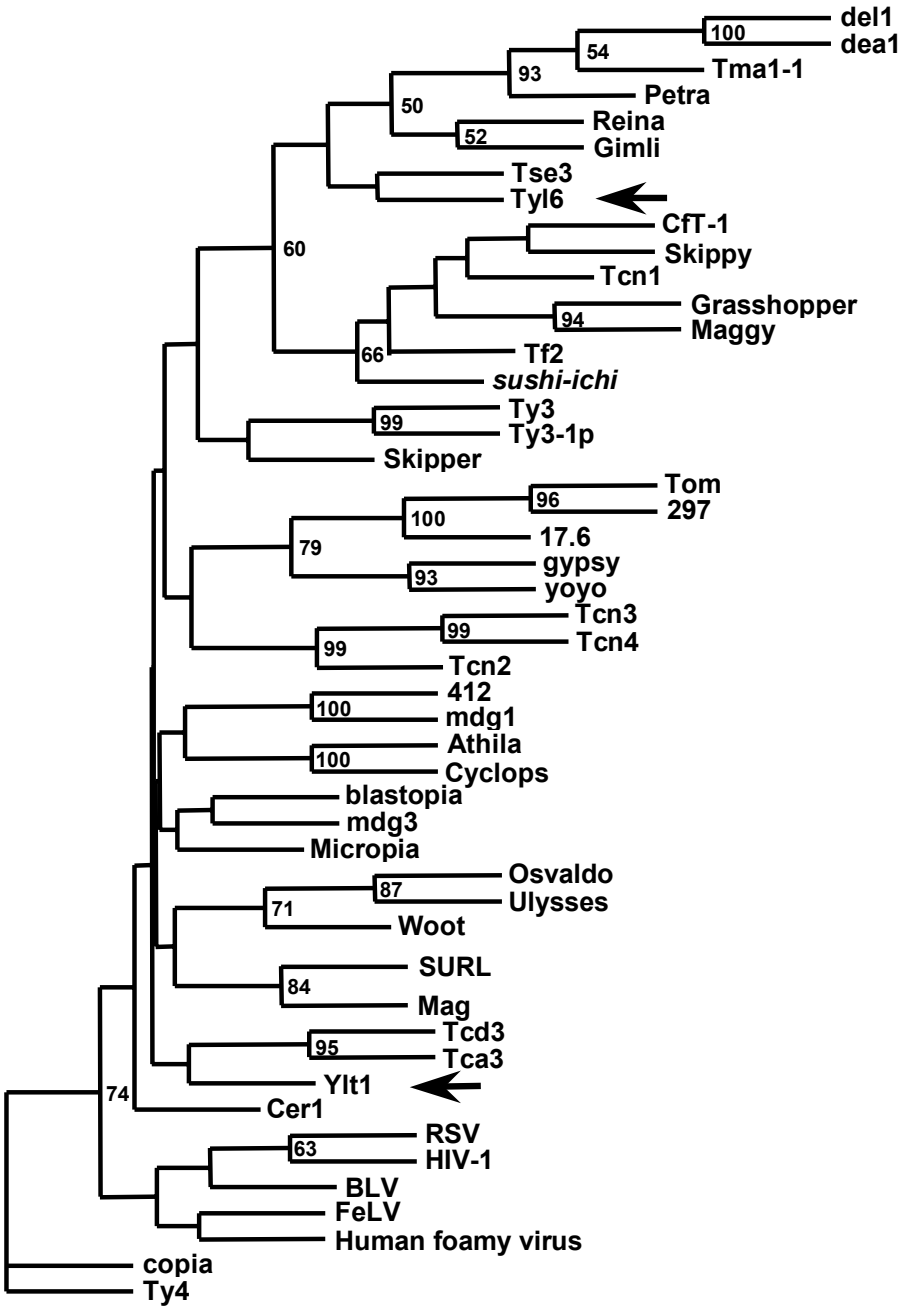

Figure 3.15 The phylogenetic tree based on the multiple alignment of the reverse transcriptase amino acid sequence. A variety of Ty3/*gypsy* retrotransposons, some retroviruses and two members of Ty1/*copia* class (as an outgroop) were included in the alignment. Numbers adjoining the branches indicate the bootstrap values from 100 bootstrap trials (only the values above 50% are shown). Positions of Tyl6 and Ylt1 are indicated by arrows. Abbreviations are as follows: RSV, Rous sarcoma virus; HIV-1, human immunodeficiency virus type 1; BLV, bovine leukemia virus; FeLV, feline leukemia virus. For the host organisms, see the chapter 2.8.2.

Unfortunately, the absence of the reverse transcriptase-encoding sequence of the third *Y. lipolytica* Ty3/*gypsy* element, Tyl3, in the public databases does not allow its inclusion into the analysis. The available Tyl3 sequence contains only the regions encoding for the core region and the C-terminal part of an integrase [Neuveglise *et al.*, 2002]. However, the corresponding amino acid sequences of Tyl6 and Tyl3 display significant similarity, and it is suggested that they may be indeed closely related.

The whole lineage of Ty3-like elements defined by Malik and Eickbush [1999] is often referred to as chromoviruses due to the presence of the chromodomain in the C-terminal part of integrase [Marin and Llorens, 2000; Kordis, 2005]. Chromoviruses are characterized by the Eukaryota-wide distribution. Numerous members of this group were described from fungal genomes as well, both from ascomycetes and basidiomycetes. Notably, the majority of the ascomycetous chromoviruses known so far were reported from filamentous fungi [Kordis, 2005]. Further on, Ty3 elements of the genus *Saccharomyces* were the only chromoviruses found in the genomes of hemiascomycetous yeast species. The analysis performed in this work placed both Tyl6 and Tse3 elements among chromoviruses, but at the same time it showed that these elements are not closely related to Ty3. Further on, Tyl6 and Tse3 show no significant similarity with the fungal chromoviruses of the Tf1/*sushi* group. So, Tyl6 and Tse3 have a PBS complementary to $tRNA^{iMet}$, whereas the members of the Tf1/*sushi* group utilize a self-priming mechanism. The phylogenetic analysis also confirmed the separation of Tyl6 and Tse3 from Tf1/*sushi* group. Therefore, it is suggested that both Tyl6 and Tse3 constitute a novel group of chromoviruses, specific for hemiascomycetous yeasts. This group appeared in the performed analysis as a separate branch next to the plant del1-like elements.

It was noted before that del1-like elements display similarities with some fungal retrotransposons [Wright and Voytas, 1998]. Similar to Tyl6 and Tse3, del1-like elements also have a $tRNA^{iMet}$-complementary PBS. Further on, some structural properties of retrotransposons from the neighboring Reina clade resemble those of Tyl6 and Tse3. The sizes of the members of the Reina clade lie between 5 and 7 kb [Kordis, 2005], whereas Tyl6 and Tse3 are 5108 bp and 6487 bp long [Neuveglise *et al.*, 2002], correspondingly. Thus, Tyl6 and Tse3 form a novel group of fungal chromoviruses, whose next relatives are del1-related plant retrotransposons.

The characteristic feature of the chromoviruses is a presence of the chromodomain. This is a 40–50 amino acids long domain, which is present in various eukaryotic proteins involved in chromatin remodelling and the regulation of gene expression. It may function as protein-interacting, RNA-interacting or DNA-binding module [Kordis, 2005]. Searches for chromodomains in the deduced amino acid sequences of both Tyl6 and Tse3 were performed;

however, no chromodomain-related sequences were found. Thus, although the phylogenetic analysis based on the amino acid sequence of the reverse transcriptase placed Tyl6 and Tse3 among chromoviruses, they apparently do not contain chromodomains.

A similar discrepancy was reported for the Ty3 element of *S. cerevisiae* [Malik and Eickbush, 1999]. Thus, all three retrotransposons of hemiascomycetous yeasts (Ty3, Tse3 and Tyl6) belonging to the chromoviruses are apparently lacking the chromodomain. However, all these elements do bear a C-terminal module of the integrase of approximately the same size. The alignment of these modules was carried out. Ty3-related element of *Saccharomyces paradoxus* (*Sp*Ty3-1) and the retrotransposon Tyl3 from the genome of *Y. lipolytica* strain W29 were also included in this alignment. Despite the fact that Ty3 and *Sp*Ty3-1, on one hand, and Tse3 and Tyl6, on the other hand, are not closely related (Fig. 3.15), a significant degree of similarity between C-terminal parts of their integrases was observed (Fig. 3.16). It should be noted that both Tyl3 and Tyl6 bear the amino acid motif TWE, which is present in most of the analyzed chromodomains [Kordis, 2005]. However, this motif was absent from the sequences of Ty3, *Sp*Ty3-1 and Tse3. Further on, no other conserved amino acid residues of the chromodomain were detected. Taking into account the striking similarity of the integration preferences of Ty3, *Sp*Ty3-1, Tse3 and Tyl6, these data support the model proposed by Malik and Eickbush [1999]. Namely, they proposed that the C-terminal part of the integrase of retroelements determines their

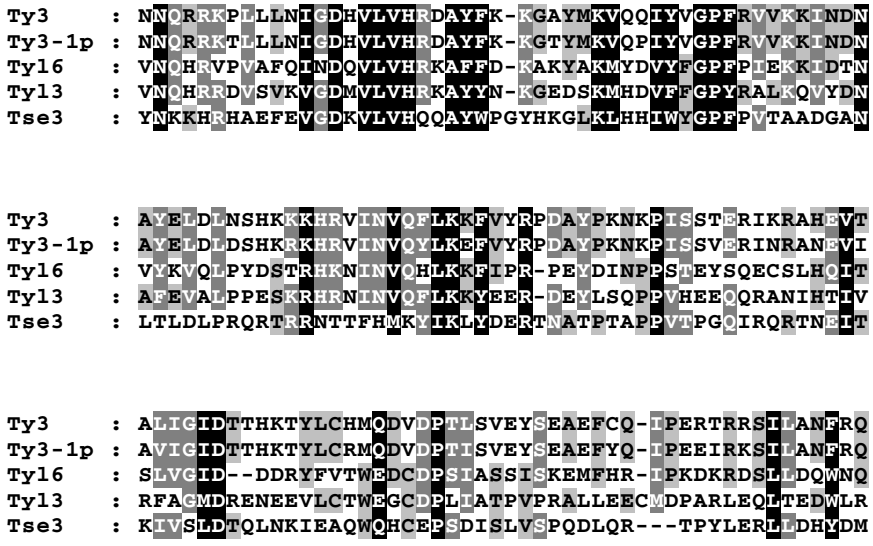

Figure 3.16 Comparison of the C-terminal integrase sequences of yeast retrotransposons, integrating in the close vicinity of tRNA genes. Presented elements form two phylogenetic groups (see Fig. 3.14), which are only distantly related, but their GPY/F module sequences display remarkable similarity, probably due to their integration specialization (compare with Malik and Eickbush, 1999). The integrase sequence of Tyl3 retroelement is included in the alignment as well, to show its similarity to Tyl6.

integration specificity. Further on, they suggested that the chromodomain of Ty3 may have been replaced or become specialized as a result of an adaptation of Ty3 to the genome of its host. The observed similarity of the corresponding regions of Ty3, Tse3 and Tyl6 retrotransposons suggests that this specialization is a general property of the chromoviruses of hemiascomycetous yeasts, which is linked with their preferred integration within 1-2 bases of PolIII transcription start site. It is also suggests the common mechanisms of target site choice used by analyzed elements. Finally, the presence of the TWE motif in the amino acid sequences of the Tyl3 and Tyl6 supports the model of the chromodomain specialization rather than its replacement in the chromoviruses of the hemiascomycetous yeasts.

4 Discussions

The aim of this work was the further characterization of the *Y. lipolytica* retrotransposon Ylt1. Proteins encoded by Ylt1 were detected, and their expression was characterized. An assay allowing *in vivo* observation of the Ylt1 transposition was developed, and the strains obtained in this assay were examined. Data obtained in the course of the genome sequencing of *Y. lipolytica* were analyzed. The insertion preferences of Ylt1 and its genome localization were specifically investigated. Also, a novel LTR retrotransposon was found in the genome of *Y. lipolytica*. Its structural properties, insertion preferences and distribution among *Y. lipolytica* strains were investigated.

4.1 Regulation of the Ylt1 expression

4.1.1 Expression of Gag protein of Ylt1

The expression of transposable elements is often tightly regulated in order to control their propagation. A variety of expression patterns was described for different retrotransposons. Both transcription and translation steps may be the subject of such regulation. Initiation of transcription is expected to be a key step limiting the frequency of transposition for many LTR retrotransposons [Boeke and Stoye, 1997]. On the other hand, the low expression of Ty1 proteins is achieved by means of post-transcriptional regulation, since Ty1-mRNA is highly abundant in *S. cerevisiae* cells [Ciriacy, 1995].

In this work, the expression of Ylt1 was studied using HA-tagged Gag protein, which was expressed under the control of the LTR promoter of Ylt1 and under the control of the *ICL1* promoter of *Y. lipolytica*. The presence of the HA-tagged Gag protein in *Y. lipolytica* cells was confirmed in both cases. It was easily detectable even when it was expressed under control of the weak LTR promoter. It was shown that LTR-driven expression of the Gag protein occurs on all tested carbon sources. Obtained data suggest that its expression occurs preferentially during the logarithmic growth phase, whereas transition to the stationary phase is accompanied by the significant reduction of the amount of Gag protein. Further on, addition of copper sulfate has a negative effect on its expression.

The proteins of other yeast retrotransposons are often poorly expressed, so they have almost exclusively been investigated by overexpression of individual elements [Garfinkel *et al.*, 1985; Adams *et al.*, 1987; Hansen *et al.*, 1992; Atwood *et al.*, 1996]. Conversely, the expression of many retrotransposons is activated by various stress factors, including heavy metals and heat shock [Strand and McDonald, 1985; Ziarczyk and Best-Belpomme, 1991; Ratner *et al.*, 1992; Mhiri *et al.*, 1997; Ikeda *et al.*, 2001]. So, the expression pattern of the Gag protein of Ylt1,

which was revealed in this work, has some unusual features. The Gag protein of Ylt1 is produced in detectable amounts in log-phase cells. In contrast, it is degraded upon transition to the stationary phase, and its expression is inhibited by copper sulfate. To our knowledge, such expression pattern was not reported previously for other LTR retrotransposons.

4.1.2 Expression of integrase and regulation of the Gag:Pol ratio

HA-tagged Gag protein of Ylt1 was readily detected in *Y. lipolytica* cell extracts even when it was expressed from the weak LTR promoter. However, no specific signals were observed when the expression of HA-tagged integrase of Ylt1 was analyzed in the same way. On the other hand, the overexpression of the HA-tagged integrase from the *ICL1* promoter resulted in the detection of a putative precursor, whose molecular weight corresponds to a protein containing both reverse transcriptase and integrase. An integrase alone was not observed in these experiments.

The HA tag used for the labeling of the integrase was readily detected in the case of the HA-tagged Gag protein. Therefore, obtained results suggest that the integrase of Ylt1 is unstable and is degraded rapidly. The observed instability may be caused by the introduction of the HA tag. However, it is also very likely that this instability is a natural property of the Ylt1 integrase, and that the integrase is degraded in order to achieve an adequate Gag:Pol ratio.

The molar excess of Gag protein is needed for the normal replication of LTR retrotransposons and vertebrate retroviruses. Usually, they express 10- to 50-fold more Gag than Pol protein. Different approaches were developed by retroelements to achieve the required Gag:Pol ratio. The most common ones are the use of rarely occurring ribosomal frame-shift or stop codon read-through for the expression of Pol protein [Gao *et al.*, 2003]. On the other hand, the retroelements expressing Gag and Pol in a single reading frame should have evolved alternative mechanisms, which allow them achieve a molar excess of Gag.

Little is known about the regulation of Gag:Pol ratio in yeast retroelements with a single ORF. Ty5 element of *S. cerevisiae* and Tf1 element of *Sz. pombe* are among few known examples of such retroelements. It is still not clear whether retrotransposons Tca3 and Tcd3, which are closely related to Ylt1, have a frame-shift between *gag* and *pol* [Goodwin et al., 2003]. The mechanism involved in the regulation of Gag:Pol ratio in the case of Tf1 was described by Atwood *et al.* [1996]. It is completely different from the alternative splicing mechanism used by the *copia* element of *D. melanogaster* [Brierley and Flavell, 1990].

The lower stability of integrase is exploited by the retrotransposon Tf1 to achieve the molar excess of Gag over Pol [Atwood *et al.*, 1996]. Like Ylt1, Tf1 expresses its Gag and Pol protein in a single reading frame, so they are produced in equimolar amounts. Indeed, in the case of Tf1

equal amounts of Gag and integrase are present during the log-phase [Atwood *et al.*, 1996]. However, integrase of Tf1 is rapidly degraded in stationary-phase cells, so that 26-fold excess of Gag is observed.

A related mechanism of Gag:Pol ratio regulation may be proposed for Ylt1. In this case, the rapid degradation of integrase would explain the difficulties in the detection of this protein.

An important difference between Tf1 and Ylt1 expression is the stability of Gag protein. The amount of Gag protein in the case of Tf1 remains at the same level and does not depends on the growth phase [Atwood *et al.*, 1996]. In contrast, Gag protein of Ylt1 was degraded in stationary phase cells. The difference may be partially explained by the kind of promoter used in these experiments. The strong *nmt1* promoter was used for the expression of Tf1 proteins [Atwood *et al.*, 1996], whereas the Gag protein of Ylt1 was expressed under control of the weak LTR promoter. The *nmt1* promoter remains active until thiamine is added to the cell culture, whereas the activity of the LTR promoter has a well-defined expression maximum [Senam, 2004]. However, the use of the LTR promoter had the important advantage, as LTR-driven expression better reflects the natural dynamics of Gag production. It remains unclear whether degradation of Gag in stationary phase cells is a result of cell defense against retroelement's propagation and whether this degradation has a certain role in the life cycle of Ylt1.

4.2 Ylt1 proteins

Ylt1 is by far the largest currently known fungal LTR retrotransposon. The most LTR retrotransposons described from yeasts usually do not exceed 6.0 kb, and only a few, like Tcn2 from *Cryptococcus neoformans*, have more than 7 kb in length. Thus, Ylt1 is approximately 1.5-fold as long as an average yeast LTR retrotransposon. A question arose, whether this length differences are reflected in the size of proteins encoded by Ylt1. Analysis of the nucleotide sequence revealed a huge single ORF, encoding a deduced polyprotein from 2621 aa [Senam, 2004]. However, the sizes of individual proteins could be only roughly estimated, since recognition sites of proteases of different retroelements show only little conservation. Further on, proteolytic processing of individual proteins during their maturation could not be excluded.

The apparent size of the HA-tagged Gag protein of Ylt1, as determined in this work, is 82 kDa. Taking into account that three tandemly repeated HA tags have a molecular weight of 4.5 kDa, the size of Gag protein alone should be about 78 kDa. Following apparent sizes were reported for other yeast LTR retrotransposons: 51-55 kDa (Ty1) [Boeke and Sandmeyer, 1991], 26 kDa (Ty3) [Hansen *et al.*, 1992] and 27 kDa (Tf1) [Atwood *et al.*, 1996]. Thus, the Gag protein of Ylt1 is 1.5- to 3-fold larger than other studied Gag proteins from yeasts. The differences are

especially pronounced if one compares the Gag proteins of Ty3/*gypsy* elements (Ylt1 vs. Ty3 and Tf1).

Gag proteins are generally proteolitically cleaved by retroelement-encoded proteases, so that C-terminal peptide is released. It is believed to be degraded in the case of Ty1, since it has never been directly observed. However, in the case of Ty3 this peptide has homology with retroviral nucleocapsid (NC) proteins, and it was detected in the virus-like particles produced by Ty3 element. The detection of HA-tagged Gag protein of Ylt1 reveals the presence of two bands having apparent size of approximately 80 and 82 kDa. This observation suggests that Gag protein of Ylt1 is proteolytically processed in the same way as Gag proteins of Ty1 and Ty3. The significance of the putative C-terminal fragment for the life cycle of Ylt1 remains unknown.

A weak high-molecular band of 140 kDa seen during the detection of HA-tagged Gag protein is proposed to be an intermediate product of polyprotein processing. Its size suggests that it contains both Gag protein and protease. On the other hand, the 140 kDa protein observed during the detection of HA-tagged integrase may contain both reverse transcriptase and integrase. Therefore, the following model of proteolytic processing of Ylt1 polyprotein is proposed (Fig. 4.1). The large polyprotein with a predicted molecular weight 289 kDa is cleaved by Ylt1 protease, and two fragments of approximately 140 kDa are released. Both these fragments were observed in this work. The N-terminal one is further processed to give rise to Gag protein (appr. 78-80 kDa) and protease. How the C-terminal fragment is processed is currently unclear. The Gag protein can be cleaved once again, so that C-terminal peptide will be released. However, obtained data suggest that such processed proteins constitute only a small fraction of the total Gag protein.

4.3 Transposition and target site preferences of Ylt1

Insertions of LTR retrotransposons are potentially deleterious to their host cells. This is especially true for yeast retroelements. Yeast genomes are often compact, they lack large intragenic and intergenic DNA regions, and many yeast species spend a large part of their life cycle in the haploid phase. Therefore, both the frequency and the specificity of transposition of yeast retrotransposons should be precisely adjusted to the genomes of their host organisms.

The copy number of yeast LTR retrotransposons rarely exceeds 30 elements per haploid genome. Further on, their transposition frequency is tightly regulated. So, the transposition frequency of Ty1 element does not exceed $5 \cdot 10^{-6}$ even when it is expressed under the control of the strong *GAL1* promoter [Ciriacy, 1995]. On the other hand, *de novo* transpositions of Tf1 and Ty3 elements were detected in up to 3 % of cells [Hansen et al., 1988; Levin and Boeke, 1992]. The

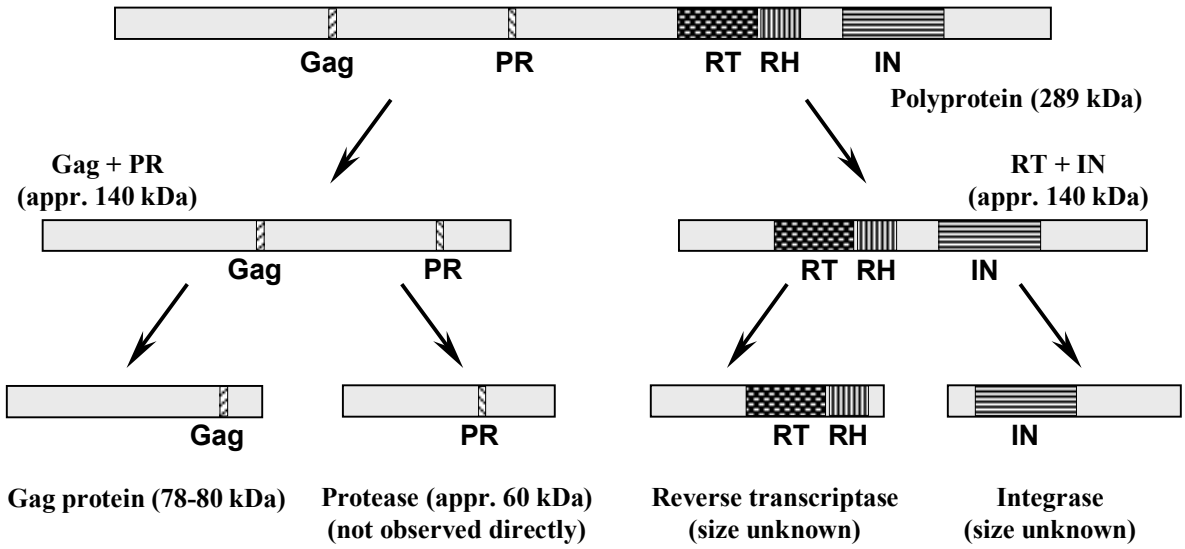

Figure 4.1 Scheme illustrating a proposed model of the processing of Ylt1 polyprotein. Large polyprotein with a predicted molecular weight of 289 kDa is synthesized during the translation of Ylt1 mRNA. The cleavage by Ylt1 protease produces two fragments of approximately 140 kDa. N-terminal fragment is further processed to give rise to Gag protein and protease. C-terminal fragment contains reverse transcriptase and integrase, but the size of individual fragments remains unknown.

copy number of transposable elements may play an important role in the regulation of their transposition. So, the transposition frequency of Ty1 decreased in a copy dependent manner over a 4800-fold range when Ty1 copy number was increased from 0 to 20 [Garfinkel *et al.*, 2003].

The transposition frequency of Ylt1, as determined in this work, reaches $1.4 \cdot 10^{-5}$ when the LTR promoter was used for expression of the marked Ylt1 element. It could be increased 2-fold by overexpression of the marked Ylt1 element under control of the *ICL1* promoter. Observed transposition frequency of the marked Ylt1 element is higher than the frequency of Ty1 transposition, but it is significantly lower than the values reported for Tf1 and Ty3. The following factors may contribute to the relatively low frequency of Ylt1 transposition. First of all, reverse transcriptases of LTR retrotransposons and retroviruses are characterized by low processivity [Huber *et al.*, 1989; Katz and Skalka, 1994], which may result in the inefficient reverse transcription of such long templates as mRNA of Ylt1. Further on, addition of the *SUC2* that was used as a marker gene may affect the level of Ylt1 transposition.

Obtained data indicate that both the available carbon source and the growth temperature have an effect on the transposition frequency of Ylt1. The influence of other factors on the trasnposition of Ylt1 should be investigated in future experiments. Transposition level of the marked Ylt1 element was 7-fold higher in acetate-grown cells than in cells grown on glycerol. The reasons for

the observed differences remain unclear, because no significant differences in the Gag expression were observed between acetate- and glycerol-grown cells.

On the other hand, temperature sensitivity of the retrotranspostion process was reported for other retrotransposons, too. Several authors reported that the transposition rate of Ty1 increase 100-fold when the cells where grown at 15-20°C instead of 30°C [Paquin and Williamson, 1984; Garfinkel *et al.*, 1985; Boeke *et al.*, 1986]. Temperature sensitivity of reverse transcriptase or protease and instability of tRNA-PBS duplex are proposed to cause this difference [Garfinkel *et al.*, 1985; Lauermann and Boeke, 1994; Lawler *et al.*, 2002]. It is currently thought that reduction of protease activity affects the proteolytic processing of Pol protein and thus influences the activity of reverse transcriptase. The activity of Ty1 protease is largely reduced at the temperatures above 30°C and is abolished completely at 37°C [Lawler *et al.*, 2002]. At the same time, the stability of tRNA-PBS duplex preferentially determines the frequency of Ty1 transposition at the temperatures between 20 and 30°C [Lauermann and Boeke, 1997].

Among more than 120 $Suc^{+}Ura^{-}$ clones obtained during this work, only two strains were found in cultures induced at 28°C. Following analysis showed that both strains contained Ylt1 elements arranged in tandem repeats, which were separated by a single LTR. Such arrangement suggests that the integration process of Ylt1 elements was affected. Indeed, similar multimeric arrays of retroelements separated by a single LTR were observed when integrase-mediated integration of Ty1 element was blocked by mutations in the integrase-encoding sequence or in LTR terminal sequences recognized by integrase [Sharon *et al.*, 1994]. In this case, the integration of Ty1 cDNA was mediated by homologous recombination. Further on, it was shown that the retrotransposon Tf2 of fission yeast mobilizes primarily through homologous cDNA recombination independently of a functional Tf2 integrase [Hoff *et al.*, 1998]. Therefore, homologous recombination can be regarded as an alternative mechanism, which mediates the integration of retroelement's cDNA when true integration is blocked.

Genome analysis of *Y. lipolytica* does not reveal the presence of Ylt1 tandem repeats. Therefore, the formation of tandem repeats may be an attribute of the developed *in vivo* transposition assay. Tandem repeat formation may have been caused by the high induction temperature, suggesting the temperature sensitivity of Ylt1 integrase. On the other hand, introduction of the marker *SUC2* gene may have affected the integration process.

Another prominent feature of yeast retrotransposons is their integration specificity. Disruption of protein-coding genes in haploid yeast genomes may have a deleterious effect on the host fitness. Thus, the integration of the majority of known yeast retroelements is targeted to "safe havens", i.e. to the loci whose disruption could be well tolerated by host. Many yeast LTR

retrotransposons are associated with tRNA and other PolIII-transcribed genes [Kim *et al.*, 1998; Hani and Feldmann, 1998]. However, the analysis of the *Y. lipolytica* genome sequence demonstrated that Ylt1 insertions are not grouped together with tRNA genes, as only one of the 29 insertions was found in the vicinity of a $tRNA^{Gly}$ gene. Further on, no association of Ylt1 insertions with centromeres or telomeres was detected. The Ylt1 copies were evenly distributed throughout the genome of the *Y. lipolytica* E150 strain instead. Ylt1 elements were preferentially found in the intergenic regions, often at a significant distance (500-3000 bp) from coding regions. Some bias for the regions upstream of PolII promoters was suggested. Thus, the distribution pattern of Ylt1 resembles to a certain extent the pattern described for Tf1 and Tf2 elements of *Sz. pombe* [Singleton and Levin, 2002; Bowen *et al.*, 2003]. The marked difference between distributions of Ylt1 and fission yeast elements is the distance between the insertion site and the downstream coding region. This distance mostly exceeds 500 bp in the case of Ylt1, whereas Tf1 and Tf2 insertions are clustered at distance of 100-400 bp.

A remarkable feature of Ylt1 is a well-defined sequence specificity of its integration. LTR retrotransposons produce short duplications of target sequence upon their integration, so-called target site duplications (TSDs). Therefore, a sequence specificity of a certain retroelement can be defined based on the analysis of its TSDs. More than 82 % of Ylt1 TSDs in *Y. lipolytica* genome fit the consensus sequence C/T-A-T-G/A. Furthermore, second and third positions of TSDs were absolutely invariantly occupied by A and T in all Ylt1 insertions. Integration of other yeast LTR retrotransposons is usually not sequence-specific, although some bias for A/T bases was observed [Kim et al., 1998]. The observed sequence specificity of the Ylt1 integration emphasizes the isolated position of Ylt1 elements among known yeast retrotransposons.

The consensus sequence of Ylt1 TSDs suggests that the Ylt1 integration may occur into the start codon sequence ATG. However, genome-wide analysis has not revealed any Ylt1 element integrated in such position. The most likely explanation for this observation is that, despite well-defined sequence specificity of Ylt1 integration, other factors also play an important role in the selection of an integration locus. The nature of these factors was not investigated during this work, but it is proposed that chromatin structure or interactions of the Ylt1 integrase with host proteins may contribute to the choice of integration site. Next, the integration of Ylt1 into start codons would often have a deleterious impact on the host fitness, so it is likely that cells with such insertions were rapidly eliminated from the population.

4.4 Sequence analysis of Tyl6

4.4.1 General features

The analysis of a nucleotide sequence of a novel LTR retrotransposon may contribute significantly to understanding of its biology. First of all, LTR retrotransposons are characterized by the presence of several well-conserved structural elements, which play an essential role in their life cycle. Further on, retrotransposons rely on a limited number of self-encoded proteins, whose amino acid sequences can be used to reconstruct the phylogenetic relationships between different retroelements.

The novel element Tyl6 has a size of 5108 bp, and its LTRs are 276 bp long. A putative TATA box was identified in the LTR sequence of Tyl6. LTRs of Tyl6 are flanked by inverted repeats (TGTAA…TTACA), which contain universal dinucleotides (TG…CA).

The structural features of Tyl6 are common for yeast retroelements. Sizes of the full-length element and its LTRs agree well with the size range reported for the majority of known yeast LTR retrotransposons (Table 4.1). Tyl6 is flanked by dinucleotides (TG…CA). These short sequences are ubiquitously present in LTRs of retrotransposons and retroviruses, where they play an important role in the recognition of newly synthesized cDNA by the integrase [Khan *et al.*, 1991].

Transcription of LTR retroelements is generally carried out by host RNA polymerase II. Initiation and termination of the transcription occurs within their LTRs [Boeke and Stoye, 1997]. Thus, LTR elements display many features of promoters recognized by PolII. They contain TATA box and a set of upstream and downstream activating sequences [Boeke and Sandmeyer, 1991].

Table 4.1 Structural properties of the selected LTR retrotransposons from ascomycetous yeast species.

Element	Host organism	Group	Size (bp)	Copy number	ORF1 size (aa)	ORF2 size (aa)	Frame-shift	LTR size (bp)	Terminal inverted repeats
Ty1	*S. cerevisiae*	Ty1/*copia*	5917	32	435	1321	+1	332	TG...CA
Ty3	*S. cerevisiae*	Ty3/*gypsy*	5351	2	285	1262	+1	340	TGTTGTAT...ATACAACA
Ty4	*S. cerevisiae*	Ty1/*copia*	6227	3	363	1440	+1	371	TGTTG...CAACA
Ty5	*S. paradoxus*	Ty1/*copia*	5376	10-15	-	1698	-	251	TGTTGA...TCAACA
Tca2	*C. albicans*	Ty1/*copia*	6426	5-10	324	1576	Stop codon	280	TGTTGG...CCAACA
Tca3	*C. albicans*	Ty3/*gypsy*	6134	0-5	510	1273	+1	313	TGA...TCA
Tse3	*S. exiguus*	Ty3/*gypsy*	6487	10-15	257	1181	+1	945-947	TGTAAC...GTTACA
Tf1	*Sz. pombe*	Ty3/*gypsy*	4941	20-40	-	1330	-	358	TGT...AGA
Tyl6	*Y. lipolytica*	Ty3/*gypsy*	5108	1	252	1235	-1	276	TGTAA...TTACA
Ylt1	*Y. lipolytica*	Ty3/*gypsy*	9453	30-35	-	2621	-	714	TGT...ACA

4.4.2 Primer-binding site

One of the key steps in the life cycle of retroelements is an initiation of reverse transcription. Interestingly, several different modes of the initiation were developed by LTR retrotransposons. Nevertheless, reverse transcription is always initiated a few bases downstream from 5´ LTR, and the analysis of the structure of the corresponding region allows the prediction of the mode of initiation used by certain retroelement.

The majority of known retrotransposons and retroviruses rely on specific host tRNA molecules for the initiation of the reverse transcription. In this case, a short sequence downstream from 5´ LTR (primer-binding site, or PBS) is complementary to a part of a certain host tRNA. However, another mechanism of the reverse transcription initiation was reported for retrotransposons of the Tf1/*sushi* group [Levin, 1995; Levin, 1996; Lin and Levin, 1997a]. These elements initiate their reverse transcription by the unusual self-priming mechanism and do not require tRNA molecules. The structure of primer-binding sites of some LTR retrotransposons (e.g., *skipper*) suggest that the mechanism used by these elements to initiate their reverse transcription differs from described above ones [Leng *et al.*, 1998], but the nature of this novel mechanism remains currently unknown.

The primer-binding site of Tyl6 element was found to be complementary to the predicted *Y. lipolytica* $tRNA^{iMet}$. It contains the characteristic triplet TGG, which interacts with the universal CCA sequence of a tRNA acceptor stem. The Tyl6 PBS also shows a high degree of similarity with PBS of other yeast retroelements, whose reverse transcription is primed with the initiator $tRNA^{iMet}$ (Table 4.2). The length of Tyl6 PBS (12 nt) agrees well into the size range reported for PBS of other yeast LTR retrotransposons [Neuveglise *et al.*, 2002].

Different yeast LTR retrotransposons use various host tRNA molecules to initiate synthesis of their cDNA (Table 4.2). PBS complementary to $tRNA^{iMet}$, $tRNA^{Arg}$, $tRNA^{Asn}$, $tRNA^{Ala}$, $tRNA^{Ile}$ and $tRNA^{Gln}$ have been reported [Voytas and Boeke, 1993; Goodwin and Poulter, 2000; Neuveglise *et al.*, 2002]. Interestingly, some of LTR retrotransposons have a PBS, which is complimentary to an internal fragment of a specific tRNA molecule, so that processing of the tRNA is required prior to the cDNA synthesis [Voytas and Boeke, 1993; Ke *et al.*, 1999]. Despite the great diversity of primer tRNA, many yeast LTR retrotransposons rely on the initiator $tRNA^{iMet}$. They include, for example, Ty1 and Ty3 elements of *S. cerevisiae*, Tse3 element of *S. exiguus* and a number of related elements [Neuveglise *et al.*, 2002]. Tyl6 also belongs to this group. However, the use of the same tRNA species does not necessary reflects close relationships between retroelements. So, both Ty1 and Ty3 elements of *S. cerevisiae* use the same $tRNA^{iMet}$, but they belong to two distinct, distantly related groups.

Table 4.2 Primer tRNA and primer-binding sites (PBS) of the selected LTR retrotransposons from ascomycetous yeast species.

Element	Host organism	Group	Primer tRNA	PBS
Ty1	*S. cerevisiae*	Ty1/*copia*	$tRNA^{iMet}$	TGGTAGCGCC
Ty3	*S. cerevisiae*	Ty3/*gypsy*	$tRNA^{iMet}$	TGGTAGCG
Ty4	*S. cerevisiae*	Ty1/*copia*	$tRNA^{Asn}$	TGGCGACCCCAGTGAGGG
Ty5	*S. cerevisiae*	Ty1/*copia*	$tRNA^{iMet}$ (anticodon stem-loop)	GGTTATGAGCCCT
Tca2	*C. albicans*	Ty1/*copia*	$tRNA^{Arg}$ (anticodon stem-loop)	GATTAGAAGTC
Tca13	*C. albicans*	Ty3/*gypsy*	$tRNA^{Ala}$	TGGTGGACGAGATGAGAG
Tse1	*S. exiguus*	Ty1/*copia*	$tRNA^{iMet}$	TGGTAGCGCCGC
Tse3	*S. exiguus*	Ty3/*gypsy*	$tRNA^{iMet}$	TGGTAGCGCCGC
Tyl6	*Y. lipolytica*	Ty3/*gypsy*	$tRNA^{iMet}$	TGGTAGCGATGC
Ylt1	*Y. lipolytica*	Ty3/*gypsy*	$tRNA^{Ala}$	TGGTGGACGACACC

4.4.3 Organization of coding sequences

4.4.3.1 Arrangement of *gag* and *pol* reading frames

Most of the known LTR retrotransposons have two coding regions, *gag* and *pol*, and there are several basic modes of their organization. As mentioned above, a molar excess of the *gag* gene product is required for the propagation of LTR retrotransposons. Therefore, many retroelements have *gag* and *pol* genes in different reading frames, so that a ribosomal frame-shift is required for the translation of the *pol* gene. However, this is not the only way to achieve the required Gag:Pol ratio. Some retroelements have *gag* and *pol* reading frames separated by a stop codon (ribosomal read-through is required in this case). Alternatively, elements can express *gag* and *pol* in a single reading frame, so that the regulation of Gag:Pol ratio occurs post-translationally.

In the case of Tyl6, *gag* and *pol* ORFs are separated by a -1 translational frame-shift. This type of coding sequence arrangement commonly occurs among Ty3/*gypsy* retrotransposons [Gao *et al.*, 2003], however, it is the first example of the -1 frame-shift in the retroelements of ascomycetous yeasts. At the same time, a -1 frame-shift is commonly found in LTR retrotransposons of ascomycetous filamentous fungi [Gao *et al.*, 2003].

Interestingly, the apparent avoidance of -1 frame-shifting by yeast LTR retrotransposons is not due to the inability of the yeast translation machinery to perform a -1 frame-shift. Indeed, the yeast "killer" virus L-A uses a -1 frame-shift [Dinman *et al.*, 1991], and HIV-1 *gag-pol* polyprotein, which requires -1 frame-shift for its expression, could be correctly translated in yeast [Wilson *et al.*, 1988]. Why yeast retrotransposons of the Ty3/*gypsy* group preferentially

use a +1 frame-shift remains currently unknown. One of the possible explanations for the observed bias is a limited number of yeast Ty3/*gypsy* retrotransposons studied so far. Further on, many of them are presented in public databases by partial sequences only, so that no conclusions about organization of their coding sequences could be drawn. The presence of a -1 frame-shift in Tyl6 indicates that there is no absolute requirement for a +1 frame-shift in yeast retroelements. Definitely, further work is required to clarify this issue.

4.4.3.2 Proteins encoded by Tyl6

Analysis of the nucleotide sequence of Tyl6 revealed the presence of two reading frames, *gag* and *pol*. The first reading frame encodes a predicted structural Gag protein. Regions showing a similarity with aspartic protease, reverse transcriptase, RNaseH and integrase were detected in the deduced polyprotein encoded by the *pol* gene. Thus, Tyl6 encodes a set of proteins, which is common for the majority of LTR retrotransposons.

Gag protein is a main structural component of virus-like particles (VLPs), where the replication of LTR retrotransposons occurs. Further on, its C-terminal fragment, so-called nucleocapsid protein (NC), is presumably involved in the scaffolding of genomic RNA inside VLPs [Roth, 2000]. Gag proteins of different retroelements often show a limited degree of sequence conservation, so the only well-conserved region is a zinc-finger motif of NC proteins. However, even this motif is not necessarily essential for the function of a Gag protein, as it was not found in the amino acid sequence of the Gag protein of the Ty1 element [Clare and Farabaugh, 1985].

The Gag protein of Tyl6 element has a lot in common with known Gag proteins. Its predicted size (252 aa; 29 kDa) fits well into the size range reported for other Ty3/*gypsy* retrotransposons (Table 4.1). Further on, it contains a well-defined C-terminal zinc-finger motif (Fig. 3.12), and a motif common for Ty3/*gypsy* Gag proteins was detected in its N-terminal part. These results confirm that the retrotransposon Tyl6 is closely related to members of Ty3/*gypsy* group. However, a low level of sequence conservation among different Gag proteins hampers their use for the reconstruction of phylogenetic relationships between retroelements.

A further protein encoded by Tyl6 element is an aspartic protease. These enzymes are characterized by the presence of an aspartate residue in their active center. They perform the proteolytic processing of polyproteins produced by retroelemets. The processing of the Gag-Pol fusion protein generally results in the release of structural CA and NC proteins, a protease, an integrase, and a reverse transcriptase. Aspartic proteases of LTR retrotransposons and retroviruses have the motif (hydrophobic residue)$_2$-D-S/T-G-A/S in common [Dunn *et al.*, 2002]. However, in the case of Tyl6, this conserved motif contains a cysteine residue instead of

canonical serine/threonine (Fig. 3.12). Interestingly, the same substitution (S → C) was observed in the conserved motif of RNaseH of Tyl6 (Fig. 3.12). Described motifs of protease and RNaseH are characterized by a very high level of sequence conservation, so one wonders whether corresponding enzymes of Tyl6 element retained their enzymatic activity. Notably, conserved motifs of protease and RNaseH of Ylt1, another Ty3/*gypsy* retrotransposon from *Y. lipolytica* genome, do not deviate from the canonical models [Senam, 2004].

Reverse transcriptases are the most conserved proteins encoded by retroelements [Xiong and Eickbush, 1990]. Hence, a predicted reverse transcriptase (RT) of Tyl6 element is highly similar to reverse transcriptases of other Ty3/*gypsy* retrotransposons (Fig. 3.12), supporting the placement of Tyl6 element into this group. A phylogenetic analysis of Tyl6 relationships was based on the comparison of RT sequences.

The integrase domain occupies the C-terminal part of the Pol polyprotein. Three integrase subdomains are generally recognized [Khan *et al.*, 1991; Malik and Eickbush, 1999]. The N-terminal subdomain contains a zinc-finger motif HH-CC implicated in binding to LTR sequences. The central core domain contains the catalytic D,D_{35}-E motif. The function of the C-terminal subdomain is less understood, but its possible role in the target specificity of retroelement's integration was proposed [Malik and Eickbush, 1999]. This domain is the least conserved of the three subdomains, however, the majority of Ty3/*gypsy* retrotransposons and vertebrate retroviruses are characterized by the presence of a GPY/F module in the C-terminal region of their integrases [Malik and Eickbush, 1999].

All three subdomains were identified in the deduced amino acid sequence of Tyl6 integrase. The presence of the GPY/F module in the C-terminal subdomain of Tyl6 integrase supports its position among Ty3/*gypsy* elements. Performed phylogenetic analysis placed Tyl6 among chromoviruses, a group of Ty3/*gypsy* retrotransposons, which are characterized by the presence of a chromodomain within their integrases [Malik and Eickbush, 1999; Kordis, 2005]. However, none of the tools used for the prediction of functional domains in protein sequences revealed the presence of chromodomain in the Tyl6 integrase. Interestingly, other yeast retrotransposons closely related to chromoviruses (Ty3 and Tse3) lack the chromodomain, too. However, C-terminal parts of their integrases displays a significant sequence similarity (Fig 3.16). Another characteristic feature of yeast chromoviruses known so far is a striking specificity of their integration. All three elements preferentially target genes transcribed by RNA polymerase III, and their insertions are almost exclusively located within 1-2 bp upstream or downstream of the transcription start. Therefore, it is proposed that the observed target specificity of yeast chromoviruses has evolved as a result of their adaptation for the compact genomes of their hosts,

and this adaptation was accompanied by the specialization of C-terminal domains of their integrases.

4.5 Distribution of Ylt1 and Tyl6 among *Y. lipolytica* strains

Natural isolates of *Y. lipolytica* appeared to have widely divergent genetic structure. A pronounced chromosome length polymorphism and a poor conservation of linkage groups between different lines were reported [Barth and Gaillardin, 1996]. The differences between different lineages of *Y. lipolytica* are reflected in the distribution of the retrotransposon Ylt1. Until now, Ylt1 was detected only in *Y. lipolytica* strains derived from the Wickerham's isolate *Y. lipolytica* YB423 [Juretzek *et al.*, 2001; Senam, 2004]. *Y. lipolytica* YB423 is a natural diploid isolate, and its meiotic segregant YB423-12 was widely used in several inbreeding programs. Therefore, some important laboratory strains of *Y. lipolytica* (e.g. E150, E129 and B204-12C-20) also contain Ylt1 elements in their genomes. No other natural isolates (e.g. H222 and W29) tested so far contain this element.

Results obtained in the course of this work indicated that Tyl6 has the same distribution pattern as Ylt1, namely, it is present exclusively in YB423-12-derived strains, but not in strains of independent origin. Further on, this distribution pattern was reported for a novel DNA transposon found in *Y. lipolytica* genome, Mutator-like element Mutyl [Neuveglise *et al.*, 2005], too. On the other hand, the LTR retrotransposon Tyl3 was found in the genome of *Y. lipolytica* W29, but it is absent in strain E150. Finally, non-LTR retrotransposon Ylli is ubiquitously present in all *Y. lipolytica* strains tested so far (Table 4.3) [Casaregola *et al.*, 2002].

Table 4.3 Distribution of transposable elements among *Y. lipolytica* strains

Element	Type	*Yarrowia lipolytica* strains				References
		YB423-12	**H222**	**W29, PO1d**	**E150**	
Ylt1	Ty3/*gypsy*	+	-	-	+	Juretzek *et al.*, 2001; Senam, 2004.
Tyl6	Ty3/*gypsy*	+	-	-	+	This work
Tyl3	Ty3/*gypsy*	n.d.	n.d.	+	-	Casaregola *et al.*, 2000
Ylli	LINE	+	+	+	+	Casaregola *et al.*, 2002
Mutyl	Mutator	+	-	-	+	Neuveglise *et al.*, 2005

Two models of colonization of the *Y. lipolytica* genome by transposable elements are proposed here. The first model suggests that different lineages of *Y. lipolytica* acquired their transposable elements independently. The only exception constitutes the non-LTR retrotransposon Ylli, which is present in all *Y. lipolytica* strains studied so far. Non-LTR retrotransposons are characterized by a vertical mode of transmission, and they are generally accepted as the most ancient group of eukaryotic retroelements [Malik *et al.*, 1999]. Therefore it is likely that Ylli occupied the *Y. lipolytica* genome prior to the divergence of individual lines. On the other hand, independent invasion of other known transposable elements that occurred after the divergence of modern *Y. lipolytica* lineages would explain the existing differences between different isolates.

The second model proposes that divergence of modern *Y. lipolytica* lineages occurred after acquisition of transposable elements by the *Y. lipolytica* genome. The observed differences between individual isolates in this case may be explained by different rates of "junk DNA" elimination. A strong argument against this model is a requirement for strikingly different rates of genome evolution in various strains of the same species. However, most natural isolates of *Y. lipolytica* are haploid, whereas Wickerham's isolate YB423 was diploid. A selective pressure against new insertions of transposable elements in a diploid organism is significantly lower than in haploid ones. Therefore, it is possible that numerous transposable elements persist in the genome of diploid strains, whereas they were eliminated from haploid genomes. This model is also supported by the differences in retrotransposon diversity observed between different yeast species. Baker's yeast *S. cerevisiae* exists a larger part of their life cycle in a haploid phase, and it is characterized by a limited number of transposable elements. For example, this species does not contain non-LTR retrotransposons or DNA transposons. On the other hand, *C. albicans* is a diploid yeast species, which is characterized by a remarkable diversity of transposable elements. Thus, the *C. albicans* genome contains 34 LTR families and several families of non-LTR retrotransposons. These data indicate that the genome ploidy level may contribute significantly to the diversity of transposable elements in the genome.

The retrotransposons Ylt1 and Tca3 found in the genomes of *Y. lipolytica* and *C. albicans*, respectively, were placed together on one branch in a phylogenetic tree of Ty3/*gypsy* elements, but no related elements were found in the genomes of other studied yeast species. This fact raises an intriguing possibility that both Ylt1 and Tca3 are remnants of an ancient group of yeast retrotransposons, whose members were eliminated from the genomes of other species with a higher rate of genome evolution. If this suggestion is correct, than further elements belonging to this group should be found in yeast species existing predominantly in a diploid phase.

Without a doubt, analysis of further natural isolates of *Y. lipolytica* will be required to clarify the evolutionary history of transposable elements in the *Y. lipolytica* genome. Nevertheless, it is already clear that the diversity of transposable elements in the *Y. lipolytica* genome is unique for hemiascomycetous yeasts studied so far. *Y. lipolytica* is the only currently known yeast species, whose genome contains DNA transposons. Further on, yeast non-LTR retrotransposons are currently found only in the genomes of *C. albicans* and *Y. lipolytica*. Finally, known LTR retrotransposons of *Y. lipolytica* are only distantly related. Tyl3 and Tyl6 elements belong to the large group of chromoviruses, whereas the unusual element Ylt1 displays some similarity with *C. albicans* LTR retrotransposons.

5 Summaries

The genome of the dimorphic fungus *Yarrowia lipolytica* is characterized by the diversity of transposable elements, which is unique among hemiascomycetous yeasts. The LTR retrotransposon Ylt1 was the first element described from the genome of *Y. lipolytica*. Its sequence analysis has demonstrated that Ylt1 differs significantly from the yeast retrotransposons characterized previously [Senam *et al.*, 2004]. The aim of this work was the further characterization of the properties of Ylt1, with a focus on the Ylt1 interactions with its host. Following results were obtained during this work:

1. The HA-tagged Gag protein of Ylt1 was expressed under control of the strong *ICL1* promoter and under control of the weak LTR promoter. Presence of the Gag protein in *Y. lipolytica* cells was confirmed in both cases. The tagged Gag protein expressed under the control of the LTR promoter was used to study the regulation of the LTR promoter activity. It was shown that the Gag protein is produced in log-phase cells on different carbon sources, and that it is degraded in stationary phase cells. Such expression pattern was not previously reported for other yeast retrotransposons.
2. The sizes of the Gag protein of Ylt1 (78-80 kDa) and of its putative precursor (about 140 kDa) were estimated. A putative precursor of the integrase was also observed, but the integrase alone has not been detected in this work. The model for the processing of the polyprotein encoded by Ylt1 was created based on the data obtained in these experiments. Also, it turned out that the Gag protein of Ylt1 is 1.5- to 3-fold larger than Gag proteins of other yeast retrotransposons. Thus, a size differences between Ylt1 and other yeast retrotransposons are reflected in the sizes of Gag proteins encoded by these elements.
3. A system allowing the detection of Ylt1 transposition events *in vivo* was developed. The transposition of the marked Ylt1 element from the autonomous plasmid into the genome of *Y. lipolytica* was demonstrated. Also, it was shown that the available carbon source and the growth temperature have an effect on the rate of this transposition. Formation of tandem repeats by newly inserted Ylt1 elements was observed when the cultivation was performed at increased temperature. It is suggested that integrase function was affected in this case, and that the integration was mediated by homologous recombination.
4. Analysis of the Ylt1 insertion specificity and of the Ylt1 distribution in the genome of *Y. lipolytica* E150 was done. The remarkable sequence specificity of Ylt1 insertions, which is unusual for LTR retrotransposons, was revealed during this analysis. The consensus sequence of the Ylt1 target site (C/T-A-T-G/A) suggested by Schmid-Berger *et al.* [1994]

was confirmed. Also, it was shown that Ylt1 insertions are found mainly in intergenic regions, often at a significant distance (>500 bp) from the next reading frame. No association of Ylt1 insertions with tRNA genes was observed.

5. The novel Ty3/*gypsy* element Tyl6 was found in the genome of *Y. lipolytica* E150. The sequence analysis of this element was performed. It was shown that structural properties of Tyl6 resemble the properties of the Ty3 element of *S. cerevisiae*. However, two reading frames of Tyl6 (*gag* and *pol*) are separated by -1 frame-shift, which was not previously reported for retrotransposons of hemiascomycetous yeasts. Phylogenetic analysis placed Tyl6 within chromoviruses, and the Tse3 element of *S. exiguus* was shown to be the closest relative of Tyl6. The distribution of Tyl6 among *Y. lipolytica* strains was analyzed. Interestingly, the novel element was found only in strains derived from the strain YB423-12. The strains of independent origin included in the analysis were shown to be Tyl6-free. The same distribution was previously reported for the retrotransposon Ylt1 and for the DNA transposon Mutyl. Two models of the evolution of transposable elements in *Y. lipolytica* genome were proposed based on these results. A significance of the natural diploid isolate YB423 for the maintenance of transposable elements of *Y. lipolytica* was emphasized.

6 References

Adams, S. E., Mellor, J., Gull, K., Sim, R. B., Tuite, M. F., Kingsman, S. M. and Kingsman, A. J. (1987). The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell 49(1): 111-9.

Akins, R. A., Grant, D. M., Stohl, L. L., Bottorff, D. A., Nargang, F. E. and Lambowitz, A. M. (1988). Nucleotide sequence of the Varkud mitochondrial plasmid of *Neurospora* and synthesis of a hybrid transcript with a 5' leader derived from mitochondrial RNA. J Mol Biol 204(1): 1-25.

Akins, R. A., Kelley, R. L. and Lambowitz, A. M. (1986). Mitochondrial plasmids of *Neurospora*: integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. Cell 47(4): 505-16.

Akins, R. A., Kelley, R. L. and Lambowitz, A. M. (1989). Characterization of mutant mitochondrial plasmids of *Neurospora spp*. that have incorporated tRNAs by reverse transcription. Mol Cell Biol 9(2): 678-91.

Aksoy, S., Lalor, T. M., Martin, J., Van der Ploeg, L. H. and Richards, F. F. (1987). Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. Embo J 6(12): 3819-26.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17): 3389-402.

Anaya, N. and Roncero, M. I. (1995). Skippy, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. Mol Gen Genet 249(6): 637-47.

Arkhipova, I. R., Pyatkov, K. I., Meselson, M. and Evgen'ev, M. B. (2003). Retroelements containing introns in diverse invertebrate taxa. Nat Genet 33(2): 123-4.

Atwood, A., Lin, J. H. and Levin, H. L. (1996). The retrotransposon Tf1 assembles virus-like particles that contain excess Gag relative to integrase because of a regulated degradation process. Mol Cell Biol 16(1): 338-46.

Augstein, A., Barth, K., Gentsch, M., Kohlwein, S. D. and Barth, G. (2003). Characterization, localization and functional analysis of Gpr1p, a protein affecting sensitivity to acetic acid in the yeast *Yarrowia lipolytica*. Microbiology 149(Pt 3): 589-600.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1997). Current Protocols in Molecular Biology. New York, John Wiley & Sons, Inc.

Avedisov, S. N., Cherkasova, V. A. and Il'in Iu, V. (1990). [Features of the structural organization of the MDG1 retrotransposon of *Drosophila*, revealed during its sequencing]. Genetika 26(11): 1905-14.

Avramova, Z., Tikhonov, A., SanMiguel, P., Jin, Y. K., Liu, C., Woo, S. S., Wing, R. A. and Bennetzen, J. L. (1996). Gene identification in a complex chromosomal continuum by local genomic cross-referencing. Plant J 10(6): 1163-8.

Barth, G., Beckerich, J.-M., Dominguez, A., Kerscher, S., Ogrydziak, D., Titorenko, V. and

Gaillardin, C. (2003). Functional genetics of *Yarrowia lipolytica*. In: Functional Genetics of Industrial Yeasts. J. H. de Winde (ed.) Berlin, Heidelberg, Springer. 2**:** 227-271.

Barth, G. and Gaillardin, C. (1996). *Yarrowia lipolytica*. In: Nonconventional yeasts in biotechnology: a handbook. K. Wolf (ed.) Berlin, Heidelberg, Springer**:** 313-388.

Barth, G. and Gaillardin, C. (1997). Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol Rev 19(4): 219-37.

Beeman, R. W., Thomson, M. S., Clark, J. M., DeCamillis, M. A., Brown, S. J. and Denell, R. E. (1996). Woot, an active gypsy-class retrotransposon in the flour beetle, *Tribolium castaneum*, is associated with a recent mutation. Genetics 143(1): 417-26.

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(23): 4709-16.

Boeke, J. D. and Devine, S. E. (1998). Yeast retrotransposons: finding a nice quiet neighborhood. Cell 93(7): 1087-9.

Boeke, J. D. and Sandmeyer, S. B. (1991). Yeast transposable elements. In: The molecular and cellular biology of the yeast *Saccharomyces*. J. Broach, E. Jones and J. Pringle (eds.) Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. 1**:** 193-261.

Boeke, J. D. and Stoye, J. P. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus (eds.) New York, Cold Spring Harbor Laboratory Press**:** 343-435.

Boeke, J. D., Styles, C. A. and Fink, G. R. (1986). *Saccharomyces cerevisiae* SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol Cell Biol 6(11): 3575-81.

Bowen, N. J., Jordan, I. K., Epstein, J. A., Wood, V. and Levin, H. L. (2003). Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of *Schizosaccharomyces pombe*. Genome Res 13(9): 1984-97.

Bradshaw, V. A. and McEntee, K. (1989). DNA damage activates transcription and transposition of yeast Ty retrotransposons. Mol Gen Genet 218(3): 465-74.

Brierley, C. and Flavell, A. J. (1990). The retrotransposon *copia* controls the relative levels of its gene products post-transcriptionally by differential expression from its two major mRNAs. Nucleic Acids Res 18(10): 2947-51.

Butler, M., Goodwin, T., Simpson, M., Singh, M. and Poulter, R. (2001). Vertebrate LTR retrotransposons of the Tf1/*sushi* group. J Mol Evol 52(3): 260-74.

Cambareri, E. B., Aisner, R. and Carbon, J. (1998). Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. Mol Cell Biol 18(9): 5465-77.

Cameron, J. R., Loh, E. Y. and Davis, R. W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. Cell 16(4): 739-51.

Cappello, J., Handelsman, K., Cohen, S. M. and Lodish, H. F. (1985). Structure and regulated transcription of DIRS-1: an apparent retrotransposon of *Dictyostelium discoideum*. Cold Spring Harb Symp Quant Biol 50: 759-67.

Capy, P., Bazin, C., Higuet, D. and Langin, T. (1998). Dynamics and Evolution of Transposable Elements. New York, Springer.

Casaregola, S., Neuveglise, C., Bon, E. and Gaillardin, C. (2002). Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. Mol Biol Evol 19(5): 664-77.

Casaregola, S., Neuveglise, C., Lepingle, A., Bon, E., Feynerol, C., Artiguenave, F., Wincker, P. and Gaillardin, C. (2000). Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. FEBS Lett 487(1): 95-100.

Casaregola, S., Neuveglise, C., Wincker, P., Anthouard, V., Weissenbach, J., Gaillardin, C. (2003). The dimorphic yeast *Yarrowia lypolytica* genome sequencing project. YEAST 20: S277.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. Trends Genet 7(5): 145-8.

Chalker, D. L. and Sandmeyer, S. B. (1990). Transfer RNA genes are genomic targets for de novo transposition of the yeast retrotransposon Ty3. Genetics 126(4): 837-50.

Chalker, D. L. and Sandmeyer, S. B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev 6(1): 117-28.

Chalvet, F., Grimaldi, C., Kaper, F., Langin, T. and Daboussi, M. J. (2003). Hop, an active Mutator-like element in the genome of the fungus *Fusarium oxysporum*. Mol Biol Evol 20(8): 1362-75.

Chavanne, F., Zhang, D. X., Liaud, M. F. and Cerff, R. (1998). Structure and evolution of Cyclops: a novel giant retrotransposon of the Ty3/*gypsy* family highly amplified in pea and other legume species. Plant Mol Biol 37(2): 363-75.

Chen, H., Bechtel, M. K., Shi, Y., Phipps, A., Mathes, L. E., Hayes, K. A. and Roy-Burman, P. (1998). Pathogenicity induced by feline leukemia virus, Rickard strain, subgroup A plasmid DNA (pFRA). J Virol 72(9): 7048-56.

Ciriacy, M. (1995). Yeast retrotransposons. In: The Mycota II. Genetics and Biotechnology. U. Kück (ed.) Berlin, Heidelberg, Springer**:** 227-245.

Clare, J. and Farabaugh, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc Natl Acad Sci USA 82(9): 2829-33.

Coffin, J. M. (1992). Structure and classification of retroviruses. In: The Retroviridae. J. A. Levy (ed.) New York, London, Plenum Press. 1**:** 19-49.

Covey, S. N. (1986). Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res 14(2): 623-33.

Cristofari, G., Ficheux, D. and Darlix, J. L. (2000). The GAG-like protein of the yeast Ty1

retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. J Biol Chem 275(25): 19210-7.

Curcio, M. J. and Derbyshire, K. M. (2003). The outs and ins of transposition: from mu to kangaroo. Nat Rev Mol Cell Biol 4(11): 865-77.

Curcio, M. J., Hedge, A. M., Boeke, J. D. and Garfinkel, D. J. (1990). Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. Mol Gen Genet 220(2): 213-21.

Daboussi, M. J. and Capy, P. (2003). Transposable elements in filamentous fungi. Annu Rev Microbiol 57: 275-99.

Dang, V. D. and Levin, H. L. (2000). Nuclear import of the retrotransposon Tf1 is governed by a nuclear localization signal that possesses a unique requirement for the FXFG nuclear pore factor Nup124p. Mol Cell Biol 20(20): 7798-812.

de Chastonay, Y., Felder, H., Link, C., Aeby, P., Tobler, H. and Muller, F. (1992). Unusual features of the retroid element PAT from the nematode *Panagrellus redivivus*. Nucleic Acids Res 20(7): 1623-8.

Devine, S. E. and Boeke, J. D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev 10(5): 620-33.

Dhundale, A., Furuichi, T., Inouye, M. and Inouye, S. (1988). Mutations that affect production of branched RNA-linked msDNA in *Myxococcus xanthus*. J Bacteriol 170(12): 5620-4.

Dickson, L., Huang, H. R., Liu, L., Matsuura, M., Lambowitz, A. M. and Perlman, P. S. (2001). Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. Proc Natl Acad Sci USA 98(23): 13207-12.

Diez, J., Beguiristain, T., Le Tacon, F., Casacuberta, J. M. and Tagu, D. (2003). Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: evolutionary relationships and use as molecular markers. Curr Genet 43(1): 34-44.

Dinman, J. D., Icho, T. and Wickner, R. B. (1991). A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. Proc Natl Acad Sci USA 88(1): 174-8.

Diolez, A., Marches, F., Fortini, D. and Brygoo, Y. (1995). Boty, a long-terminal-repeat retroelement in the phytopathogenic fungus *Botrytis cinerea*. Appl Environ Microbiol 61(1): 103-8.

Dobinson, K. F., Harris, R. E. and Hamer, J. E. (1993). Grasshopper, a long terminal repeat (LTR) retroelement in the phytopathogenic fungus *Magnaporthe grisea*. Mol Plant Microbe Interact 6(1): 114-26.

Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16(13): 6127-45.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I. *et al.* (2004). Genome evolution in yeasts. Nature 430(6995): 35-44.

Duncan, L., Bouckaert, K., Yeh, F. and Kirk, D. L. (2002). kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons. Genetics 162(4): 1617-30.

Dunn, B. M., Goodenow, M. M., Gustchina, A. and Wlodawer, A. (2002). Retroviral proteases. Genome Biol 3(4): REVIEWS3006.

Dzhumgaliev, E. B., Mazo, A. M., Baev, A. A., Jr., Gorelova, T. V. and Arkhipova, I. R. (1986). [Structure of long terminal repeats of transcriptionally active and inactive copies of the mobile dispersed gene MDG3 in *Drosophila melanogaster*]. Genetika 22(3): 368-77.

Eickbush, T. H. (1997). Telomerase and retrotransposons: which came first? Science 277(5328): 911-2.

Evgen'ev, M. B., Corces, V. G. and Lankenau, D. H. (1992). Ulysses transposable element of *Drosophila* shows high structural similarities to functional domains of retroviruses. J Mol Biol 225(3): 917-24.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. and Bairoch, A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res 30(1): 235-8.

Farabaugh, P. J. (1995). Post-transcriptional regulation of transposition by Ty retrotransposons of *Saccharomyces cerevisiae*. J Biol Chem 270(18): 10361-4.

Farabaugh, P. J. (1996). Programmed translational frameshifting. Annu Rev Genet 30: 507-28.

Farman, M. L., Tosa, Y., Nitta, N. and Leong, S. A. (1996). MAGGY, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. Mol Gen Genet 251(6): 665-74.

Fauci, A. S. and Desrosiers, R. C. (1997). Pathogenesis of HIV and SIV. In: Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus (eds.) New York, Cold Spring Harbor Laboratory Press**:** 587-635.

Faure, E., Best-Belpomme, M. and Champion, S. (1996). UVB irradiation upregulation of the Drosophila 1731 retrotransposon LTR requires the same short sequence of U3 region in a human epithelial cell line as in *Drosophila* cells. Photochem Photobiol 64(5): 807-13.

Felsenstein, J. (1993). PHYLIP manual, version 3.5. Seattle, University of Washington.

Fingerman, E. G., Dombrowski, P. G., Francis, C. A. and Sniegowski, P. D. (2003). Distribution and sequence analysis of a novel Ty3-like element in natural *Saccharomyces paradoxus* isolates. Yeast 20(9): 761-70.

Flavell, A. J., Ruby, S. W., Toole, J. J., Roberts, B. E. and Rubin, G. M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. Proc Natl Acad Sci USA 77(12): 7107-11.

Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D. M., Yaver, D., Xuan, J. W., Peito, A., Ribet, A. M., Feynerol, C. and et al. (1993). Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. Proc Natl Acad Sci USA 90(11): 4912-6.

Frame, I. G., Cutfield, J. F. and Poulter, R. T. (2001). New BEL-like LTR-retrotransposons in

Fugu rubripes, *Caenorhabditis elegans*, and *Drosophila melanogaster*. Gene 263(1-2): 219-30.

Friant, S., Heyman, T., Bystrom, A. S., Wilhelm, M. and Wilhelm, F. X. (1998). Interactions between Ty1 retrotransposon RNA and the T and D regions of the tRNA(iMet) primer are required for initiation of reverse transcription in vivo. Mol Cell Biol 18(2): 799-806.

Frommer, G., Schuh, R. and Jackle, H. (1994). Localized expression of a novel micropia-like element in the blastoderm of *Drosophila melanogaster* is dependent on the anterior morphogen bicoid. Chromosoma 103(2): 82-9.

Gabus, C., Ficheux, D., Rau, M., Keith, G., Sandmeyer, S. and Darlix, J. L. (1998). The yeast Ty3 retrotransposon contains a 5'-3' bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. Embo J 17(16): 4873-80.

Gao, X., Havecker, E. R., Baranov, P. V., Atkins, J. F. and Voytas, D. F. (2003). Translational recoding signals between *gag* and *pol* in diverse LTR retrotransposons. Rna 9(12): 1422-30.

Garfinkel, D. J. (1992). Retroelements in microorganisms. In: The Retroviridae. J. A. Levy (ed.) New York, London, Plenum Press. 1**:** 107-158.

Garfinkel, D. J., Boeke, J. D. and Fink, G. R. (1985). Ty element transposition: reverse transcriptase and virus-like particles. Cell 42(2): 507-17.

Garfinkel, D. J., Nyswaner, K., Wang, J. and Cho, J. Y. (2003). Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics 165(1): 83-99.

Goodwin, T. J., Dalle Nogare, D. E., Butler, M. I. and Poulter, R. T. (2003). Ty3/gypsy-like retrotransposons in *Candida albicans* and *Candida dubliniensis*: Tca3 and Tcd3. Yeast 20(6): 493-508.

Goodwin, T. J., Ormandy, J. E. and Poulter, R. T. (2001). L1-like non-LTR retrotransposons in the yeast *Candida albicans*. Curr Genet 39(2): 83-91.

Goodwin, T. J. and Poulter, R. T. (2000). Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res 10(2): 174-91.

Goodwin, T. J. and Poulter, R. T. (2001). The DIRS1 group of retrotransposons. Mol Biol Evol 18(11): 2067-82.

Goodwin, T. J. and Poulter, R. T. (2001). The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. Yeast 18(9): 865-80.

Goodwin, T. J. and Poulter, R. T. (2002). A group of deuterostome Ty3/*gypsy*-like retrotransposons with Ty1/*copia*-like pol-domain orders. Mol Genet Genomics 267(4): 481-91.

Goyon, C., Rossignol, J. L. and Faugeron, G. (1996). Native DNA repeats and methylation in *Ascobolus*. Nucleic Acids Res 24(17): 3348-56.

Hamer, J. E., Farrall, L., Orbach, M. J., Valent, B. and Chumley, F. G. (1989). Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. Proc Natl Acad Sci USA 86(24): 9981-5.

Hani, J. and Feldmann, H. (1998). tRNA genes and retroelements in the yeast genome. Nucleic Acids Res 26(3): 689-96.

Hansen, L. J., Chalker, D. L., Orlinsky, K. J. and Sandmeyer, S. B. (1992). Ty3 GAG3 and POL3 genes encode the components of intracellular particles. J Virol 66(3): 1414-24.

Hansen, L. J., Chalker, D. L. and Sandmeyer, S. B. (1988). Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol Cell Biol 8(12): 5245-56.

Hansen, L. J. and Sandmeyer, S. B. (1990). Characterization of a transpositionally active Ty3 element and identification of the Ty3 integrase protein. J Virol 64(6): 2599-607.

Hauber, J., Nelbock-Hochstetter, P. and Feldmann, H. (1985). Nucleotide sequence and characteristics of a Ty element from yeast. Nucleic Acids Res 13(8): 2745-58.

Havecker, E. R., Gao, X. and Voytas, D. F. (2004). The diversity of LTR retrotransposons. Genome Biol 5(6): 225.

He, C., Nourse, J. P., Kelemu, S., Irwin, J. A. and Manners, J. M. (1996). CgT1: a non-LTR retrotransposon with restricted distribution in the fungal phytopathogen *Colletotrichum gloeosporioides*. Mol Gen Genet 252(3): 320-31.

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(11): 6839-52.

Hoffman, C. S. and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57(2-3): 267-72.

Holton, N. J., Goodwin, T. J., Butler, M. I. and Poulter, R. T. (2001). An active retrotransposon in *Candida albicans*. Nucleic Acids Res 29(19): 4014-24.

Huber, H. E., McCoy, J. M., Seehra, J. S. and Richardson, C. C. (1989). Human immunodeficiency virus 1 reverse transcriptase. Template binding, processivity, strand displacement synthesis, and template switching. J Biol Chem 264(8): 4669-78.

Hull, R. (2001). *Caulimoviridae* (Plant pararetroviruses). Encyclopedia of Life Sciences. London, Nature Publishing Group.

Ichiyanagi, K., Beauregard, A., Lawrence, S., Smith, D., Cousineau, B. and Belfort, M. (2002). Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. Mol Microbiol 46(5): 1259-72.

Ikeda, K., Nakayashiki, H., Takagi, M., Tosa, Y. and Mayama, S. (2001). Heat shock, copper sulfate and oxidative stress activate the retrotransposon MAGGY resident in the plant pathogenic fungus *Magnaporthe grisea*. Mol Genet Genomics 266(2): 318-25.

Inouye, S., Sunshine, M. G., Six, E. W. and Inouye, M. (1991). Retronphage phi R73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. Science 252(5008): 969-71.

Inouye, S., Yuki, S. and Saigo, K. (1986). Complete nucleotide sequence and genome

organization of a *Drosophila* transposable genetic element, 297. Eur J Biochem 154(2): 417-25.

Janetzky, B. and Lehle, L. (1992). Ty4, a new retrotransposon from *Saccharomyces cerevisiae*, flanked by tau-elements. J Biol Chem 267(28): 19798-805.

Jiang, Y. W. (2002). Transcriptional cosuppression of yeast Ty1 retrotransposons. Genes Dev 16(4): 467-78.

Jilbert, A. R. and Mason, W. S. (2002). Hepatitis B virus. Encyclopedia of Life Sciences. London, Nature Publishing Group.

Juretzek, T. (1999). Entwicklung von Wirts-Vektor-Systemen zur heterologen Expression von Proteinen in der nichtkonventionellen Hefe *Yarrowia lipolytica* und ihre Anwendung für die Cytochrom P450-katalysierte Stoffumwandlung. Dissertation, Institut für Mikrobiologie, Fakultät Mathematik und Naturwissenschaften, Technische Universität Dresden.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G. and Nicaud, J. (2001). Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. Yeast 18(2): 97-113.

Kaneko, I., Tanaka, A. and Tsuge, T. (2000). REAL, an LTR retrotransposon from the plant pathogenic fungus *Alternaria alternata*. Mol Gen Genet 263(4): 625-34.

Katz, R. A. and Skalka, A. M. (1994). The retroviral enzymes. Annu Rev Biochem 63: 133-73.

Kawakami, K., Pande, S., Faiola, B., Moore, D. P., Boeke, J. D., Farabaugh, P. J., Strathern, J. N., Nakamura, Y. and Garfinkel, D. J. (1993). A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in *Saccharomyces cerevisiae*. Genetics 135(2): 309-20.

Kazazian, H. H., Jr., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332(6160): 164-6.

Ke, N., Gao, X., Keeney, J. B., Boeke, J. D. and Voytas, D. F. (1999). The yeast retrotransposon Ty5 uses the anticodon stem-loop of the initiator methionine tRNA as a primer for reverse transcription. Rna 5(7): 929-38.

Kennell, J. C., Wang, H. and Lambowitz, A. M. (1994). The Mauriceville plasmid of *Neurospora spp.* uses novel mechanisms for initiating reverse transcription in vivo. Mol Cell Biol 14(5): 3094-107.

Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J. and Skalka, A. M. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res 19(4): 851-60.

Kikuchi, Y., Ando, Y. and Shiba, T. (1986). Unusual priming mechanism of RNA-directed DNA synthesis in copia retrovirus-like particles of *Drosophila*. Nature 323(6091): 824-6.

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(5): 464-78.

Kinsey, J. A. and Helber, J. (1989). Isolation of a transposable element from *Neurospora crassa*. Proc Natl Acad Sci USA 86(6): 1929-33.

Kirchner, J., Connolly, C. M. and Sandmeyer, S. B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retrovirus-like element. Science 267(5203): 1488-91.

Kirchner, J. and Sandmeyer, S. (1993). Proteolytic processing of Ty3 proteins is required for transposition. J Virol 67(1): 19-28.

Kirchner, J., Sandmeyer, S. B. and Forrest, D. B. (1992). Transposition of a Ty3 GAG3-POL3 fusion mutant is limited by availability of capsid protein. J Virol 66(10): 6081-92.

Kordis, D. (2005). A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. Gene 347(2): 161-73.

Kossack, D. S. and Kinlaw, C. S. (1999). IFG, a gypsy-like retrotransposon in *Pinus* (*Pinaceae*), has an extensive history in pines. Plant Mol Biol 39(3): 417-26.

Kribel, H. (1985). DNA sequence components of the *Pinus strobus* nuclear genome. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 15(1): 1-4.

Kujau, M., Weber, H. and Barth, G. (1992). Characterization of mutants of the yeast *Yarrowia lipolytica* defective in acetyl-coenzyme A synthetase. Yeast 8(3): 193-203.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227(5259): 680-5.

Lambowitz, A. M. and Zimmerly, S. (2004). Mobile group II introns. Annu Rev Genet 38: 1-35.

Lampson, B. C., Sun, J., Hsu, M. Y., Vallejo-Ramirez, J., Inouye, S. and Inouye, M. (1989). Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. Science 243(4894 Pt 1): 1033-8.

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(6822): 860-921.

Lankenau, D. H., Huijser, P., Jansen, E., Miedema, K. and Hennig, W. (1988). Micropia: a retrotransposon of *Drosophila* combining structural features of DNA viruses, retroviruses and non-viral transposable elements. J Mol Biol 204(2): 233-46.

Lauermann, V. and Boeke, J. D. (1994). The primer tRNA sequence is not inherited during Ty1 retrotransposition. Proc Natl Acad Sci USA 91(21): 9847-51.

Lauermann, V. and Boeke, J. D. (1997). Plus-strand strong-stop DNA transfer in yeast Ty retrotransposons. Embo J 16(21): 6603-12.

Lawler, J. F., Jr., Haeusser, D. P., Dull, A., Boeke, J. D. and Keeney, J. B. (2002). Ty1 defect in proteolysis at high temperature. J Virol 76(9): 4233-40.

Le Dall, M. T., Nicaud, J. M. and Gaillardin, C. (1994). Multiple-copy integration in the yeast *Yarrowia lipolytica*. Curr Genet 26(1): 38-44.

Leeton, P. R. and Smyth, D. R. (1993). An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol Gen Genet 237(1-2): 97-104.

Leng, P., Klatte, D. H., Schumann, G., Boeke, J. D. and Steck, T. L. (1998). Skipper, an LTR retrotransposon of *Dictyostelium*. Nucleic Acids Res 26(8): 2008-15.

Levin, H. L. (1995). A novel mechanism of self-primed reverse transcription defines a new family of retroelements. Mol Cell Biol 15(6): 3310-7.

Levin, H. L. (1996). An unusual mechanism of self-primed reverse transcription requires the RNase H domain of reverse transcriptase to cleave an RNA duplex. Mol Cell Biol 16(10): 5645-54.

Levin, H. L. and Boeke, J. D. (1992). Demonstration of retrotransposition of the Tf1 element in fission yeast. Embo J 11(3): 1145-53.

Levin, H. L., Weaver, D. C. and Boeke, J. D. (1990). Two related families of retrotransposons from *Schizosaccharomyces pombe*. Mol Cell Biol 10(12): 6791-8.

Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A. and Sheen, F. M. (1993). Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell 75(6): 1083-93.

Lim, D. and Maas, W. K. (1989). Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56(5): 891-904.

Lin, J. H. and Levin, H. L. (1997). A complex structure in the mRNA of Tf1 is recognized and cleaved to generate the primer of reverse transcription. Genes Dev 11(2): 270-85.

Lin, J. H. and Levin, H. L. (1997). Self-primed reverse transcription is a mechanism shared by several LTR- containing retrotransposons. Rna 3(9): 952-3.

Lin, S. S., Nymark-McMahon, M. H., Yieh, L. and Sandmeyer, S. B. (2001). Integrase mediates nuclear localization of Ty3. Mol Cell Biol 21(22): 7826-38.

Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D. *et al.* (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402(6763): 761-8.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. J Biol Chem 193(1): 265-75.

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(4): 595-605.

Lyozin, G. T., Makarova, K. S., Velikodvorskaja, V. V., Zelentsova, H. S., Khechumian, R. R., Kidwell, M. G., Koonin, E. V. and Evgen'ev, M. B. (2001). The structure and evolution of Penelope in the *virilis* species group of *Drosophila*: an ancient lineage of retroelements. J Mol Evol 52(5): 445-56.

Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16(6): 793-805.

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(6): 5186-90.

Malik, H. S. and Eickbush, T. H. (2001). Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11(7): 1187-97.

Malik, H. S., Henikoff, S. and Eickbush, T. H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res 10(9): 1307-18.

Marchler-Bauer, A. and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32(Web Server issue): W327-31.

Marin, I. and Llorens, C. (2000). Ty3/*Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol Biol Evol 17(7): 1040-9.

Marlor, R. L., Parkhurst, S. M. and Corces, V. G. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol Cell Biol 6(4): 1129-34.

Martoglio, B., Graf, R. and Dobberstein, B. (1997). Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. Embo J 16(22): 6636-45.

Matthews, G. D., Goodwin, T. J., Butler, M. I., Berryman, T. A. and Poulter, R. T. (1997). pCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast *Candida albicans*. J Bacteriol 179(22): 7118-28.

Mauersberger, S., Wang, H. J., Gaillardin, C., Barth, G. and Nicaud, J. M. (2001). Insertional mutagenesis in the n-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. J Bacteriol 183(17): 5102-9.

McClintock, B. (1984). The significance of responces of the genome to challenge. Science 226: 792-801.

McHale, M. T., Roberts, I. N., Noble, S. M., Beaumont, C., Whitehead, M. P., Seth, D. and Oliver, R. P. (1992). CfT-I: an LTR-retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato. Mol Gen Genet 233(3): 337-47.

Mes, J. J., Haring, M. A. and Cornelissen, B. J. (2000). Foxy: an active family of short interspersed nuclear elements from *Fusarium oxysporum*. Mol Gen Genet 263(2): 271-80.

Mhiri, C., Morel, J. B., Vernhettes, S., Casacuberta, J. M., Lucas, H. and Grandbastien, M. A. (1997). The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol 33(2): 257-66.

Michaille, J. J., Mathavan, S., Gaillard, J. and Garel, A. (1990). The complete sequence of mag, a new retrotransposon in *Bombyx mori*. Nucleic Acids Res 18(3): 674.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., Vogelstein, B. and Nakamura, Y. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52(3): 643-5.

Milanesi, L., Muselli, M. and Arrigo, P. (1996). Hamming-Clustering method for signals prediction in 5' and 3' regions of eukaryotic genes. Comput Appl Biosci 12(5): 399-404.

Miller, W. A., Bujarski, J. J., Dreher, T. W. and Hall, T. C. (1986). Minus-strand initiation by brome mosaic virus replicase within the 3' tRNA-like structure of native and modified RNA templates. J Mol Biol 187(4): 537-46.

Miyake, T., Mae, N., Shiba, T. and Kondo, S. (1987). Production of virus-like particles by the transposable genetic element, *copia*, of *Drosophila melanogaster*. Mol Gen Genet 207(1): 29-37.

Moore, S. P., Rinckel, L. A. and Garfinkel, D. J. (1998). A Ty1 integrase nuclear localization signal required for retrotransposition. Mol Cell Biol 18(2): 1105-14.

Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., Batzer, M. A. and Moran, J. V. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31(2): 159-65.

Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M. and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. Nature 333(6168): 87-90.

Mount, S. M. and Rubin, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol 5(7): 1630-8.

Murata, H., Miyazaki, Y. and Yamada, A. (2001). marY2N, a LINE-like non-long terminal repeat (non-LTR) retroelement from the ectomycorrhizal homobasidiomycete *Tricholoma matsutake*. Biosci Biotechnol Biochem 65(10): 2301-5.

Murata, H. and Yamada, A. (2000). marY1, a member of the gypsy group of long terminal repeat retroelements from the ectomycorrhizal basidiomycete *Tricholoma matsutake*. Appl Environ Microbiol 66(8): 3642-5.

Nakamura, T. M. and Cech, T. R. (1998). Reversing time: origin of telomerase. Cell 92(5): 587-90.

Nakayashiki, H., Kiyotomi, K., Tosa, Y. and Mayama, S. (1999). Transposition of the retrotransposon MAGGY in heterologous species of filamentous fungi. Genetics 153(2): 693-703.

Nargang, F. E. (1986). Conservation of a long open reading frame in two *Neurospora* mitochondrial plasmids. Mol Biol Evol 3(1): 19-28.

Nargang, F. E., Bell, J. B., Stohl, L. L. and Lambowitz, A. M. (1984). The DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggest a relationship to introns and mobile elements. Cell 38(2): 441-53.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H. and Matsuo, M. (1993). Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest 91(5): 1862-7.

Neuveglise, C., Chalvet, F., Wincker, P., Gaillardin, C. and Casaregola, S. (2005). Mutator-like element in the yeast *Yarrowia lipolytica* displays multiple alternative splicings. Eukaryot Cell 4(3): 615-24.

Neuveglise, C., Feldmann, H., Bon, E., Gaillardin, C. and Casaregola, S. (2002). Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12(6): 930-43.

Neuveglise, C., Sarfati, J., Latge, J. P. and Paris, S. (1996). Afut1, a retrotransposon-like element from *Aspergillus fumigatus*. Nucleic Acids Res 24(8): 1428-34.

Nicaud, J. M., Fabre, E. and Gaillardin, C. (1989). Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. Curr Genet 16(4): 253-60.

Nicholas, K. B., Nicholas, H. B. J. and Deerfield, D. W. I. (1997). GeneDoc: Analysis and visualization of genetic variation. EMBNEW News 4: 14.

Nishimura, M., Hayashi, N., Jwa, N. S., Lau, G. W., Hamer, J. E. and Hasebe, A. (2000). Insertion of the LINE retrotransposon MGL causes a conidiophore pattern mutation in *Magnaporthe grisea*. Mol Plant Microbe Interact 13(8): 892-4.

Page, R. D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12(4): 357-8.

Panstruga, R., Buschges, R., Piffanelli, P. and Schulze-Lefert, P. (1998). A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 26(4): 1056-62.

Pantazidis, A., Labrador, M. and Fontdevila, A. (1999). The retrotransposon Osvaldo from *Drosophila buzzatii* displays all structural features of a functional retrovirus. Mol Biol Evol 16(7): 909-21.

Paquin, C. E. and Williamson, V. M. (1984). Temperature Effects on the Rate of Ty Transposition. Science 226(4670): 53-55.

Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D. and Knowles, B. B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7(4): 597-606.

Pelissier, T., Tutois, S., Deragon, J. M., Tourmente, S., Genestier, S. and Picard, G. (1995). Athila, a new retroelement from *Arabidopsis thaliana*. Plant Mol Biol 29(3): 441-52.

Pignede, G., Wang, H. J., Fudalej, F., Seman, M., Gaillardin, C. and Nicaud, J. M. (2000). Autocloning and amplification of *LIP2* in *Yarrowia lipolytica*. Appl Environ Microbiol 66(8): 3283-9.

Plant, E. P., Goodwin, T. J. and Poulter, R. T. (2000). Tca5, a Ty5-like retrotransposon from *Candida albicans*. Yeast 16(16): 1509-18.

Pogulis, R. J., Vallejo, A. N. and Pease, L. R. (1996). In vitro recombination and mutagenesis by overlap extension PCR. Methods Mol Biol 57: 167-76.

Poulter, R. and Butler, M. (1998). A retrotransposon family from the pufferfish (fugu) *Fugu*

rubripes. Gene 215(2): 241-9.

Ratner, V. A., Zabanov, S. A., Kolesnikova, O. V. and Vasilyeva, L. A. (1992). Induction of the mobile genetic element Dm-412 transpositions in the *Drosophila* genome by heat shock treatment. Proc Natl Acad Sci USA 89(12): 5650-4.

Rawson, J. M. (2000). Transposable elements in the phytopathogenic fungus *Stagonospora nodorum*. School of Biosciences. Birmingham, The University of Birmingham. PhD thesis.

Rice, S. A., Bieber, J., Chun, J. Y., Stacey, G. and Lampson, B. C. (1993). Diversity of retron elements in a population of rhizobia and other gram-negative bacteria. J Bacteriol 175(13): 4250-4.

Rogozin, I. B., Kochetov, A. V., Kondrashov, F. A., Koonin, E. V. and Milanesi, L. (2001). Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. Bioinformatics 17(10): 890-900.

Rolfe, M., Spanos, A. and Banks, G. (1986). Induction of yeast Ty element transcription by ultraviolet light. Nature 319: 339-340.

Rosenberg, N. and Jolicoeur, P. (1997). Retroviral pathogenesis. In: Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus (eds.) New York, Cold Spring Harbor Laboratory Press**:** 475-587.

Roth, J. F. (2000). The yeast Ty virus-like particles. Yeast 16(9): 785-95.

Ruiz-Perez, V. L., Murillo, F. J. and Torres-Martinez, S. (1996). Prt1, an unusual retrotransposon-like sequence in the fungus *Phycomyces blakesleeanus*. Mol Gen Genet 253(3): 324-33.

Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985). Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc Natl Acad Sci USA 82(3): 677-81.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984). Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature 312(5995): 659-61.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74(12): 5463-7.

Sanmiguel, P. and Bennetzen, J. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. ANNALS OF BOTANY 82: 37-44.

SanMiguel, P. J., Ramakrishna, W., Bennetzen, J. L., Busso, C. S. and Dubcovsky, J. (2002). Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct Integr Genomics 2(1-2): 70-80.

Schmid-Berger, N., Schmid, B. and Barth, G. (1994). Ylt1, a highly repetitive retrotransposon in

the genome of the dimorphic fungus *Yarrowia lipolytica*. J Bacteriol 176(9): 2477-82.

Schmidt, T. (1999). LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol Biol 40(6): 903-10.

Schramke, V. and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science 301(5636): 1069-74.

Schwarz, D., Tizard, R. and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell 32: 853-869.

Senam (2004). Molekulare Charakterisierung des Retrotransposons Ylt1 der Hefe *Yarrowia lipolytica*. Dissertation, Institut für Mikrobiologie, Fakultät Mathematik und Naturwissenschaften, Technische Universität Dresden.

Sentry, J. W. and Smyth, D. R. (1989). An element with long terminal repeats and its variant arrangements in the genome of *Lilium henryi*. Mol Gen Genet 215(2): 349-54.

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(10): 6540-51.

Shiflett, A. M., Enkerli, J. and Covert, S. F. (2002). Nht2, a *copia* LTR retrotransposon from a conditionally dispensable chromosome in *Nectria haematococca*. Curr Genet 41(2): 99-106.

Singleton, T. L. and Levin, H. L. (2002). A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukaryot Cell 1(1): 44-55.

Smith, R. F., Wiese, B. A., Wojzynski, M. K., Davison, D. B. and Worley, K. C. (1996). BCM Search Launcher--an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. Genome Res 6(5): 454-62.

Smyth, D. R., Kalitsis, P., Joseph, J. L. and Sentry, J. W. (1989). Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of Drosophila. Proc Natl Acad Sci USA 86(13): 5015-9.

Springer, M. S., Davidson, E. H. and Britten, R. J. (1991). Retroviral-like element in a marine invertebrate. Proc Natl Acad Sci USA 88(19): 8401-4.

Strand, D. J. and McDonald, J. F. (1985). *Copia* is transcriptionally responsive to environmental stress. Nucleic Acids Res 13(12): 4401-10.

Stucka, R., Lochmuller, H. and Feldmann, H. (1989). Ty4, a novel low-copy number element in *Saccharomyces cerevisiae*: one copy is located in a cluster of Ty elements and tRNA genes. Nucleic Acids Res 17(13): 4993-5001.

Sun, J., Herzer, P. J., Weinstein, M. P., Lampson, B. C., Inouye, M. and Inouye, S. (1989). Extensive diversity of branched-RNA-linked multicopy single-stranded DNAs in clinical strains of *Escherichia coli*. Proc Natl Acad Sci USA 86(18): 7208-12.

Tanda, S., Mullor, J. L. and Corces, V. G. (1994). The *Drosophila* tom retrotransposon encodes an envelope protein. Mol Cell Biol 14(8): 5392-401.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The

CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24): 4876-82.

Thomson, K. G., Thomas, J. E. and Dietzgen, R. G. (1998). Retrotransposon-like sequences integrated into the genome of pineapple, *Ananas comosus*. Plant Mol Biol 38(3): 461-5.

Van Arsdell, S. W., Stetler, G. L. and Thorner, J. (1987). The yeast repeated element sigma contains a hormone-inducible promoter. Mol Cell Biol 7(2): 749-59.

Vogt, V. M. (1997). Retroviral virions and genomes. In: Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus (eds.) New York, Cold Spring Harbor Laboratory Press**:** 27-69.

Voytas, D. F. and Boeke, J. D. (1992). Yeast retrotransposon revealed. Nature 358(6389): 717.

Voytas, D. F. and Boeke, J. D. (1993). Yeast retrotransposons and tRNAs. Trends Genet 9(12): 421-7.

Wang, H. and Lambowitz, A. M. (1993). The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. Cell 75(6): 1071-81.

Warmington, J. R., Waring, R. B., Newlon, C. S., Indge, K. J. and Oliver, S. G. (1985). Nucleotide sequence characterization of Ty 1-17, a class II transposon from yeast. Nucleic Acids Res 13(18): 6679-93.

Weaver, D. C., Shpakovski, G. V., Caputo, E., Levin, H. L. and Boeke, J. D. (1993). Sequence analysis of closely related retrotransposon families from fission yeast. Gene 131(1): 135-9.

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E. and Keller, B. (2001). Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26(3): 307-16.

Wickerham, L. J., Kurtzman, C. P. and Herman, A. I. (1969). Sexuality in *Candida lipolytica*. Recent Trends in Yeast Research. New York, Miner Institute, Chazy**:** 81-92.

Wilhelm, M. and Wilhelm, F. X. (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell Mol Life Sci 58(9): 1246-62.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J. and et al. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 368(6466): 32-8.

Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M. and Kingsman, A. J. (1988). HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell 55(6): 1159-69.

Wright, D. A. and Voytas, D. F. (1998). Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/*gypsy* retrotransposons that encode envelope-like proteins. Genetics 149(2): 703-15.

Xiong, Y. and Eickbush, T. H. (1988). The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol Cell Biol 8(1): 114-23.

Xiong, Y. and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. Embo J 9(10): 3353-62.

Xu, H. and Boeke, J. D. (1990). Host genes that influence transposition in yeast: the abundance of a rare tRNA regulates Ty1 transposition frequency. Proc Natl Acad Sci USA 87(21): 8360-4.

Yee, T., Furuichi, T., Inouye, S. and Inouye, M. (1984). Multicopy single-stranded DNA isolated from a gram-negative bacterium, *Myxococcus xanthus*. Cell 38(1): 203-9.

Yieh, L., Kassavetis, G., Geiduschek, E. P. and Sandmeyer, S. B. (2000). The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. J Biol Chem 275(38): 29800-7.

Yoshioka, K., Honma, H., Zushi, M., Kondo, S., Togashi, S., Miyake, T. and Shiba, T. (1990). Virus-like particle formation of *Drosophila copia* through autocatalytic processing. Embo J 9(2): 535-41.

Yuki, S., Inouye, S., Ishimaru, S. and Saigo, K. (1986). Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. Eur J Biochem 158(2): 403-10.

Zhou, Q. and Haymer, D. S. (1997). Molecular structure of yoyo, a *gypsy*-like retrotransposon from the mediterranean fruit fly, *Ceratitis capitata*. Genetica 101(3): 167-78.

Zhu, P. and Oudemans, P. V. (2000). A long terminal repeat retrotransposon Cgret from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. Curr Genet 38(5): 241-7.

Ziarczyk, P. and Best-Belpomme, M. (1991). A short 5' region of the long terminal repeat is required for regulation by hormone and heat shock of *Drosophila* retrotransposon 1731. Nucleic Acids Res 19(20): 5689-93.

Zou, S., Wright, D. A. and Voytas, D. F. (1995). The *Saccharomyces* Ty5 retrotransposon family is associated with origins of DNA replication at the telomeres and the silent mating locus HMR. Proc Natl Acad Sci USA 92(3): 920-4.

7 Appendix

7.1 Plasmid maps

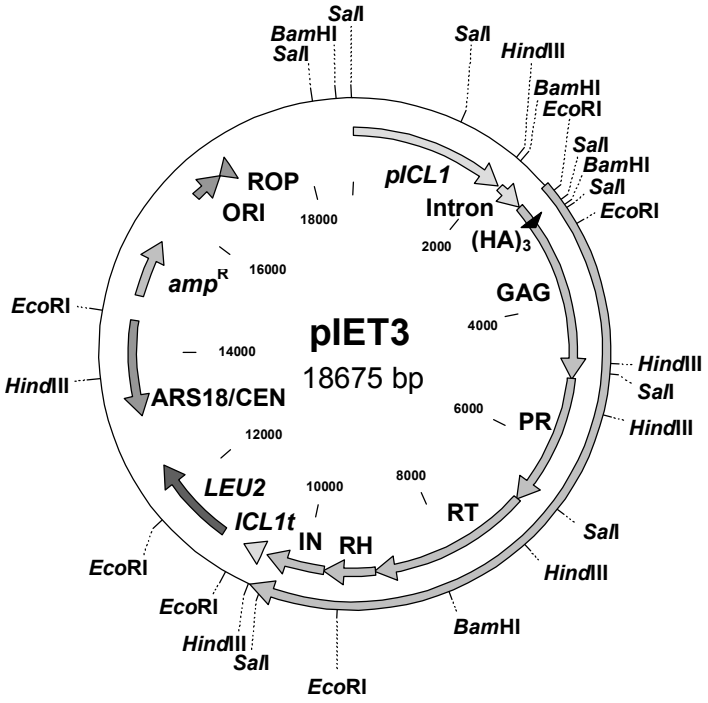

Figure 7.1 The plasmid pIET3 used for the expression of the HA-tagged Gag protein of Ylt1 under control of the *ICL1* promoter.

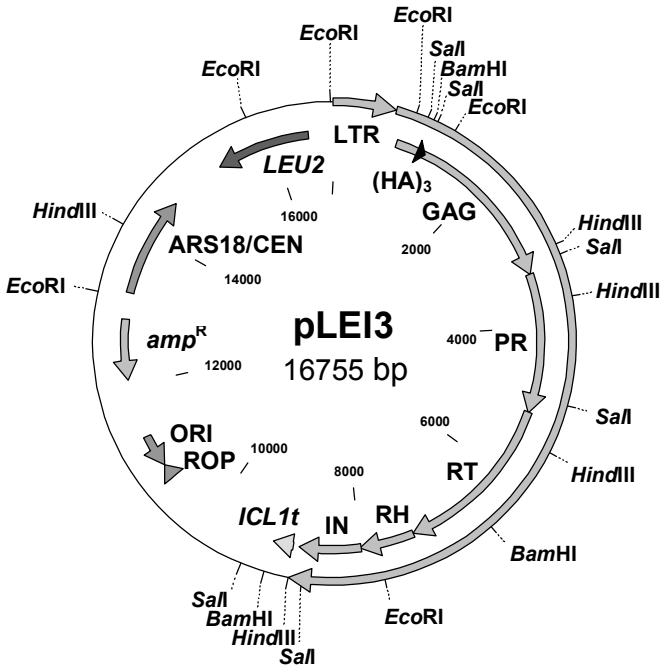

Figure 7.2 The plasmid pLEI3 used for the expression of the HA-tagged Gag protein of Ylt1 under control of the LTR promoter.

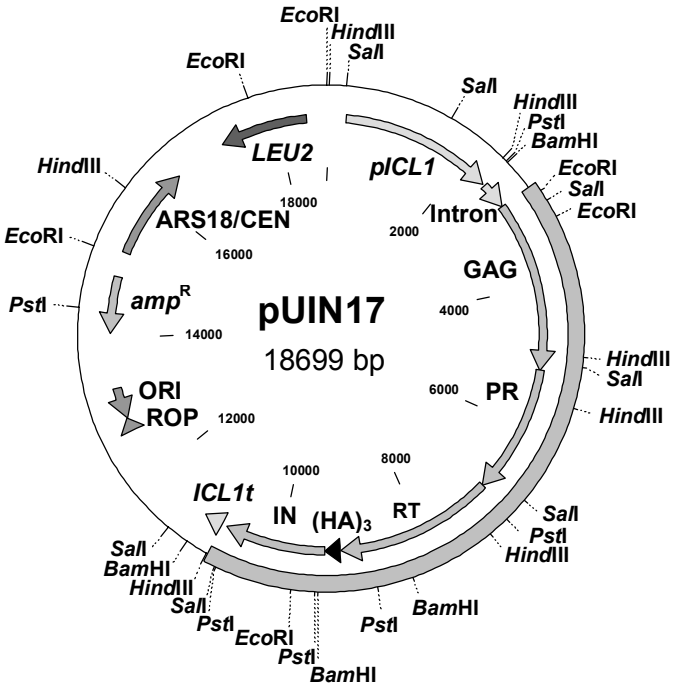

Figure 7.3 The plasmid pUIN17 used for the expression of the HA-tagged integrase of Ylt1 under control of the *ICL1* promoter.

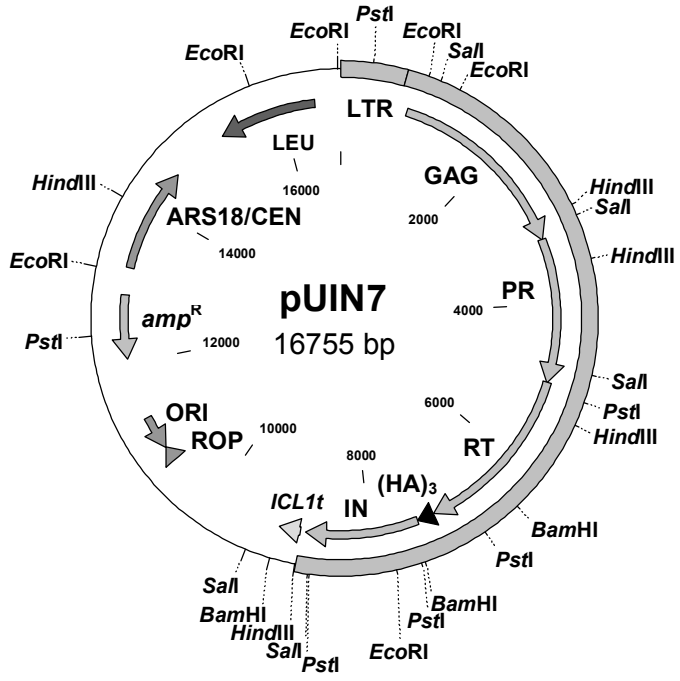

Figure 7.4 The plasmid pUIN7 used for the expression of the HA-tagged integrase of Ylt1 under control of the LTR promoter.

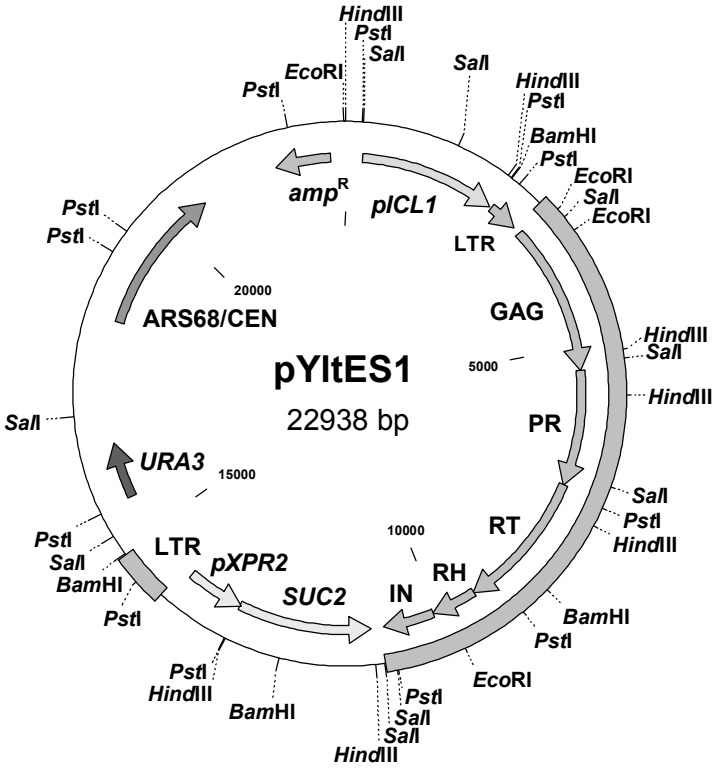

Figure 7.6 The plasmid pYltS3 used for the exprssion of the *SUC2*-marked Ylt1 element under control of the LTR promoter.

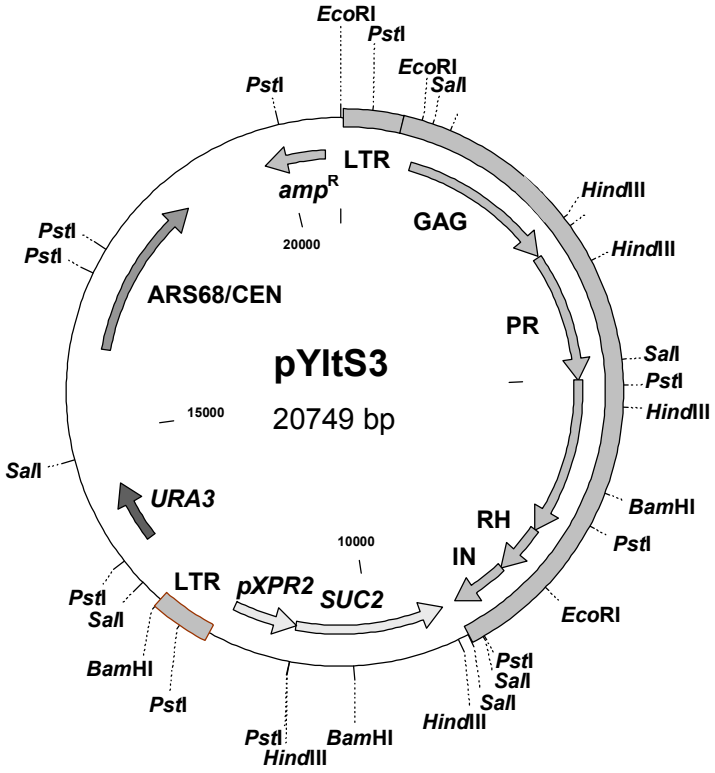

Figure 7.6 The plasmid pYltS3 used for the exprssion of the *SUC2*-marked Ylt1 element under control of the LTR promoter.

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

The thesis work was conducted from 01.09.2001 to 30.04.2005 under the supervision of Prof. Dr. Gerold Barth at the Institut für Mikrobiologie of the Technische Universität Dresden.

Dresden, 12.08.2005 Andriy Kovalchuk

I would like to thank Prof. Dr. Gerold Barth for the provided possibility to work on the interesting theme, for his permanent assistance and helpful discussions.

I also would like to thank all my colleagues from the Institute of Microbiology for their friendly support during my work.