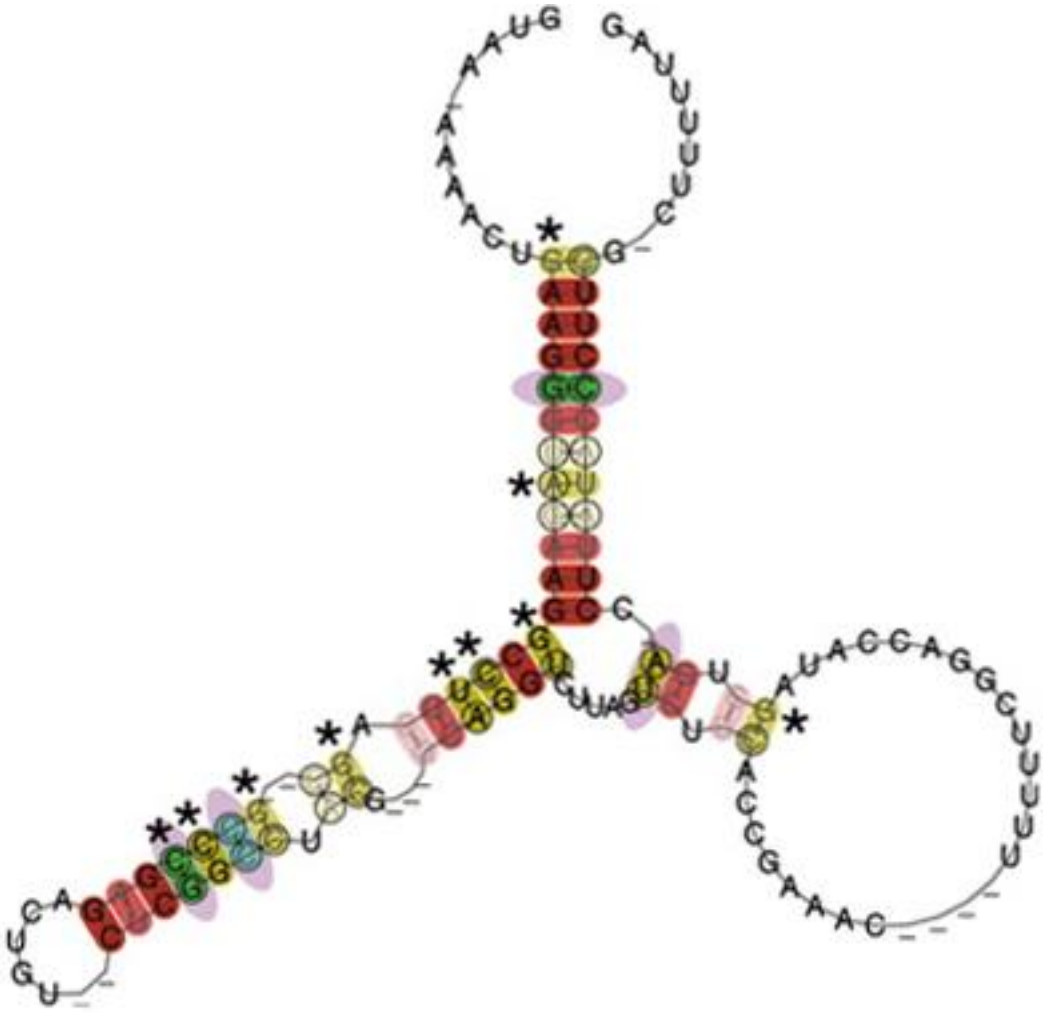
Multiplication of Introner-Like Elements in Fungi



Minor MSc Thesis Report

Ivo Koolhaas

Multiplication of Introner-Like Elements in Fungi

Ivo Koolhaas
900825-463-070

Minor MSc Thesis Phytopathology (PHP-80424)

Performed at:
Laboratory of Phytopathology
Wageningen University and Research Centre

Februari-Juni 2013

Supervisors:
Dr. Jérôme Collemare
Dr. Pierre de Wit

Examiners:
Dr. Pierre de Wit

Table of Contents

List of figures

List of tables

Abstract

Spliceosomal introns interrupt the coding sequence of eukaryotic genes and are removed from the primary transcript by the the spliceosome. Recently, a new type of spliceosomal intron was found in the genomes of six fungal species. These newfound introns were called introner-like elements (ILEs).They were found in hundreds of near-identical copies in unrelated genes, suggesting that they are able to proliferate within genomes. Because the majority of intron gains in fungi could not be associated with other types of intron duplication, it is suggested that ILE multiplication is the main mechanism behind intron gain in fungi. It is hypothesized that the mechanism behind ILE multiplication is similar to the retro-homing of self-splicing group II introns. In this research, experiments were carried out to assess the mobility of ILEs and the mechanism behind ILE multiplication. A genetic screen was performed to assess the mobility of ILEs from *Cladosporium fulvum* and *Zymoseptoria tritici* in the *Saccharomyces cerevisiae* temperature sensitive *prp2-1* mutant. However, this assay did not work properly because the ILEs were not spliced from the primary transcript. This is probably caused by a splicing incompatibility, in which the *S. cerevisiae* spliceosome is not able to recognize ILEs from *C. fulvum* and *Z. tritici*. To assess the self-splicing ability and ribozyme activity of ILEs, several *in vitro* experiments were set up. In the self-splicing assays, *sfGFP* RNA molecules containing introns from *C. fulvum* and *Z. tritici* were incubated under various conditions. Self-splicing events could not be detected for the combinations of ILEs and experimental conditions that were tested. In the ribozyme activity assays, ILE RNA molecules from *C. fulvum* and *Z. tritici* were mixed with target DNAs and incubated under various conditions. No ribozyme activity was observed for the combinations of ILEs, target DNAs and experimental conditions that were tested. Gel electrophoresis was used to detect both self-splicing events and ribozyme activity. It is possible that these processes were taking place at a very low rate, making them undetectable by gel electrophoresis.

1. Introduction

1.1 Intron classification

1.1.1 The mosaic structure of genes

Introns are non-coding sequences that interrupt the coding sequence of genes. They are removed from transcribed pre-mRNA by a process called splicing. Based on the way they are spliced, introns can be divided into different groups. Group I and II self- splicing introns are spliced by RNA catalysis (Lilley, 2010). Archaeal- and spliceosomal introns are respectively spliced by tRNA splicing endonucleases (Lykke-Andersen et al., 1997) and the spliceosome (Rogozin et al., 2012). Here, group II self-splicing introns and spliceosomal introns will be adressed because they share a similar splicing mechanism. Because of this similarity, it has been suggested that group II self-splicing introns are the predecessors of spliceosomal introns (Cavalier-Smith, 1991; Roy and Gilbert, 2006).

1.1.2 Self-splicing introns

Group II introns are mainly found in the mitochondria and chloroplasts of plants and fungi (Dai and Zimmerly, 2003). More recently, they have also been found in several eubacterial- and archaebacterial genomes (Robart and Zimmerly, 2005). While group II introns are mostly present within protein-coding genes, they have also been found in rRNA and tRNA genes. Self-splicing of a group II intron occurs through two consecutive transesterification reactions, using an adenine residue within the intron as the initiating nucleophile. Subsequently, the intron is excised in the form of a lariat. Group II introns act as mobile elements that can insert into intronless alleles in the genome through a process called retrohoming. This process requires site-specific DNA endonuclease and reverse transcriptase enzymes, which are encoded by the group II intron itself (Guo et al., 1997). First, the DNA endonuclease cuts a specific site in the recipient DNA. The excised group II intron is then reverse spliced into the exon junction, after which the complementary strand is synthesized by reverse transcription. This reverse transcription is primed by the antisense strand of the cleaved target DNA (Zimmerly et al., 1995).

1.1.3 Spliceosomal introns in eukaryotes

In contrast to self-splicing introns, spliceosomal introns only occur in the nuclear genome of eukaryotes and require help from the spliceosome to be excised from transcribed pre-mRNA (Rogozin et al., 2012). The spliceosome is a complex ribonucleoprotein that recognizes introns, removes them from the primary transcript and ligates the adjacent exons (Will and Lührmann, 2011). The site where the 5’ end of the intron is cut is called the donor splice site. Almost always, the sequence of the donor splice site ends with a highly conserved GU. The site where the 3’ end of the intron is cut is called the acceptor splice site, and almost always ends with a highly conserved AG (Burset *et al.*, 2000). Other *cis* elements facilitating the splicing process include a branch point sequence and polypyrimidine tracts (Kupfer *et al*., 2004). After recognition of a spliceosomal intron, the different subunits of the spliceosome sequentially assemble along the splice sites and two successive transesterification reactions take place. Similar to the splicing of group II introns, an adenine residue within the intron is used as the initiating nucleophile for those reactions. The intron is then released from the pre-mRNA in the form of a lariat (Will and Lührmann, 2011). Although spliceosomal introns were initially named “junk” DNA, various functions have been discovered ever since. These functions include expanding protein diversity and regulating gene expression (Le Hir *et al*., 2003).

1.2 The origin of spliceosomal introns

1.2.1 Intron gain and loss analysis

Intron size and density are highly variable between eukaryotic genomes. Eukaryotic genomes are commonly classified as either intron-poor or intron rich. The intron-poor group consists mostly of unicellular organisms, while the intron-rich one includes animals, plants and some fungi (Rogozin et al., 2012). Although spliceosomal introns have been studied for more than 30 years, their origin remains elusive. When researchers began to compare the intron-exon structures of orthologous genes of plants and animals, a strong conservation of intron position became apparent (Fedorov et al., 2002). This finding implies that higher eukaryotes did not accumulate introns in their genomes during evolution, but that their ancestors were also intron rich. Therefore, the rates of intron gain and loss must have been in approximate balance in the evolution of higher eukaryotes. When the evolution of gene architecture was reconstructed for several major groups of eukaryotes, it was suggested that intron-poor species originated from intron-rich ancestors as well. This finding implies that intron loss is the driving force behind the evolution of many eukaryotic lineages, while episodic intron gain only took place in a few major branches (Csuros et al., 2011). An intron gain and loss analysis indicated that the intron content of fungal genomes is rather balanced, suggesting that intron gains are still occurring to a large extent in this kingdom (Koonin, 2009).

1.2.2 Mechanisms of intron gain and loss

Two mechanisms of intron loss have already been identified, namely reverse transcriptase mediated intron loss (RTMIL) and genomic deletion. In the RTMIL model, a given gene is transcribed and an intron is spliced out. Reverse transcriptase then converts this spliced transcript into cDNA. Finally, recombination between this cDNA and the original genomic intron leads to loss of the intron. In the other model, a (nearly) exact genomic deletion is responsible for the loss of an intron (Roy and Gilbert, 2006). Several mechanisms have been proposed to explain intron gains, including intron transposition, transposon insertion, tandem genomic duplication, intron gain during double-strand break repair, insertion of a group II intron, intron transfer and intronization (Yenerall and Zhou, 2012).

In the mechanism of *intron transposition*, intron gain occurs through reverse splicing of an intron into its own- or another mRNA molecule. Subsequently, the mRNA needs to be reverse transcribed, and inserted into the genomic DNA through recombination with its original locus (Yenerall and Zhou, 2012) (Figure 1a). The process of reverse splicing has been proven in vitro using a mutant that is unable to release mRNA from the spliceosome (Tseng and Cheng, 2008). However, there is a problem with the intron transposition mechanism called the “rate paradox” (Roy and Irimia, 2009). The mechanism behind intron transposition is identical to the mechanism behind RTMIL, the only difference being that RTMIL does not need reverse splicing. Because reverse splicing is an extremely rare event, it is expected that there are much more introns lost through RTMIL than introns gained through intron transposition. However, there are several fungal lineages in which intron gains and losses are more or less balanced (Roy and Gilbert, 2006). In such lineages, introns are gained at a rate that cannot be explained by the intron transposition mechanism.

Transposon insertion and *tandem genomic duplication* are both mechanisms of intron gain that rely on the duplication of an AGGT sequence. In the transposon insertion mechanism, a transposable element is inserted in this AGGT sequence. The AGGT sequence is then duplicated at both sides of the transposon, generating both a donor- and an acceptor site as a result (Figure 1b). In tandem genomic duplication, these splice sites are generated by the duplication of an exonic fragment containing AGGT sequence

(Figure 1c). Of all seven proposed mechanisms of intron gain, tandem genomic duplication is the only one that has been proven *in vivo* (Hellsten et al, 2011).

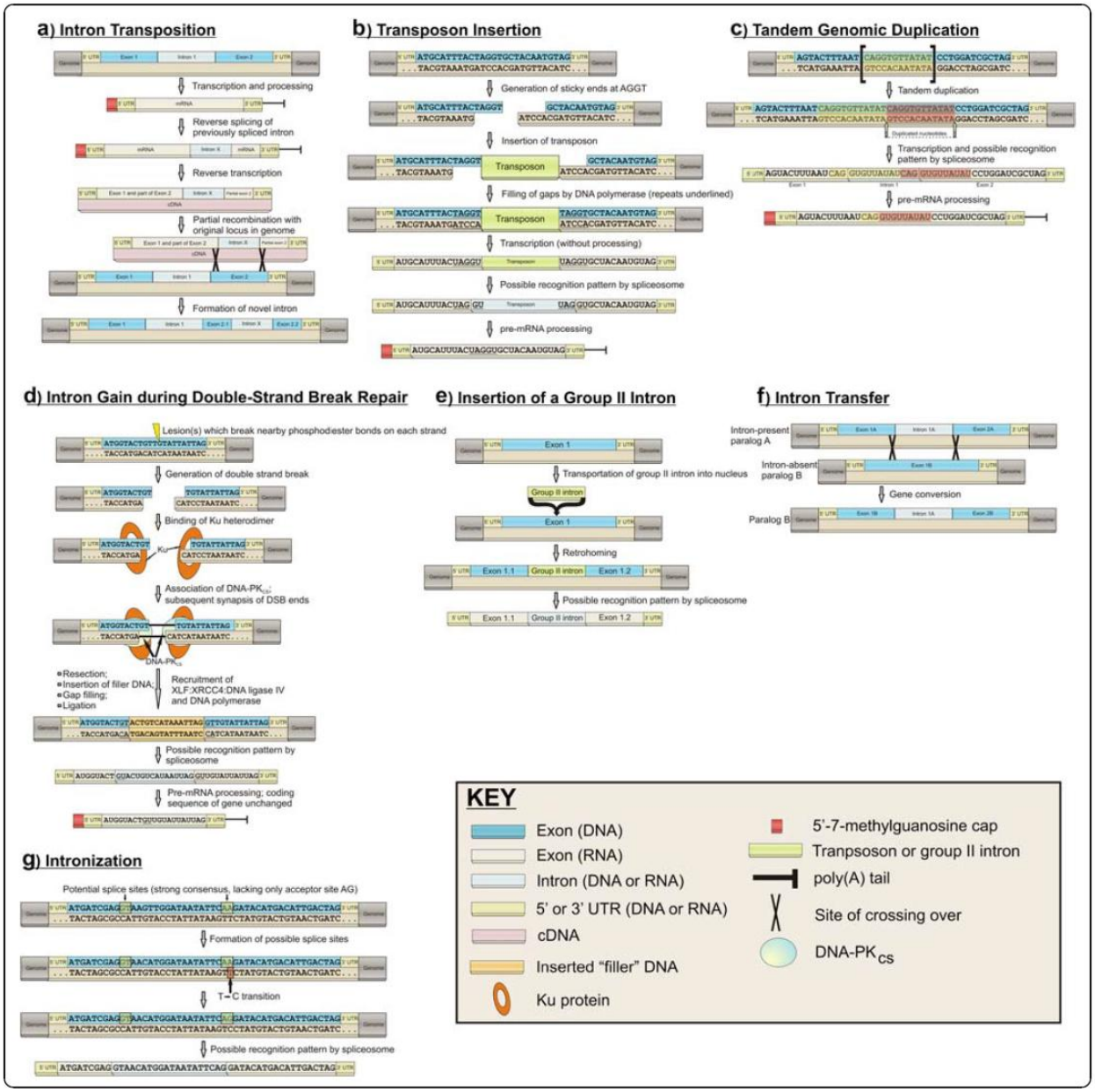


Figure 1. The seven proposed mechanisms of intron gain. Strong consensus donor and acceptor sites (following a single point mutation) are highlighted in yellow (Yenerall and Zhou, 2013).

Double-strand break repair by non-homologous end joining (NHEJ) is associated with the insertion of short stretches of exogenous DNA (Yu and Gabriel, 1999). When it was found that short repeats were flanking 43% of the introns gained in *Daphnia pulex*, it was suggested that NHEJ could be one of the mechanisms behind intron gain (Li et al., 2009). In this mechanism, the short stretch of inserted exogenous DNA would function as a splicable intron (Figure 1d).

Because of the remarkable similarities between group II- and spliceosomal introns, it has been suggested that they are evolutionary related (Cavalier-Smith, 1991; Roy and Gilbert, 2006). Intron gain might take place through the *insertion of a group II intron* into genomic DNA after self-splicing. Mobile group II introns encode for reverse transcriptase and can insert specifically into intronless alleles (Figure 1e). This process is called retro-homing and has already been proven *in vitro* for the yeast mtDNA group II introns al1 and al2 (Zimmerly et al., 1995). However, a more recent *in vivo* experiment showed a strong decrease in expression of the nuclear gene *CUP1* after the insertion of a group II intron (Chalamcharla et al., 2010). Therefore, it is improbable that group II intron insertion into a nuclear gene is one of the mechanisms behind recent spliceosomal intron gain (Yenerall and Zhou, 2012).

In *intron transfer*, an intron is first gained by a paralog. Subsequently, the intron is transferred to an intron-less location in its sister paralog through recombination, yielding an extra intron (Figure 1f). A problem with this model is that it does not give any explanation for the gain of the initial intron in the paralog (Yenerall and Zhou, 2012). In the model of *intronization*, an exonic sequence is converted into an intron by random point mutations (Irimia et al., 2008) (Figure 1g). Although this is theoretically possible, the chance of two splice sites, a branch point sequence and polypyrimidine tracts all being generated by random point mutations is very low. Therefore it is improbable that this mechanism is a major contributor to intron gain.

The majority of intron gains in fungi cannot be associated with any of these proposed mechanisms (Collemare et al., 2013). Most of them are based on rare events like reverse splicing. Moreover, tandem genomic duplication is the only mechanism that has been experimentally proven. Although they may contribute to intron gain to some extent, they are unable to explain the large intron gains in higher eukaryotes that were observed during evolution.

1.3 Mobile spliceosomal introns

1.3.1 Near-identical introns in *Micromonas*

About 10.000 near-identical introns were found in isolate CCMP1545 of the green algae *Micromonas pusilla*. Because spliceosomal introns are not conserved at the sequence level (Roy and Irimia, 2009), this discovery suggests that these introns were recently gained. These near-identical introns were named introner elements (IEs) and make up about 9% of the CCMP1545 genome. IEs were found within introns of unrelated genes and almost extended to their donor- and acceptor splice sites. It is probable that transcription was involved in their proliferation because they were always found on the coding strand of genes. The IEs did not possess any known characteristics of transposable elements. Also, IEs were not found in *M. pusilla* isolate RCC299 (Worden *et al*., 2009).

1.3.2 Introner-like elements in fungi

Recently, hundreds of near-identical introns were also found in unrelated genes of the Dothideomycete fungus *Zymoseptoria tritici* (Torriani et al., 2011). After analyzing the genomes of several other fungi with a BlastN-based method, similar near identical-introns were detected in the genomes of five more fungal species (*Cladosporium fulvum*, *Dothistroma septosporum*, *Mycosphaerella fijiensis*, *Hysterium pulicare* and *Stagnospora nodorum*)(van der Burgt et al., 2012). These newfound near-identical introns were called introner-like elements (ILEs), named after the IEs found in *M. pusilla*. ILEs present in each of the fungal genomes were categorized in one to eight different clusters based on sequence similarity. Subsequent analysis showed that all elements within a given cluster were derived from a common ancestor. Moreover, a comparison between the ILE clusters in the different fungal species with a similarity matrix suggested that they were derived from a single ancestral element (van der Burgt et al.,

2012). It was found that ILEs possess the typical splicing features that are also found in regular spliceosomal introns (RSIs). Indeed, EST and RNAseq data confirm that ILEs are spliced. They can be distinguished from RSIs because they are longer and their predicted secondary structures are more stable due to the presence of three stem loops. By aligning related ILEs, it was shown that the hairpin structures of ILEs are highly conserved by compensatory mutations. However, a positive correlation was found between sequence length, pairwise identity and stability, suggesting that ILEs are subject to degeneration. This degeneration will eventually cause ILEs to lose their stable secondary structures and ability to multiply, making them indistinguishable from RSIs. Dating of multiplication events showed that ILEs cannot be distinguished from RSIs after 100.000 years of degeneration (van der Burgt et al., 2012). Because ILEs were found in hundreds of near-identical copies in unrelated genes, it is highly probable that they are able proliferate within genomes. Indeed, it was shown that ILEs account for up to 90% of recently gained introns in these six fungal species. Altogether, these findings suggest that ILEs could be the predecessors of RSIs in at least six fungal species. Because the majority of intron gains in fungi have not been associated with other types of intron duplication, ILE multiplication is hypothesized to be the main mechanism behind intron gain in fungi (Collemare *et al*., 2013).

1.3.3 Proposed mechanism of ILE multiplication

The current models of intron gain are unable to explain the high ILE content of at least six fungal species. Therefore, a new model was suggested in which the specific characteristics of ILEs are taken into account. In this hypothetical mechanism, ILE multiplication occurs through a mechanism similar to retro-homing of self-splicing group II introns. A gene containing an ILE would first be transcribed (Figure 2.1), and the ILE would be excised from the primary transcript either by self-splicing or by the spliceosome (Figure 2.2). The predicted stable secondary structure of ILEs and their strong conservation suggest that an RNA intermediate is involved in the multiplication process (Figure 2.3). It is thought that the excised ILE then recognizes and cuts a specific region in the recipient DNA (Figure 2.4). The presence of conserved nucleotides in loops could indicate such ribozyme activity. Different ILEs were located at the same loci in different isolates of *Z. tritici* and sister species indicating parallel gains (van der Burgt, unpublished data), suggesting that ILEs tend to insert into specific hotspots. Subsequent synthesis of the complementary strand could occur through a DNA-primed reverse transcription mechanism similar to group II intron insertion. Just like IEs, ILEs are always present on the coding strand of genes (Worden et al., 2009) indicating that reverse transcription might be involved in this process. However, ILEs do not encode for any proteins so an exogenous reverse transcriptase would be needed in this case. An alternative to this DNA primed reverse transcription could be RNA-mediated DNA repair (Figure 2.5).

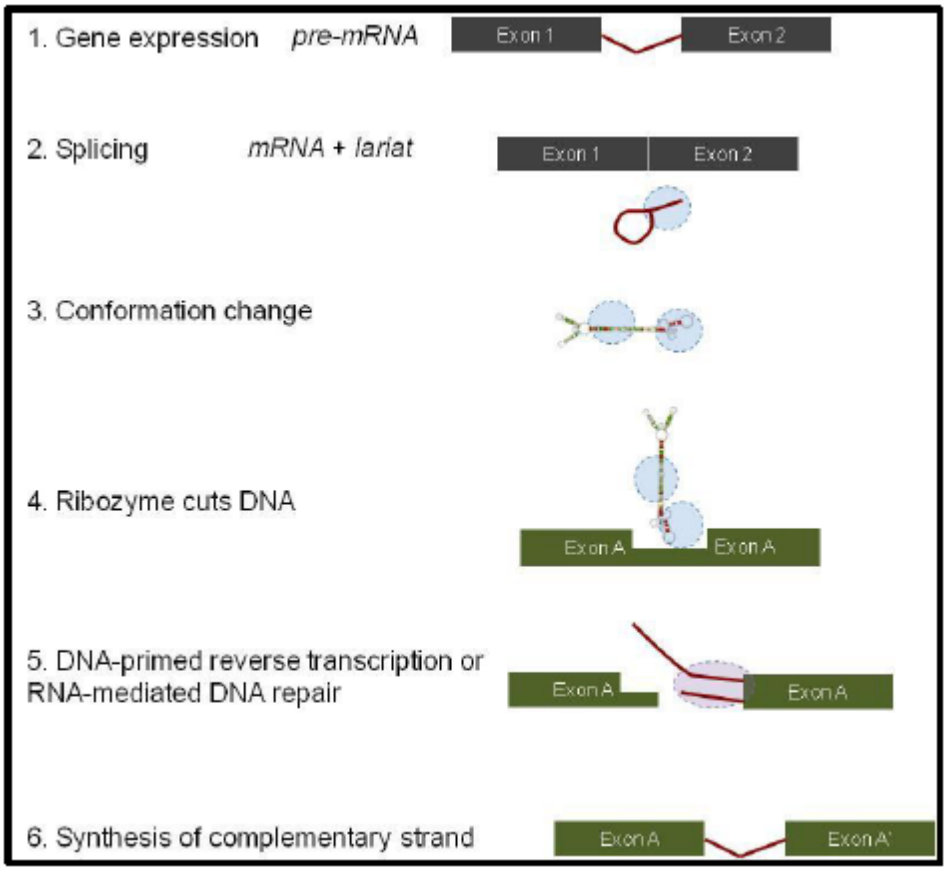


Figure 2. Proposed mechanism of ILE multiplication. A gene containing an ILE is transcribed (2.1), after which the ILE is spliced from the primary transcript (2.2). The ILE then undergoes a conformational change (2.3) cuts the recipient genomic DNA (2.4) and is inserted into the exon junction. Subsequently, DNA-primed reverse transcription or RNA-mediated DNA repair takes place (2.5) after which the complementary strand is synthesized (2.6).

2. Research objectives

Spliceosomal introns interrupt the coding sequence of eukaryotic genes and are removed from transcribed pre-mRNA by a complex ribonucleoprotein called the spliceosome. Their functions include expanding protein diversity and regulating gene expression (Le Hir *et al*., 2003). Although spliceosomal introns have been studied for more than 30 years, their origin remains elusive. Recently, a new type of spliceosomal intron was discovered in the genomes of six different fungal species, including *Cladosporium fulvum* and *Zymoseptoria tritici* (van der Burgt et al., 2012). These so called introner-like elements (ILEs) were found in hundreds of near-identical copies in unrelated genes. Similar abundant near-identical introns were previously discovered in the green alga *Micromonas pusilla,* in which they were named introner elements (IE) (Worden *et al*., 2009). ILEs possess the typical splicing features that are found in regular spliceosomal introns (RSIs), but can be distinguished because they are significantly longer and their predicted secondary structures are more stable. However, ILEs are subject to strong degeneration in length and sequence, eventually making them indistinguishable from RSIs. Because ILEs were found in hundreds of near-identical copies in unrelated genes, it is thought that they are able proliferate within genomes. Indeed, it was shown that ILEs account for up to 90% of recently gained introns in six fungal species (van der Burgt., 2012). Several mechanism to explain intron gains in eukaryotes have already been proposed, but most of them are based on rare molecular processes that occur at low frequency and lack experimental evidence (Yenerall and Zhou, 2012). Therefore, the observed intron gains in fungi cannot be associated with any of these proposed mechanisms (Collemare et al., 2013). It is hypothesized that ILE multiplication is the driving force behind intron gain in at least six fungal species. Although the mechanism of ILE multiplication is still unknown, it is hypothesized that it may occur through a mechanism similar to retro-homing of self-splicing group II introns (Roy and Irimia, 2009).

To unravel the mechanism behind ILE multiplication, an active element has to be identified. The identification of an active ILE could be hard, because fast degeneration turns them into RSIs. However, ILEs are still expected to be active in *C. fulvum* and *Z. tritici*. The highest number of ILEs was found in C. fulvum, many of which have a high identity to each other. In *Z. tritici*, a single gains analysis provides evidence for recent ILE multiplication events (van der Burgt et al., 2012).

To address the mobility of ILEs, a genetic screen was designed specifically aimed at the identification of new insertion events in *Saccharomyces cerevisiae* (Ochoa Tufiño, 2013). The self-splicing- and ribozyme activity of ILEs from both *Z. tritici* and *C. fulvum* will be assessed *in vitro*. To test self-splicing ability, different ILEs will be inserted into *sfGFP* and incubated under various conditions. Self-splicing events can then be detected on gel. To test ribozyme activity, different ILEs will be mixed with target DNA and incubated under various conditions. Whether or not this target DNA is cut can also be detected on gel.

3. Materials & Methods

3.1 Bait plasmid construction

Two PCR reactions were performed to amplify the *URA3* promoter and coding sequence from the pEYA2 vector (provided by Dr. Colin Lazarus). The primers VO32 and VO33 (Table 1) were used to amplify the *URA3* promoter. To amplify the *URA3* coding sequence, the primers VO2, VO3, VO4, VO6, VO13, VO15, VO16 and VO17 were used in combination with JC164 (Table 1). In both PCRs, 50 ng of pEYA2 DNA was combined with Phusion® Flash High-Fidelity PCR Master Mix (Thermo Scientific). The used PCR conditions were 1 min 98°C, 30 cycles of 1 sec 98°C, 5 sec 60°C, 15 sec/kb 72°C and a final step of 1 min at 72°C. Bait sequences with tails homologous to the *URA3* promoter (Table 2) were previously synthesized. Each bait sequence, the *URA3* promoter and coding sequence were used in a ratio 1:1:1 (10:50:100 ng) to perform a double-joint PCR. The PCR was performed using Pfu DNA Polymerase (Promega) to assemble the fragments. The used PCR conditions were 2 min 95°C, 15 cycles of 1 min 95°C, 10 min 60°C, 2:20 min 72°C and a final step of 5 min at 72°C. The assembled fragment was then amplified using the primers VO32 and VO33 (Table 1). 5 µl of target DNA was combined with Phusion® Flash High-Fidelity PCR Master Mix (Thermo Scientific). Two PCR reactions were performed to amplify the *AOX1* promoter and terminator sequences from the pPIC9 vector, using JC165, VO34, VO35 and VO36 as primers. Both PCRs were performed using Phusion® Flash High-Fidelity PCR Master Mix (Thermo Scientific). pEYA2 vector DNA was digested using *Nde*I and *Eco*RV. In the reaction mixture, 10 µl of pEYA2 vector DNA was mixed with 2 µl of buffer D (Promega), 1 µl of NdeI (Promega), 1 µl of EcoRV (Promega) and 6 µl of H_20. The mixture was incubated at 37°C for 1 hour. The *URA3*-bait-*URA3*, *AOX1* promoter, *AOX1* terminator and cut pEYA2 fragments were all purified from gel. The bands were cut from the gel and 3 volumes of 6M sodium iodide were added for each 1 volume of gel slice. The agarose was melted at 60°C, after which 40 µl of resuspended silica powder was added to the mixture. This mixture was incubated for 10 min at 55°C. The resulting silica powder/DNA complex was then spun down, washed and resuspended in 10-20 µl H_2O. The tubes were then incubated at 60°C for 8 min and centrifuged, after which the supernatant was transferred into a new tube.

The *URA3*-bait-*URA3*, *AOX1* promoter, *AOX1* terminator and digested pEYA2 fragments were then assembled through homologous recombination in yeast (Gibson et al., 2008). Four overnight cultures were started for the *Δura*3 yeast mutant in 50 ml of YPD medium (10 g yeast extract, 20 g peptone and 20 g D-glucose) at 30°C and 230 rpm. The next day, 2 mL of each culture was diluted in 50 mL of YPD and grown for five hours. The cells were pelleted, washed and resuspended in sterile water. For each reaction, 50 µL of yeast cells were mixed with 50 µL of boiled salmon sperm DNA (2 µg/µL), 32 µL of 1 M lithium acetate, 240 µL of freshly made 50% PEG 4000 (Sigma-Aldrich) and 300 ng of each fragment. The samples were first incubated at 30°C for 30 minutes and then at 45°C for 15 minutes. Subsequently, the cells were pelleted, resuspended and plated on yeast synthetic drop-out medium (1.7 g yeast nitrogen base without amino acids (Fluka), 5 g ammonium sulphate, 5 g casein hydrolysate (Fluka), 20 mg adenine (Sigma) and 20 g glucose per 1000 ml). After 3 days of incubation at 30°C, recombinant plasmids were isolated from the yeast colonies using a Zymoprep™ Yeast Plasmid Miniprep II Kit (Zymo Research). For each construct, 5 µL of isolated plasmid was mixed with 50 µL of chemically competent DH5α cells and a heat shock was performed at 42°C for 45 seconds (Sambrook et al., 1989). The transformed cells were plated on LB/chloramphenicol medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, 35 µg/mL chloramphenicol) and incubated at 37°C. Two resistant colonies per construct were picked to start overnight cultures in LB medium supplemented with 35 µg/mL chloramphenicol. Recombinant plasmids were isolated from the cultures using a Plasmid Mini Kit (QIAGEN) and sent for sequencing (Macrogen).

Table 1. List of primers used in this study. Bold letters indicate overlapping regions or restriction sites.

Number	Sequence
VO2	**TTTGGATCCGCTCTCGAGATCCTC**TCGAAAGCTACATATAAGGAACGTGC
VO3	**ACATGATCGCGTTCGAGCACATTCC**TCGAAAGCTACATATAAGGAACGTGC
VO4	**AGATGGAGGGGGCCGGCTTTCGCGT**TCGAAAGCTACATATAAGGAACGTGC
VO6	**ATGGGTATTGGAAACCTCATCTTG**TCGAAAGCTACATATAAGGAACGTGC
VO13	**atgatccaaggtggtgacccaacc**TCGAAAGCTACATATAAGGAACGTGC
VO15	**TTTCCCGCAGCGATGATTACTCGT**TCGAAAGCTACATATAAGGAACGTGC
VO16	**ATCAAAGGAATCTACGGGATGAAT**TCGAAAGCTACATATAAGGAACGTGC
VO17	**GCTACCCCAAACGAATCCCAACAGC**TCGAAAGCTACATATAAGGAACGTGC
VO32	AGCGGAAGTGTATCGTACAG
VO33	CATGATTTATCTTCGTTTCCTGCAG
VO34	AATTAACTTCAGCCTCT**GGCGCGCC**ATCTTCTCAAGTTGTCGTTAAAAGTCG
VO35	CGCGCCAGAGGCTGAAG**TTAATTAA**GCGGCCGCGAATTAATTCGCCTTAGACATG
VO36	**TAATACGACTCACTATAGGGGAT**TAAGCTTGCACAAACGAACTTCTCAC
JC164	ACTGTTATTTTTCAGATCTTACCGCTGTTGAGATCCAG
JC165	**CAACAGCGGTAAGAT**CTGAAAAATAACAGTTATTATTCGAG
JC200	**TAATACGACTCACTATAG**Ggtaagcaaactgaagggtgctag
JC201	**TAATACGACTCACTATAG**Ggtgagtaaatccgaagggcagaag
JC202	**TAATACGACTCACTATAG**Ggtacgcaccactacacgatctatg
JC203	CTAAACATCATATATCAGTAC
JC204	**TAATACGACTCACTATAG**Ggtaagaaattgaagtgcagaag
JC205	**TAATACGACTCACTATAG**Ggtaagaaaaccgaagggcagaag
JC206	**TAATACGACTCACTATAG**Ggtaagtagaccggaagggagagcg
JC207	CTATCAAAGCCGAAGAGGC
JC208	**TAATACGACTCACTATAG**GGatgagcaaaggcgaagaactgtttac
JC209	GGTAATGCCCGCCGCGGTC
qcf01-R	CTAAAAGCAAAGGGTACAAG
qcf02-R	CTAAACAGAAGGGTGTTAGC
qmg01-R	CTGCAAAGCAAAGCGTATAA
qmg02-R	CTAAAACAGAAGGGCAGAAG

Table 2. List of Introner-Like Elements (ILEs), Regular Spliceosomal Introns (RSIs) and bait sequences used in this study. Cf refers to *Cladosporium fulvum* and mg refers to *Zymoseptoria tritici*. Cf01, cf02, mg01 and mg02 are ILEs. CfRSI02 and mgRSI02 are RSIs. Bold letters indicate overlapping regions.

Name	Sequence
cf01	gtaagcaaactgaagggtgctaggcatcttcgcgtgtctgagacatctgctgaccttgtaccctttgcttttag
cf02	gtgagtaaatccgaagggcagaaggcatcatctgtctgaaacatgcgcgcggcatcccagctcagtagcagttgctgagacattgccacgtggtgttcgagtatggctaacaccttctgtttag
cfRSI02	gtacgcaccactacacgatctatgacccatacgagctttcgagaggcgtggaatgagacacgatgtgtgggaatgtgtactgatatatgatgtttag
mg01	gtaagaaattgaagtgcagaaggcctccattgtccgagactgtctcggacgccgggccttcgtccattccacactgtctacgacatccgctgaccttatacgctttgctttgcag
mg02	gtaagaaaaccgaagggcagaaggcctgcagctgtccgagattgtctctcgggctatgccagcaggtctcagtctgccttataccctgtgtccgaaccattgctgaccttctgccttctgttttag
mgRSI02	gtaagtagaccggaagggagagcgtgtctccagcttgccatgaccatgccttcgatcagctagggatgtatcatcgacttcactaaccctgcctcttcggctttgatag
bait-cf01011	**ctgcaggaa acgaa gataa atcatg**TATATTCCCGACCGGGTGGCATGGCCTCGT
bait-cf01020	**ctgcaggaa acgaa gataa atcatg**atcatcaagaacttcatgatccaagggtggtgacccaacc
bait-cf02043	**ctgcaggaa acgaa gataa atcatg**TTCGTTAGGTACCGGGCAGTGCAATGTGAAGGCGTA
bait-cf03014	**ctgcaggaa acgaa gataa atcatg**CTCATCACACCATGGAATTTTCCCGCAGCGATGATTACTCGT
bait-cf03030	**ctgcaggaa acgaa gataa atcatg**GATGCGATCAAAGGAATCTACGGGATGAAT
bait-cf07087	**ctgcaggaa acgaa gataa atcatg**CACTGGTGTAGGCTACCCCAAACGAATCCCAACAGC
bait-mg067	**ctgcaggaa acgaa gataa atcatg**gccatcatcttcttcggggtgttaatgaacgcgtttggatccgctctcgagatcctc
bait-mg214	**ctgcaggaa acgaa gataa atcatg**gcacttcgcagcggtggcgggagacatgatcgcgttcgagcacattcc
bait-mg224	**ctgcaggaa acgaa gataa atcatg**gcaggagatggaggggcccggctttcgcgt
bait-mg272	**ctgcaggaa acgaa gataa atcatg**ccaaagcatgtaatggttttctgattggaat
bait-mg326	**ctgcaggaa acgaa gataa atcatg**gacatcacggctctcatgacataccgtccatgttcatgggtattggaaacctcatcttg
bait-mg331	**ctgcaggaa acgaa gataa atcatg**ctcatctgctgggagcacgatgcccctccatgac

3.2 Mobility assay

The expression vectors pBait02043-cf01, pBait02043-cf02, pBait02043-cfRSI, pBait331-mg01, pBait331-mg02 and pBait331-mgRSI were available in the laboratory (Ochoa Tufiño, 2013). The first part of the vector's name refers to the bait sequence that it contains, and the second part refers to the ILE/RSI that is cloned in it (Table 2). For example the expression vector pBait02043-cf01 contains the bait sequence "bait-cf02043" and the ILE cf01. Those vectors were transformed into the temperature sensitive *prp2-1* yeast mutant (provided by Pr. Lührmann). *Prp2-1* yeast cells were cultured overnight in YPD medium (10 g yeast extract, 20 g peptone and 20 g D-glucose per 1000 ml) at 25°C and 230 rpm. The next day, the cells were pelleted, washed and resuspended in sterile water. For each reaction, 50 µL of *prp2-1* cells were mixed with 50 µL of salmon sperm DNA (2 µg/µL), 150 ng of plasmid DNA and 250 µL of freshly made 50% PEG 4000. The mixture was incubated at 42°C for 15 min and electroporation was performed (1,5 kV, 25 µF, 200 ohm). Subsequently, the cells were pelleted, resuspended and plated on yeast synthetic drop-out medium. Resistant colonies were screened by colony PCR using the primers JC208 and JC209 (Table 1). The colony PCRs were performed using GoTaq® DNA Polymerase (Promega). The used PCR conditions were 2 min 95°C, 40 cycles of 1 min 95°C, 1 min 60°C, 1 min 72°C and a final step of 5 min at 72°C.

Positive *prp2-1* transformants were cultured overnight in 25 mL of BMGY (10 g yeast extract, 20 g peptone, 100 mL 1M potassium phosphate buffer pH 6.0, 100 mL 10x YNB, 2 mL 500x biotin, 100mL 10x glycerol per 1000 ml) supplied with 200 µg/mL ampicillin at 25°C and 230 rpm. The next day, cultures were diluted to an OD of 1.0 in 6 mL of BMMY (BMGY medium in which the 10x glycerol is replaced by 10x methanol) supplemented with 200 µg/mL ampicillin. Every day the cultures were supplemented with 1% methanol. Samples of 1 mL were taken at 6, 24 and 48 hours after induction. The cells were pelleted and screened under long UV light (365 nm) for GFP fluorescence. RNA was extracted from the samples using a NucleoSpin RNA extraction kit (Macherev-Nagel). Subsequently, cDNA was synthesized from the resulting RNA samples using a M-MLV reverse transcriptase kit (Promega). A mixture containing 2 µl of oligo DTs and 13 µl of RNA sample was incubated at 70°C for 5 minutes. After cooling the mixture on ice, 5 µl of 5x reaction buffer, 1,25 µl of dATP, 1,25 µl of dCTP, 1,25 µl of dGTP, 1,25 µl of dTTP, 0.5 µl of RNAsin and 1 µl of NLV-RT were added. The samples were then incubated at 37°C for 60 minutes. A RT-PCR was then performed on the resulting cDNA samples using the primers JC208 and JC209 (Table 1). GoTaq® DNA Polymerase (Promega) was used to perform the RT-PCR.

3.3 *In vitro* transcription

SfGFP genes containing the ILEs cf01, cf02, cfrsi, mg01, mg02 and mgrsi (Table 2) were amplified from the corresponding expression vectors using JC208 and JC209 (Table 1) as primers. GFP without any insert was also amplified from pFLAG-NLS-sfGFP using the same primers. The ILEs/RSIs cf01, cf02, cfRSI, mg01, mg02 and mgRSI (Table 2) were amplified from the corresponding expression vectors using JC200, JC201, JC202, JC203, JC204, JC205, JC206, JC207, qcf01-R, qcf02-R, qmg01-R, and qmg02-R (Table 1) as primers. Phusion® Flash High-Fidelity PCR Master Mix (Thermo Scientific) was used to perform all the PCRs. 7 µL of unpurified PCR product was used to perform *in vitro* transcription with the HiScribe™ T7 *In Vitro* Transcription Kit (New England Biolabs) following the manufacturer's recommendations. The ILEs within *sfGFP* and the synthetic molecules were respectively incubated for 4 hours and overnight at 42°C. Ethanol precipitation was used to purify the resulting RNA samples.

3.4 *In vitro* self-splicing assay

500 ng of the amplified ILEs within *sfGFP* was mixed with 5 µL 2x reaction buffer (80 mM Tris-HCl, 200 mM $MgCl_2$ and 2 M NH_4Cl) and RNAse free water (to 10 µL). GFP without any insert was also tested as a

control. This mixture was incubated at 45°C, after which it was run on gel. The self-splicing assay was performed using $MgCl_2$ concentrations of 10 and 100 mM, pH 7.5 and 8.5, incubation temperatures of 25°C and 42°C and incubation times of 0, 60 and 180 minutes.

3.5 *In vitro* ribozyme activity assay

The bait sequences cf02043 and mg331 (Table 2) were amplified from the corresponding bait vectors using the primers VO32 and JC164 (Table 1). The forward and reverse strand were also amplified separately using asymmetric PCR. To amplify the forward strand, 1 µL of 10 nM VO32 was used in combination with 1 µL of 0.1 nM JC164. The reverse strand was amplified by combining 1 µL of 10 nM VO32 with 0.1 µL of 1 nM JC164. The PCR reactions were performed using GoTaq® DNA Polymerase (Promega).

10 pmol of the amplified ILEs/RSIs was incubated at 65°C for 2 minutes. It was then cooled on ice for 2 minutes, and incubated at 37°C for 5 minutes. 500 ng of the target DNA was incubated at 95°C for 5 minutes. The amplified ILEs/RSIs were then mixed with the target DNA, 20 µl of 2x reaction buffer (200 mM KCl, 100 mM Tris-HCl and 10 mM $MgCl_2$) and RNAse free water (to 40 µL). This mixture was incubated at 37°C, after which it was run on gel. The ribozyme activity assay was performed using double- and single-stranded target DNA, circular and linear target DNA, $MgCl_2$ concentrations of 2 and 20 mM, pH 7.5 and 8.5 and incubation times of 0, 30, 60 and 120 minutes and overnight.

4. Results

Introner-Like Elements (ILEs) are a new type of spliceosomal introns that appear to multiply in unrelated genes in at least six fungal species (van der Burgt et al., 2012). Because the majority of intron gains in fungi could not be associated with other types of intron duplication, ILE multiplication is suggested to be the main mechanism behind intron gain in fungi (Collemare *et al*., 2013). However, the mechanism is still unknown. A proposed mechanism for ILE multiplication is similar to retro-homing of self-splicing group II introns, which are ribozyme RNA molecules that insert directly into double-stranded DNA (Guo et al., 1997; Zimmerly et al., 1995). The aim of this research is to prove the mobility of ILEs *in vivo*, and assess the self-splicing- and ribozyme activity of ILEs *in vitro*. ILEs present in the genomes of *C. fulvum* and *Z. tritici* were grouped in different clusters based on sequence identity. It is expected that active elements are long and share a high pairwise identity. Therefore, the consensus sequence of the ILE clusters cf01, cf02, mg01 and mg02 will be used in all experiments. The abbreviations cf and mg refer to *C. fulvum* and *Z. tritici* respectively.

4.1 Mobility assay

In order to prove the mobility of ILEs *in vivo*, a genetic screen has been developed to identify new insertion events. Destination vectors for a mobility assay in *Saccharomyces cerevisiae* were previously constructed (Ochoa Tufiño, 2013) (Figure 3). Because ILE insertion events are expected to be rare, it is necessary to make use of a positive selection marker. However, the selection marker gene will be disrupted when an ILE is inserted into it. Therefore the *URA3* gene was chosen to serve as a dual selection marker. The *URA3* gene encodes for the enzyme orotidine-5'-phosphate decarboxylase that is involved in the biosynthesis of uracil, but is also responsible for the conversion of 5-fluoro-orotic acid (5-FOA) into 5-fluoroacil, which causes cell death by interfering with DNA and RNA metabolism (Boeke et al., 1987). Therefore, *S. cereviseae* cells expressing the *URA3* gene will be uracil prototroph and 5-FOA sensitive. When the *URA3* gene is disrupted, the yeast cell will lose uracil prototrophy and gain 5-FOA resistance. The destination vector also contains the methanol inducible alcohol oxidase (AOX) promoter and terminator. The gateway cassette is used to clone ILEs within *sfGFP* into the vector through LR reactions. After the expression vectors have been transformed into *S. cereviseae*, expression of these ILEs within *sfGFP* can be induced by the addition of methanol. In order to increase the chance of an ILE inserting into the *URA3* gene, different bait sequences corresponding to ILE insertion hotspots were cloned downstream of the start codon of the *URA3* gene (Ochoa-Tufiño, 2013). A bait sequence is a short stretch of 30-60 nucleotides that has been identified as containing several different introns, including ILEs, in the same or different isolates of one or more given species.

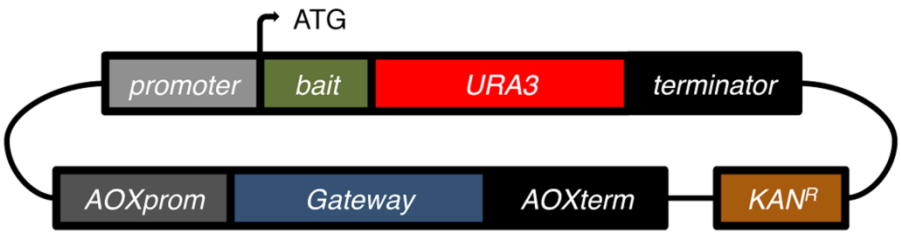


Figure 3. Destination vector for ILE mobility assay in yeast (Ochoa Tufiño, 2013).
The features of this destination vector include an *URA3* promoter and coding sequence interrupted by a bait sequence and a methanol inducible alcohol oxidase (AOX) promoter followed up by a gateway cassette and a terminator. The gateway cassette consists of a chloramphenicol resistance gene and the lethal *ccdB* gene, flanked by attR recombination sites.

ILEs are typical spliceosomal introns (van der Burgt et al., 2012), meaning that they are spliced from the primary transcript. Therefore, the insertion of an ILE into a gene should not disrupt gene function. To solve this problem, the genetic screen will be carried out in the temperature sensitive yeast mutant *prp2-1*. This mutant suffers from a splicing defect at temperatures above 30°C (Kim et al., 1999). When the *prp2-1* mutant is transformed with an expression vector, grown at 25°C and *sfGFP* expression is induced by the addition of methanol, the ILE inserted into *sfGFP* should be spliced from the primary transcript resulting in the production of a functional *sfGFP* protein. The excised ILE can then insert into the bait sequence. Subsequently, ILE insertions can be detected by growing the yeast colonies on 5-FOA at restrictive temperature. When an ILE is inserted into the bait sequence, *URA3* translation will be prevented by an early stop codon, causing the *S. cereviseae* cells to lose uracil auxotrophy and gain 5-FOA resistance. Survival of a colony indicates the capture of an ILE in the *URA3* gene.

The expression vectors pBait02043-cf01, pBait02043-cf02, pBait02043-cfRSI, pBait331-mg01, pBait331-mg02 and pBait331-mgRSI were already available in the laboratory (Ochoa Tufiño, 2013). *Prp2-1* mutant cells were transformed with these vectors and selected on a medium without uracil. The transformation was only successful for pBait02043-cf01 and pBait331-mg02. Transcription of *sfGFP* containing the ILEs was induced by culturing the transformant *prp2-1* cells at 25°C in the presence of methanol. Samples were taken at 0, 6, 24 and 48 hours after incubation (hai). These samples were centrifuged and the pellets were screened under UV light. As expected, no GFP signal was detected in any of the samples at 0 hai. Because the *prp2-1* strain does not suffer from a splicing defect at 25°C, it was expected that the ILEs would be spliced and the *prp2-1* cells would produce a functional GFP protein, resulting in fluorescence in the 6, 24 and 48 hai samples. However, no GFP signal could be detected in any of them. To check that *sfGFP* containing ILEs was indeed expressed, RNA was isolated from the samples to perform RT-PCR. The PCRs were also carried out using the RNA samples as a template to check for genomic DNA contamination (Figure 4). On the gel bands of 816 and 869 bp were visible for the RNA control of the samples pBait02043-cf01 6 hai, 48 hai and pBait331-mg02 24 hai respectively, indicating that they were contaminated with genomic DNA. Similar bands were visible for the samples pBait02043-cf01 24 hai and pBait331-mg02 6 hai. The size of the bands corresponds to the size of the *sfGFP* molecule that still contains the ILE. In case of a splicing event, the PCR fragment would have been smaller. Therefore it can be concluded that *sfGFP* containing an ILE is indeed expressed in the samples pBait02043-cf01 24 hai and pBait331-mg02 6 hai, but the ILE is not spliced from the primary transcript. No band was visible for pBait331-mg02, probably because the cDNA concentration in this sample was too low.

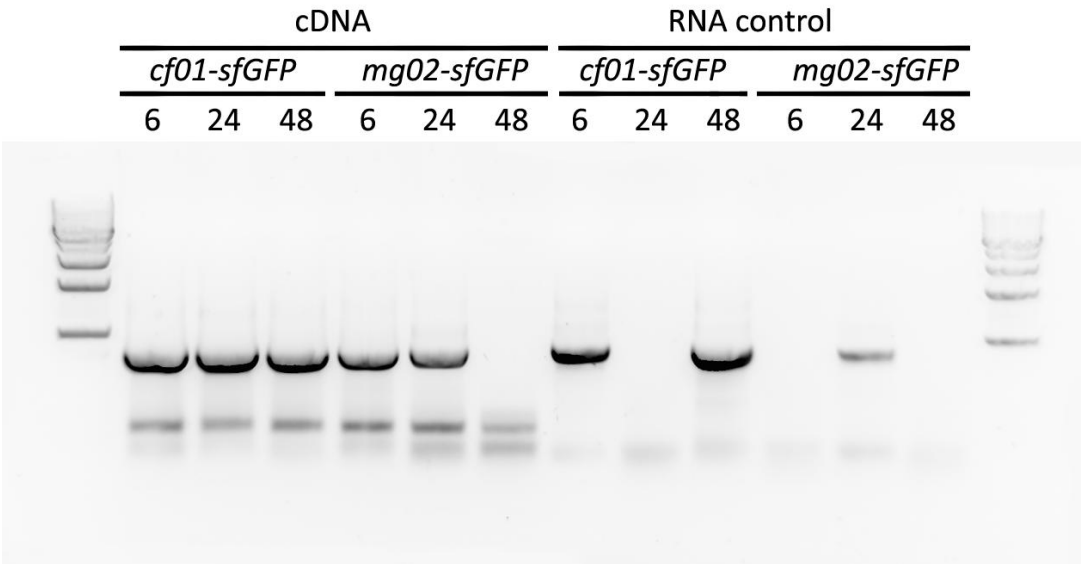


Figure 4. RT-PCR on *prp2-1* transformants.
Prp2-1 transformants carrying the expression vector pBait02043-cf01 or pBait331-mg01 were grown at 25°C and transcription of *sfGFP* containing the ILEs was started by the addition of methanol. Samples were taken at 6, 24 and 48 hours after incubation (hai). RNA was extracted from the samples and used to perform RT-PCR. To check for genomic DNA contamination, RNA controls were also included. An 1kb step ladder (Promega) was used. Bands of 816 and 869bp are visible for the RNA control of the samples pBait02043-cf01 6 hai, 48 hai and pBait331-mg02 24 hai respectively, indicating genomic DNA contamination. The size of the bands corresponds to the size of the *sfGFP* molecule that still contains the ILE. Bands of the same size are visible for the samples pBait02043-cf01 24 hai and pBait331-mg02 6 hai, indicating that the *sfGFP* is indeed expressed in those samples, but the ILE is not spliced from the primary transcript. In case of a splicing event, a band of 740bp corresponding to the size of *sfGFP* without an ILE is expected.

4.2 *In vitro* self-splicing assay

Group II introns are suggested to be the predecessors of spliceosomal introns (Cavalier-Smith, 1991; Roy and Gilbert, 2006) and it has been hypothesized that ILEs multiply through a mechanism similar to the retro-homing of group II introns. This raises the question if ILEs are self-splicing. Therefore, an *in vitro* assay was set up to test the self-splicing ability of ILEs. Experimental conditions were based on previous *in vitro* self-splicing assays conducted on group II introns (Hebbar et al., 1992). The optimal self-splicing temperatures for group IIA and group IIB self-splicing introns in yeast are respectively 42°C and 45°C, while reactions are essentially devoid at 30°C (Hebbar et al., 1992). Self-splicing activity was observed in group IIB introns at Mg^{2+} concentrations of 10mM, while group IIA introns required much higher Mg^{2+} concentrations (Hebbar et al., 1992). *sfGFP* RNA molecules containing the ILEs cf01, cf02, mg01 and mg02 were obtained using *in vitro* transcription. Subsequently, they were mixed with a 2x reaction buffer and incubated at the previously described conditions. In addition, a different temperature (25°C) and Mg^{2+} concentration (100mM) were tested because ILEs might require slightly different conditions. Although pH does not seem to play a significant role in self-splicing of group II introns (between 6,5 and 8,5; Hebbar et al., 1992), pHs of 7,5 and 8,5 were tested. Samples were taken at various time points and run on gel to screen for self-splicing events (Figure 5). A single band can be observed for all samples, which corresponds to the size of the *sfGFP* molecule containing an ILE. If a self-splicing event would have taken place in one of the samples, multiple bands of smaller sizes would have been present on the gel. Therefore, it can be concluded that no self-splicing event took place in any of the samples.

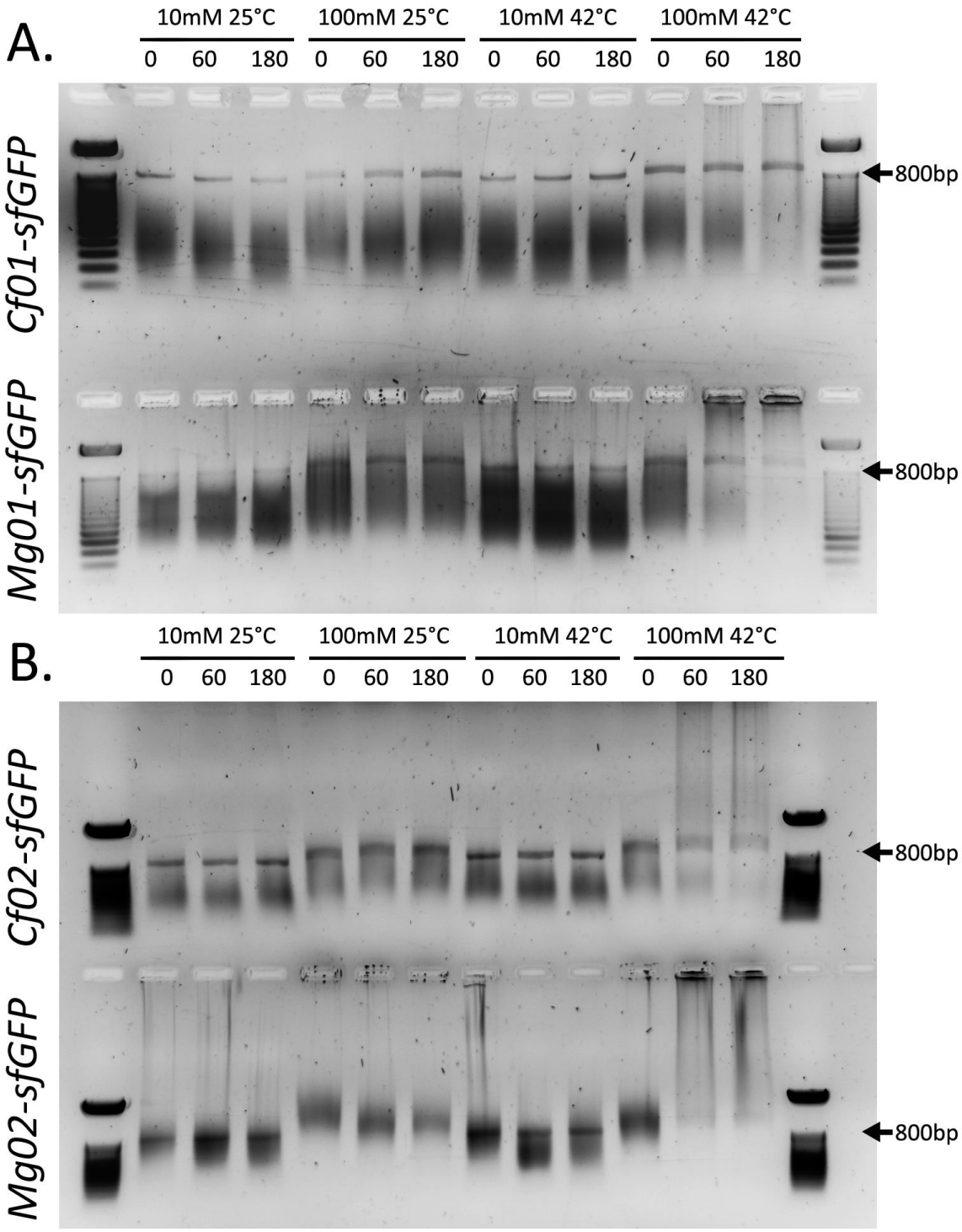


Figure 5. *In vitro* self-splicing assay.

sfGFP RNA containing either cf01, cf02, mg01 or mg02 were mixed with a 2x reaction buffer. $MgCl_2$ concentrations of 10mM and 100mM, incubation temperatures of 25°C and 42°C and incubation times of 0, 60 and 180 minutes were tested. Also, pHs of 7,5 (**A**) and 8,5 (**B**) were tested. The samples were run on gel to screen for self-splicing events. A 50bp step ladder (Promega) was used as a reference. For the samples cf01, cf02, mg01 and mg02 single bands with respective sizes 816bp, 866bp, 855bp and 869bp are visible on the gel. These sizes correspond to *sfGFP* still containing an ILE, indicating that no self-splicing event took place.

4.3 *In vitro* ribozyme activity assay

If ILEs do indeed multiply through a mechanism similar to retro-homing of self-splicing group II introns, cutting of the recipient DNA is an essential step. Therefore, an *in vitro* assay was set up to test if ILEs possess ribozyme activity. The bait sequences cf02043 and mg331 were amplified from the corresponding expression vectors using both normal and asymmetric PCR (Poddar, 2000), resulting in double- and single-stranded target DNA respectively. Circular target DNA was also included in the experiment. The ILEs cf01 and mg01 were amplified from their corresponding expression vectors and transcribed into RNA. Subsequently, they were mixed with the various target DNAs and incubated under various conditions. The optimal conditions for *in vitro* ribozyme assays with group II introns have previously been defined as 20mM $MgCl_2$, pH 6,5-8,5 and temperatures of 30°C-37°C (Zimmerly et al., 1995). Because it is possible that ILEs behave somewhat different from group II introns, it was decided to test a concentration of 2mM $MgCl_2$ in addition to the 20mM $MgCl_2$ concentration. Although pH does not seem to play a significant role in ribozyme activity of group II introns (between 6,5 and 8,5) several pHs within this range have been used in successful *in vitro* ribozyme activity assays (Zimmerly et al., 1995; Liu et al.,2002). Therefore it was decided to test both pHs of 7,5 and 8,5. It was decided to use an incubation temperature of 37°C for all assays because that is the most commonly used temperature in previous *in vitro* ribozyme activity experiments (Zimmerly et al., 1995; Liu et al., 2002). Samples were taken at various time points and run on gel to detect ribozyme activity (Figure 6). For the assays using plasmid target DNA (Figure 6a) two bands are visible on the gel. Those two bands represent different conformations of the plasmid, namely the supercoiled- and open-circular conformation (Voordouw et al., 1978). If the target DNA would have been cut, a single band of about 7,4 kb corresponding to the linearized vector would have been present on the gel. For the assays using linear target DNA (Figure 6b) single bands of about 1,5 kb were visible on the gel both for cf02043 and mg331. These bands correspond to the size of the uncut targets. If the target DNA would have been cut, two bands of 1100 bp and 350 bp would have been present on the gel. Therefore it can be concluded that none of the target DNA was cut in any of the samples.

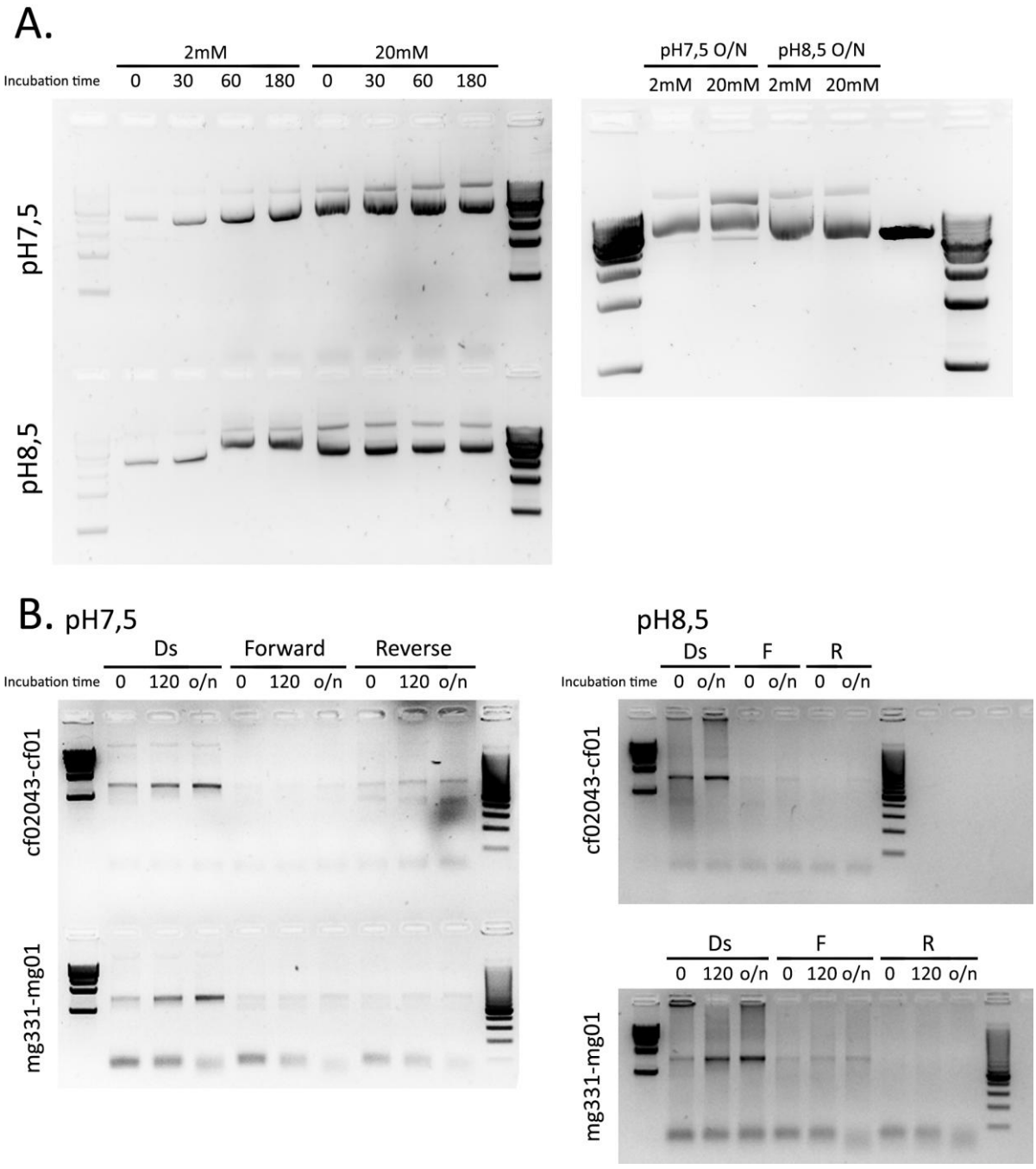


Figure 6. *In vitro* ribozyme activity assay using circular (A) and linear DNA (B) as a target.
The ILEs cf01 and mg01 were mixed with cf02043 and mg331 respectively. Samples were mixed with a 2x reaction buffer and incubated at 37°C. $MgCl_2$ concentrations of 2mM and 20mM, pHs of 7,5 and 8,5 and incubation times of 0, 30, 60, 180 minutes and overnight (o/n) were tested. The samples were run on gel to check for ribozyme activity. **A:** the ILE mg01 was incubated with mg331 double-stranded plasmid DNA as a target. On the gel, two bands corresponding to two conformations of the plasmid are visible for all the samples, indicating that the target DNA was not cut. **B:** the ILEs cf01 and mg01 were incubated with linear cf02043 and mg331 DNA as a target. Double-stranded and single stranded (forward and reverse) target DNA was tested. The $MgCl_2$ concentration was 20mM. On the gel, single bands are visible for all the samples, indicating that the target DNA was not cut. If the target DNA would have been cut, multiple bands of a smaller size would have been present on the gel.

5. Discussion

5.1 Mobility assay

The characteristics of ILEs suggest that they are mobile genetic elements that are able to proliferate within genomes. It is hypothesized that they are the predecessors of RSIs in at least six fungal species (van der Burgt, 2012). To assess the mobility of ILEs *in vivo*, a genetic screen has been developed (Ochoa Tufiño, 2013). This screen makes use of a construct containing the dual selection marker URA3 with a bait sequence cloned downstream of its start codon. The construct also contains ILEs/RSIs within *sfGFP* under the control of the methanol inducible alcohol oxidase (AOX) promoter. Constructs containing different combinations of ILEs and bait sequences were transformed into the *S. cerevisiae prp2-1* mutant, which suffers from a splicing defect at restrictive temperatures. After inducing *sfGFP* expression in the *prp2-1* transformants under permissive temperature, it was expected that the ILE would be spliced from the primary transcript, resulting in a functional *sfGFP* protein and fluorescence. However, no *sfGFP* activity could be detected in any of the samples in the first 48 hours after the transcription was started. The RT-PCR results suggested that *sfGFP* containing the ILEs was indeed expressed, but the ILE was not spliced from the precursor RNA. When the ILE is not spliced from the precursor RNA, an early stop codon within the ILE prevents *sfGFP* from being translated. This explains why no *sfGFP* fluorescence could be detected in any of the samples. A possible explanation for the absence of splicing is that the *S. cerevisiae* spliceosome is not able to recognize introns from *C. fulvum* and *Z. tritici*. While the ability of eukaryotes to splice introns has been conserved throughout evolution, the signals by which they are recognized have differentiated considerably. At some point, the splicing mechanisms of distantly related species can become incompatible with each other (Iwata and Gotoh, 2011). For example, introns from the rabbit *β*-globin gene were shown not to be spliced in *S. cerevisiae* either (Beggs et al., 1980). In another study in *S. cerevisiae*, four introns were shown not to be spliced from the *A. oryzae* glucoamylase gene (Innis et al., 1985).

The finding that ILEs are not spliced from the primary transcript in *S. cerevisiae* allows for a different mobility assay. Because ILEs are not spliced, the insertion of an ILE into the bait sequence should disrupt the *URA3* gene. There is no need to work with the temperature sensitive *prp2-1* mutant anymore. *S. cerevisieae* cells can be transformed with a similar expression vector in which the ILEs are not inserted into *sfGFP*. In this way, the ILEs are produced directly after which they can insert into the bait sequence. Insertion events can then be detected by culturing the *S. cerevisieae* cells on 5-FOA. The genome of *S. cerevisiae* is known to contain only 294 introns, none of which can be identified as ILEs (Neuvéglise et al., 2011). It is possible that *S. cerevisiae* lacks one or more proteins that are essential to splice ILEs. Therefore, another approach is to carry out a similar mobility assay in an eukaryotic organism with a genome that contains ILEs. Because there is evidence for recent ILE multiplication events in *Z. tritici* (Torriani et al., 2011; van der Burgt et al., 2012), this fungus is very well suited for a follow up screen. Because ILEs are spliced in *Z. tritici*, the *prp2-1* mutation should be introduced to make the assay work.

5.2. Mechanism of ILE multiplication

The proposed mechanism for ILE multiplication is based on the retro-homing mechanism of self-splicing group II introns. However, it is not known if ILEs are self-splicing like group II introns or if they require the help of the spliceosome to be spliced from the precursor RNA. Also, cutting the recipient DNA is an essential step if the multiplication mechanism of ILEs resembles the retro-homing of group II introns. Therefore, *in vitro* assays were set up to research the self-splicing ability and ribozyme activity of ILEs.

5.2.1 Splicing mechanism

Previous research on the self-splicing ability of ILEs *in vivo* suggested that ILEs are not self-splicing (Ochoa Tufiño, 2013). *sfGFP* containing different ILEs and RSIs was expressed in the prokaryotic bacteria *E. coli*. Because this organism lacks the spliceosome, only self-splicing introns are able to splice from the primary transcript. The results of this experiment showed that the *sfGFP* genes containing the ILEs or RSIs were expressed, but not translated into a functional protein. The introns were not spliced from the primary transcript, causing the translation to be abolished by an early stop codon in the ILEs/RSIs. Similarly, ILEs were also not spliced during the mobility assay in *S. cerevisiae*, providing further evidence that ILEs are not self-splicing. An *in vitro* assay was set up to further address the self-splicing ability of ILEs. In this assay, *sfGFP* RNA molecules containing different ILEs were incubated under different conditions. The experimental conditions were based on *in vitro* self-splicing assays on group II introns (Hebbar et al., 1992). On gel, single bands were visible for all samples, indicating that none of the ILEs was spliced from the *sfGFP* RNA. In case of a self-splicing event, multiple bands of a smaller size would have been present on the gel. These results provide further evidence that ILEs are not self-splicing. Self-splicing assays with *sfGFP*-cf01 and *sfGFP*-mg01 were only carried out at pH7,5, while pH8,5 was only tested for *sfGFP*-cf02 and *sfGFP*-mg02. To confidently conclude that ILEs are not self-splicing, both pHs should be tested. Moreover, additional ILEs should be tested in addition to cf01, cf02, mg01 and mg02. Lastly, it is possible that self-splicing events did take place during the assays but the efficiency was so low that they could not be detected on gel. To check if this is the case a northern blot could be used. This method is much more sensitive because it makes use of a hybridization probe (Alwine et al., 1977).

5.2.2 Ribozyme activity

In the *in vitro* ribozyme activity assay, ILEs were mixed with different types of target DNA including plasmids, double-stranded linear DNA and single-stranded linear DNA. The mixture was incubated under various conditions based on previous *in vitro* ribozyme activity research (Zimmerly et al., 1995; Liu et al.,2002). On gel, double bands were visible for the samples using plasmids as target DNA. These two bands correspond to two different conformations of the plasmid DNA, indicating that the target DNA was not cut. If the target DNA would have been cut, a single band corresponding to the size of the linearized plasmid would have been present on gel. For the assays using linearized DNA as a target, single bands where visible on the gel for all the samples, indicating that the target DNA was not cut. If the target DNA would have been cut, multiple smaller bands would have been present on the gel. The results of this assay could not confirm that ILEs possess any ribozyme activity. However, this is not enough to conclude that ILEs do not possess any ribozyme activity. The only tested plasmid/ILE combination was mg331-mg01. Additional plasmid/ILE combinations should be tested, not only from *Z. tritici* but also from *C. fulvum*. Single stranded target DNAs were only tested under a $MgCl_2$ concentration of 20mM, because this concentration was previously described as the optimal concentration for *in vitro* ribozyme assays with group II introns. However, it is possible that ILEs behave somewhat different from group II introns. Therefore the assays should also be carried out using a $MgCl_2$ concentration of 2mM. The only combinations of linear target DNA and ILEs that were tested where mg331-mg01 and cf02043-cf01. Additional combinations should be tested before drawing a final conclusion on the ribozyme activity of ILEs *in vitro*. Lastly, it is possible that target DNAs were cut during the assays but the efficiency was so low that it could not be detected on gel. To check if this is the case a Southern blot could be used. This method is much more sensitive because it makes use of a hybridization probe (Southern, 1975).

6. Conclusion and perspectives

ILE multiplication is the main mechanism behind intron gain in at least six fungal species (van der Burgt., 2012). The purpose of this study was to unravel the mechanism behind ILE multiplication. This mechanism is hypothesized to be similar to the retro-homing of self-splicing group II introns. However, this study suggests that ILEs are not self-splicing like group II introns, but require help from the spliceosome to be excised from the primary transcript. Although additional research is needed to confirm this, it would explain the fact that ILEs harbor typical splicing features that are also found in RSIs. An essential step in the proposed mechanism of ILE multiplication is the cutting of the recipient genomic DNA. It was hypothesized that ILEs are able to recognize and cut specific regions in the target DNA. However, this study suggests that ILEs do not possess such ribozyme activity. If this is indeed the case, the recipient DNA is probably cut in a different way. For example, it is possible that an exogenous restriction enzyme is involved in the process. Organisms whose genomes do not contain ILEs like *S. cerevisiae* will most likely lack such a protein, meaning that future *in vivo* mobility assays should be performed in a fungus with a genome that possesses ILEs.

References

Alwine JC, Kemp DJ and Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. PNAS 74(12): 5350-5354

Beggs JD, van den Berg J, van Ooyen A and Weissmann C (1980) Abnormal expression of chromosomal rabbit *β*-globin gene in *Saccharomyces cerevisiae*. Nature 283: 835-840

Boeke JD, La Croute F and Fink GR (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Molecular and General Genetics 197: 345-346

Burset M, Seledtsov IA, Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Research 28: 4364-4375

Bonen L and Vogel J (2001) The ins and outs of group II introns. Trends in Genetics 17: 322-331

Cavalier-Smith T (1991) Intron phylogeny: A new hypothesis. Trends in Genetics 7: 145–148

Chalamcharla VR, Curcio MJ and Belfort M (2010) Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. Genes & Development 24: 827-836

Collemare J, van der Burgt A and de Wit PJGM (2013) At the origin of spliceosomal introns: Is multiplication of introner-like elements the main mechanism of intron gain in fungi? Communicative & Integrative Biology 6(2): e23147

Csuros M, Rogozin IB and Koonin EV (2011) A Detailed History of Intron-rich Eukaryotic Ancestors Inferred from a Global Survey of 100 Complete Genomes. PLOS Computational Biology 7: e1002150

Dai L and Zimmerly S (2003) ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. RNA 9: 14-19

Fedorov A, Merican AF and Gilbert W (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. PNAS 99: 16128–16133

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proceedings of the National Academy of Sciences 105: 20404-20409

Guo H, Zimmerly S, Perlman PS and Lambowitz AM (1997) Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO Journal 16: 6835-6848

Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. Trends in Genetics 21: 111–119

Hebbar SK, Belcher SM and Perlman PS (1992) A maturase-encoding group IIA intron of yeast mitochondria self-splices in vitro. Nucleic Acids Research 20: 1747-1754

Hellsten U, Aspden JL, Rio DC and Rokhsar DS (2011) A segmental genomic duplication generates a functional intron. Nature Communications 2

Innis MA, Holland MJ, McCabe PC, Cole GE, Wittman VP and Tal R (1985) Expression, glycosylation, and secretion of an *Aspergillus*glucoamylase by *Saccharomyces cerevisiae*. Science 228: 21-26

Irimia M, Rukov JL, Penny D, Vinther J, Garcia-Fernandez J and Roy SW (2008) Origin of introns by intronization of exonic sequences. Trends in Genetics 24: 378-381

Iwata H and Gotoh O (2011) Comparative analysis of information contents relevant to recognition of introns in many species. BMC Genomics 10: 12-45

Kim DH, Edwalds-Gilbert G, Ren C, Lin RJ (1999) A mutation in a methionine tRNA gene suppresses the *prp2-1* Ts mutation and causes a pre-mRNA splicing defect in Saccharomyces cerevisiae. Genetics 153(3): 1105-15

Koonin EV (2009) Intron-Dominated Genomes of Early Ancestors of Eukaryotes. Journal of Heredity 100(5): 618-623

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy JW (2004) Introns and splicing elements of five diverse fungi. Eukaryotic Cell 3: 1088–1100

Lambowitz AM and Zimmerly S (2004) Mobile Group II Introns. Annual Review of Genetics 38: 31-35

Lambowitz AM and Zimmerly S (2011) Group II Introns: Mobile Ribozymes that Invade DNA. Cold Springer Harbor Perspective in Biology 3: a003616

Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. Trends in Biochemical Science 28: 215–220

Li W, Tucker AE, Sung W, Thomas WK and Lynch M (2009) Extensive, Recent Intron Gains in *Daphnia* Populations. Science 326: 1260-1262

Lilley DMJ (2010) Mechanisms of RNA catalysis. Philosophical Transactions of the Royal Society: Biological Sciences 366: 2910-2917

Liu R, Rohe B, Carson DD and Farach-Carson MC (2002) A Rapid and Simple Nonradioactive Method for *In Vitro* Testing of Ribozyme Activity. Antisense And Nucleic Acid Drug Development 12: 283–288

Lykke-Andersen J, Aagaard C, Semionenkov M and Garrett RA (1997) Archaeal introns: splicing, intercellular mobility and evolution. Trends in Biochemical Sciences 22: 326-331

Malardier L, Daboussi MJ, Julien J, Roussel F, Scazzocchio C, Brygoo Y (1989) Cloning of the nitrate reductase gene (niaD) of Aspergillus nidulans and its use for transformation of Fusarium oxysporum. Gene 78(1): 147-56

Neuvéglise C, Marck C and Gaillardin C (2011) The intronome of budding yeasts. Comptes Rendus Biologies 334: 662-670

Ochoa Tufiño AV (2013) Mobility of fungal Introner-Like Elements. Master thesis.

Tseng C and Cheng S (2008) Both Catalytic Steps of Nuclear Pre-mRNA Splicing Are Reversible. Science 320: 1782-1784

Voordouw G, Kam Z, Borochov N and Eisenberg H (1978) Isolation and physical studies of the intact supercoiled: The open circular and the linear forms of $CoIE_1$-plasmid DNA. Biophysical Chemistry 8(2): 171-189

Poddar SK (2000) Symmetric vs asymmetric PCR and molecular beacon probe in the detection of a target gene of adenovirus. Molecular and Cellular Probes 14(1): 25–32

Robart AR and Zimmerly S (2005) Group II intron retroelements: function and diversity. Cytogenetic and Genome Research 110: 589–597

Rogers J (1985) Exon shuffling and intron insertion in serine protease genes. Nature 315: 458-459

Rogozin IB, Carmel L, Csuros M and Koonin EV (2012) Origin and evolution of spliceosomal introns. Biology Direct 7(11)

Roy SW and Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. Nature Reviews Genetics 7: 211-221

Roy SW and Irimia M (2009) Mystery of intron gain: new data and new models. Trends in Genetics 25: 67–73

Roy SW and Irimia M (2009) Splicing in the eukaryotic ancestor: form, function and dysfunction. Trends in Ecology & Evolution 24: 447-455

Roy SW and Irimia M (2012) Genome Evolution: Where Do New Introns Come From? Current Biology 22: R529–R531

Sambrook J, Fritsch EF, & Maniatis, T (1989). Commonly used techniques in molecular cloning. Molecular cloning: a laboratory manual, 3.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98(3): 503–517

Torriani SFF, Stukenbrock EH, Brunner PC, McDonald BA and Croll D (2011) Evidence for Extensive Recent Intron Transposition in Closely Related Fungi. Current Biology 21(23): 2017–2022

van der Burgt A, Severing E, de Wit PJGM and Collemare J (2012) Birth of New Spliceosomal Introns in Fungi by Multiplication of Introner-like Elements. Current Biology 22(13): 1260–1265

Will CL, Lührmann R (2011) Spliceosome structure and function. Cold Springer Harbor Perspectives in Biology 3(7): pii: a003707

Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryote Micromonas. Science 324(5924): 268-272

Yenerall P and Zhou L (2012) Identifying the mechanisms of intron gain: progress and trends. Biology Direct 7(29): 1745-6150

Yu X and Gabriel A (1999) Patching Broken Chromosomes with Extranuclear Cellular DNA. Molecular Cell 4: 873–881

Zimmerly S, Guo H, Perlman PS, Lambowltz AM (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82: 545-554

Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS and Lambowitz AM (1995) A Group II lntron RNA Is a Catalytic Component of a DNA Endonuclease Involved in lntron Mobility. Cell 83: 529-538