The Dissertation Committee for Sujin Lee certifies that this is the approved version of the following dissertation:

Exploring intron mobilization and detection of an intron gain event *via* intron transposition using a novel intron gain and loss reporter

Committee:

_____
Scott Stevens, Supervisor

_____
Dean Appling

_____
Karen Browning

_____
Arlen Johnson

_____
Rick Russell

Exploring intron mobilization and detection of an intron gain event *via* intron transposition using a novel intron gain and loss reporter

by

Sujin Lee, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2015

Dedication

To my husband Yeonjong and my family, for their endless love.

Acknowledgements

This dissertation would have not been possible without the guidance of my graduate advisor Dr. Scott Stevens. I would like to express my gratitude to him for his support throughout my graduate study. I greatly appreciate his direction and encouragement in the course of completing this study. And I would like to thank Dr. Arlen Johnson for his thoughtful suggestions and advice. I would also like to express my appreciation to my committee members, Dr. Dean Appling, Dr. Karen Browning, and Dr. Rick Russell, for helpful ideas and scientific advice.

I would like to thank the members of the Stevens laboratory who have supported me in many ways during my graduate study: Champ Gupton, Rea Lardelli, Jennifer Hennigan, Matthew Sorenson, Sanjeev Namjoshi, Rachel Chirayil, and Caroline Davis, for their friendship and help. My thanks also go out to Albert MacKrell and the students in the Vertebrate Interactome Stream of the UT Freshman Research Initiative for help with the screen. I am also grateful to the members of the Johnson laboratory who have always been kind and helpful.

Finally I would like to express special thanks to my husband, my family, and my friends for the love and support.

Exploring intron mobilization and detection of an intron gain event *via* intron transposition using a novel intron gain and loss reporter

by

Sujin Lee, Ph.D.

The University of Texas at Austin, 2015

Supervisor: Scott Stevens

Eukaryotic nuclear genes are discontinuous with the presence of intervening sequences termed spliceosomal introns. Once the DNA coding sequences are transcribed into pre-mRNA, these spliceosomal introns must be removed within the ribonucleoprotein complex called the spliceosome. The processed mRNA is exported from nucleus to cytoplasm where it is translated into protein. Therefore the removal of spliceosomal introns, pre-mRNA splicing, is an essential process for appropriate gene expression in eukaryotes.

Given the importance of pre-mRNA splicing and diversity of intron densities across eukaryotic genomes, numerous studies have been conducted to understand the origin and evolution of spliceosomal introns. Although several models based on the phylogenetic analyses exist which address the molecular mechanism of the intron gain and loss, validation of these models is restricted due to the lack of experimental evidence.

In this dissertation, we report the use of a novel strategy which detects selected intron gain and loss events. Our reporter is designed to produce an intronic RNA containing a selectable marker that detects its incorporation into the yeast genome. We have experimentally verified the first demonstration of intron gain *via* intron transposition in any organism. The intron RNA derived from the reporter was perfectly transposed in the yeast gene *RPL8B* and remains stable and intact within the genome. This novel allele, *RPL8Bint*, is functional when overexpressed in a deletion strain of *RPL8A*, a paralog of *RPL8B*, demonstrating that the newly formed intron is successfully removed by the spliceosome.

To address the mechanism of this intron transposition, we investigated the involvement of the known cellular genes in intron transposition using the intron gain and loss reporter. A number of deletion strains of the spliceosome-related genes and recombination-related genes were employed in addition to the conditional mutants of splicing helicases. The results from these mutational analyses provided evidence to further understand the mechanism of intron mobilization with highlighting the importance of *RAD52* and Ty transposable elements.

Altogether this dissertation describes the development and validation of a novel reporter detecting *in vivo* intron gain and loss and the utilization of the reporter in understanding the mechanism of intron mobilization in *S. cerevisiae*.

Table of Contents

List of tables xiii

List of figures xiv

Chapter 1. Introduction 1

1.1 Eukaryotic gene expression 1

1.2 Pre-mRNA splicing 2

1.2.1 Spliceosome 3

1.2.2 Pre-mRNA splicing mechanism 4

1.2.3 Splicing cycle 6

1.2.3.1 Dynamics in spliceosome assembly, catalysis, and disassembly 7

1.2.3.2 Penta-snRNP model of spliceosomal assembly 11

1.3 Origin and evolution of spliceosomal introns 12

1.3.1 Group II self-splicing introns 13

1.3.2 Establishment of the spliceosomal system 16

1.3.3 Evolution of the spliceosomal introns 18

1.3.3.1 Mechanisms of intron loss 20

1.3.3.2 Mechanisms of intron gain 23

1.4 Dissertation objectives 26

Chapter 2. Detection of the first intron transposition-induced intron gain using a novel reporter system 27

2.1 Background 27

2.1.1 Reversibility of pre-mRNA splicing 28
2.2 Materials and methods 29
2.2.1 Construction of a reporter system to detect intron gain and loss 29
2.2.2 Yeast strains and DNA oligonucleotides 29
2.2.3 Yeast transformation with the reporter plasmid 30
2.2.4 Intron transposition assay 34
2.2.5 Diagnostic PCR to determine genomic integration 34
2.2.6 Confirmation of intron loss 34
2.2.7 Inverse PCR analysis 34
2.2.8 *RPL8Bint* identification and confirmation by PCR 35
2.2.9 TAP tagging of *RPL8Bint* 35
2.2.10 *RPL8Bint* promoter switch 36
2.2.11 Yeast sporulation and tetrad dissection 36
2.2.12 Protein isolation and western blot analysis 36
2.2.13 Total RNA extraction 38
2.2.14 RT-PCR analysis to detect recursive splicing products 38
2.2.15 Serial dilutions and spotting test 38
2.3 Results 39
2.3.1 Construction and validation of the reporter system 39
2.3.2 Detection of recursive splicing products 42
2.3.3 Formation of *his+* prototroph 45

2.3.4 Detection of intron loss events 45
2.3.5 Detection of an intron gain event 47
2.3.6 *RPL8Bint* is capable of splicing and expression 50
2.3.7 Overexpression of *RPL8Bint* restores the expression level 50
2.3.8 Rpl8 expressed from *GPD-RPL8Bint* strain is sufficient enough to overcome deletion of *RPL8A* 51
2.3.9 *GPD-RPL8Bint* strain has no growth defect 51
2.4 Discussion 58
Chapter 3. The involvement of splicing and recombination genes in intron mobilization 64
3.1 Background 64
3.1.1 Spliceosome-related genes: *DBR1*, *SNU66*, and *BUD13* 64
3.1.2 Homologous recombination and repair genes: *RAD52* epistasis group and *RAD1* 65
3.1.3 RNA helicase genes: *PRP16*, *PRP22*, and *PRP43* 67
3.2 Materials and methods 71
3.2.1 Plasmids 71
3.2.2 Yeast strains 71
3.2.3 Construction of strains 71
3.2.3.1 *prp16 and prp22* cold-sensitive mutant strains 71
3.2.3.2 *prp43* cold-sensitive mutant strains 72
3.2.3.3 *prp43* cold-sensitive diploid strain 72

3.2.3.4 Wild-type diploid strain 73
3.2.4 DNA oligonucleotides 76
3.2.5 Yeast transformation with the reporter plasmid 78
3.2.6 Intron transposition assay 78
3.2.7 Rate determination of *his*+ prototroph formation 78
3.2.8 Diagnostic PCR and inverse PCR 78
3.3 Results 79
3.3.1 Formation of *his*+ prototroph is rare in *dbr1*Δ 79
3.3.2 *his*+ prototroph formation increases in splicing defective mutants, *bud13*Δ and *snu66*Δ 79
3.3.3 The effect of defective recombination on *his*+ prototroph formation 80
3.3.4 The effect of splicing helicase mutants in *his*+ prototroph formation 84
3.3.5 Formation of *his*+ is inhibited in the diploid strains 87
3.3.6 Chromosomal events resulted in the pseudogene formation, not intron gain 87
3.4 Discussion 91
Chapter 4. Significance and future directions 94
4.1 Significance 94
4.2 Future directions 95
4.2.1 Involvement of Ty transposable element 95

4.2.2 Connection between *RAD52* and Ty transposable element . 96

4.2.3 Reversibility of pre-mRNA splicing 97

References 98

VITA 116

List of Tables

Table 2.1 Yeast strains used in Chapter 2 31

Table 2.2 DNA oligonucleotides used in Chapter 2..... 32

Table 3.1 Yeast strains used in Chapter 3 74

Table 3.2 DNA Oligonucleotides used in Chapter 3..... 76

Table 3.3 *his*+ prototroph formation in *dbr1*Δ..... 81

Table 3.4 *his*+ prototroph formation in the splicing defective mutants 81

Table 3.5 *his*+ prototroph formation in the recombination defective mutants..... 82

Table 3.6 *his*+ prototroph formation in the splicing helicase mutants..... 85

Table 3.7 *his*+ prototroph formation in the diploid strains 90

List of Figures

Figure 1.1 Pre-mRNA splicing mechanism 5

Figure 1.2 Splicing cycle 8

Figure 1.3 Group II self-splicing mechanism 15

Figure 1.4 Reverse transcriptase-mediated intron loss 22

Figure 1.5 Intron transposition involving reverse splicing 25

Figure 2.1 Design and validation of the intron gain and loss reporter 41

Figure 2.2 Detection of recursive intron splicing 43

Figure 2.3 Plasmid-borne intron loss events 46

Figure 2.4 Inverse PCR analysis 48

Figure 2.5 Transposition of the eGFP intron into the *RPL8B* locus 49

Figure 2.6 *RPL8Bint* expression 52

Figure 2.7 *RPL8Bint* expression with the GPD promoter 53

Figure 2.8 PCR confirmation of *rpl8a*Δ */ GPD-RPL8Bint* strain 54

Figure 2.9 Quantitative Western blot 55

Figure 2.10 Overexpressed *RPL8Bint* allele is functionally comparable to the wild -type strain 56

Figure 2.11 Microhomology-mediated end joning rules preclude recombination as a means of this intron acquisition 61

Figure 2.12 Predicted base-pairing interactions between the *RPL8B* transcript and the U5 snRNA at the new splice site 62

Figure 2.13 Proposed model for the mechanism of *RPL8B* intron gain 63

Figure 3.1 Proposed model of *trans*-reverse splicing facilitated by a mutant prp43 70

Figure 3.2 Rate of *his*+ prototroph formation in the wild-type, the splicing defective mutants, and the recombination defective mutants 83

Figure 3.3 Rate of *his*+ prototroph formation in the splicing helicase mutants 85

Figure 3.4 Proposed model of *cis*-reverse splicing facilitated by a mutant prp22 86

Figure 3.5 Proposed model of the plasmid-related events leading to the pseudogene additions 89

Figure 3.6 Chromosomal integration of the reporter intron resulted in pseudogene formation 90

Chapter 1. Introduction

1.1 Eukaryotic gene expression

The genomes of all living organisms encode gene products which determine cell function. During gene expression, each step in the flow of genetic information, DNA to RNA to proteins, needs to be precisely regulated to produce RNAs and proteins in the correct amounts. While the basic features of gene expression are commonly shared between prokaryotes and eukaryotes, gene expression is achieved in a much more intricate manner in eukaryotes. One big difference between prokaryotes and eukaryotes is that eukaryotic transcription and translation take place in different cellular compartments, the nucleus and cytoplasm respectively, while these processes are more closely coupled in prokaryotes. The presence of a membrane-enclosed nucleus in eukaryotes provides an additional level of complexity. Another major difference is that all eukaryotes contain introns which need to be removed from the messenger RNA (mRNA) for accurate gene expression. This feature requires another layer of complexity in eukaryotic RNA processing.

Gene expression begins with transcription of mRNA from the DNA template in the nucleus (Lee and Young, 2000). While prokaryotes have one RNA polymerase, eukaryotes have three kinds of RNA polymerases, RNA polymerase I, II and III, and RNA polymerase II catalyzes mRNA transcription. When RNA polymerase II initiates transcription, it creates a large DNA-protein complex with

transcription factors on the promoter sequence of the DNA template for accurate transcription. Pre-mRNAs being produced by RNA polymerase II undergo extensive processing including 5' cap addition, pre-mRNA splicing, and 3' polyadenylation. Capping is a process that adds an inverted 7-methylguanosine to the first transcribed nucleotide by triphosphate linkage to stabilize the mRNA and promote transcription, downstream processing and nuclear export (Rasmussen and Lis, 1993). Eukaryotic transcripts interrupted by non-coding sequences called introns are further processed within the spliceosome to remove the introns (Matera and Wang, 2014). Polyadenylation catalyzes the addition of a poly(A) tail to the 3' end of the cleaved transcript which enhances mRNA stability and promotes nuclear export of the mRNA to the cytoplasm and translational efficiency (Colgan and Manley, 1997). Mature mRNAs are then exported through the nuclear pore complex to the cytoplasm where they are translated into proteins by a large ribonucleoprotein complex termed the ribosome. Gene expression is also subject to post-translational regulation including both RNA and protein modifications (Steitz, 2008).

Together, the spatial and temporal separation of transcription and translation and the presence of introns allow eukaryotes to regulate gene expression at several different levels, contributing to expanding eukaryotic function.

1.2 Pre-mRNA splicing

One of the defining features of all eukaryotic organisms is the presence of

intervening sequences termed introns in at least some nuclear genes. The removal of introns from eukaryotic pre-mRNA is essential for accurate gene expression and dependent upon a ribonucleoprotein complex called the spliceosome leading to ligated coding sequences. This following introduction will detail the splicing machinery, the mechanism of pre-mRNA splicing and the dynamics of spliceosome assembly, catalysis, and disassembly.

1.2.1 Spliceosome

Spliceosome is a large RNA-protein complex, which is composed of five small nuclear ribonucleoprotein complexes (U1, U2, U4, U5 and U6 snRNPs) and over one hundred proteins which are required for accurate splicing reaction (Jurica and Moore, 2003). The spliceosome is highly conserved from yeast to human and almost 85% of splicing factors in yeast have been shown to have a human ortholog (Cordin and Beggs, 2013). Metazoans contain additional *cis*-acting splicing elements and *trans*-acting splicing factors due to numerous alternative splicing mechanisms. All eukaryotes contain spliceosomes which catalyze the U2-type intron removal and the yeast spliceosome is thought to encode what is considered a basic, core spliceosome. In yeast, the spliceosome was shown biochemically to be contained in a 40S complex (Brody and Abelson, 1985).

More recently, several metazoan species have been shown to have a second spliceosome termed the minor spliceosome (Tarn and Steitz, 1996a, b). Although both spliceosomes are analogous in terms of their functions, they are

different in their composition and target introns. Only the U5 snRNP is shared by these two types of spliceosomes. U1, U2, U4, and U6 snRNPs of the major spliceosome are replaced by U11, U12, U4atac, and U6atac snRNPs, respectively in the minor spliceosome (Turunen et al., 2013). Similar to U4/U6•U5 tri-snRNP, U4atac, U6atac, and U5 snRNPs are also associated into a tri-snRNP (Turunen et al., 2013). As my work in this dissertation was performed in *Saccharomyces cerevisiae*, the following introduction will focus on the U2-type spliceosome present in yeast.

1.2.2 Pre-mRNA splicing mechanism

Spliceosomal introns are characterized by short sequences at the 5' splice site, branchpoint site, and 3' splice site. These sequences are highly conserved in budding yeast while in metazoans they are more variable with additional available splicing elements and factors.

Extensive studies on the splicing mechanism showed that the splicing reaction occurs by means of two sequential transesterification reactions, ligating the exons and releasing the excised intron (Fig. 1.1) (Domdey et al., 1984; Grabowski et al., 1984; Lin et al., 1985; Padgett et al., 1984; Ruskin et al., 1984). In the first catalytic splicing reaction, the 2' hydroxyl group of the branchpoint adenosine residue (UACUA**A**C in budding yeast) nucleophilically attacks the phosphodiester bond of 5' splice site (GU), yielding a free 5' exon and a lariat-structured intron-3' exon intermediate. The branchpoint adenosine is linked to the

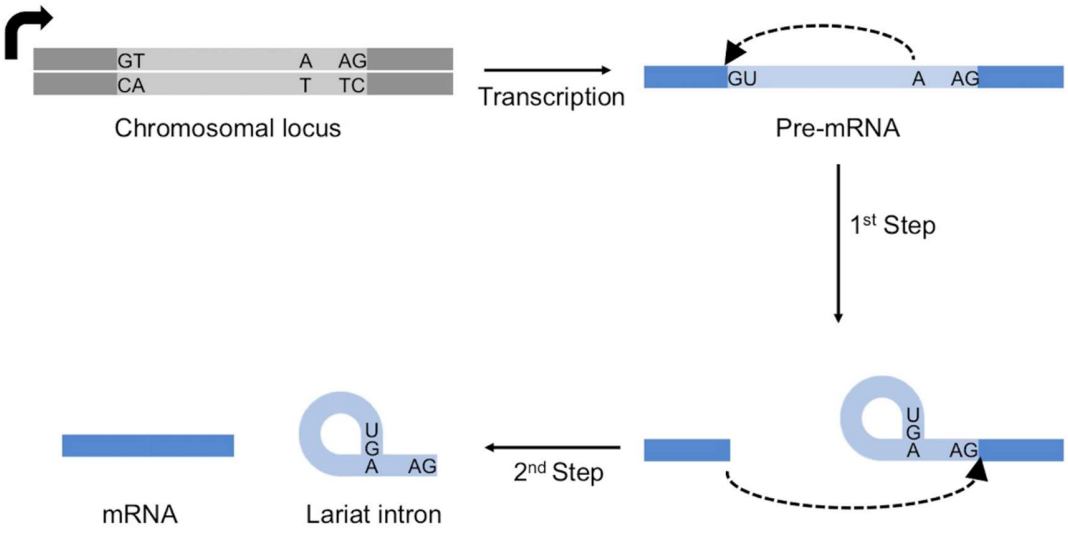

Figure 1.1 Pre-mRNA splicing mechanism

Pre-mRNA splicing occurs by two ordered transesterification steps. In the first step, the 2'-OH of the branchpoint adenosine performs a nucleophilic attack at the 5' splice site, yielding a lariat intermediate RNA and a 5' exon. The second step takes place when the 3'-OH of the 5' exon attacks the 3' splice site, ligating the exons and releasing the lariat intron.

guanosine from the 5' splice site by an unusual 2'-5'-phosphodiester bond. These products are repositioned as the substrates for the second reaction since the spliceosome has one active site for transesterification reactions. The second catalytic step of the splicing reaction subsequently occurs when the 3' splice site (AG) is nucleophilically attacked by the 3' hydroxyl group of 5' exon. This ligates the 5' and 3' exons together and produces the excised lariat-shaped intron RNA.

1.2.3 Splicing cycle

The spliceosome is assembled on the substrate pre-mRNA in an ordered process. Five snRNPs and non-snRNP splicing factors interact with each other and with changes in RNA structures and also association and release of splicing factors throughout the splicing cycle. Once the splicing reaction is completed, the spliceosome is disassembled for another cycle of splicing. The snRNPs are recycled and the excised intron is degraded after debranching the 2'-5' branchpoint linkage.

In the step-wise model of spliceosome assembly, U1 and U2 snRNPs and pre-formed U4/U6•U5 tri-snRNP are sequentially added to the pre-mRNA by forming base-paring between the splice sites of pre-mRNA and snRNAs (Matera and Wang, 2014). In an alternative penta-snRNP model, all five snRNPs assemble as a holoenzyme first prior to associating with the pre-mRNA (Stevens et al., 2002).

1.2.3.1 Dynamics in spliceosome assembly, catalysis, and disassembly

The U1, U2, U4/U6, and U5 snRNPs are the main building blocks of the spliceosome, each of which is composed of a snRNA and many protein factors. These snRNPs do not have a static active center and are subject to remodeling during the course of splicing reaction. Stepwise spliceosome assembly begins with the ATP-independent binding of the U1 snRNP on the pre-mRNA by base-pairing between 5' end of U1 snRNP and 5' splice site of the intron (Fig. 1.2) (Seraphin et al., 1988; Seraphin and Rosbash, 1989). In higher eukaryotes, this interaction is stabilized by the members of serine-arginine-rich (SR) protein family and U1 snRNP protein factors. Additionally, the branchpoint site and the polypyrimidine tract just downstream of the branchpoint site are recognized by Msl5 and Mud2. Together, these interactions form the commitment complex (complex E) and have crucial roles in the initial recognition of the 5' splice site and 3' splice site of intron.

Subsequently, U2 snRNP is recruited in ATP-dependent manner by base-pairing between U2 snRNA and the branchpoint site of the pre-mRNA, leading to the formation of complex A (Legrain et al., 1988). Msl5 and Mud2 are displaced by U2 snRNP binding. Two RNA helicases, Prp5 and Sub2, are required at the time of U2 snRNP binding to the branchpoint site. The requirement for ATP and Prp5 suggests their possible roles in inducing a conformational change in the U2 snRNP to expose the U2 snRNP branchpoint recognition region (Dalbadie-McFarland and Abelson, 1990; O'Day et al., 1996; Ruby et al., 1993). ATP-bound Sub2 is

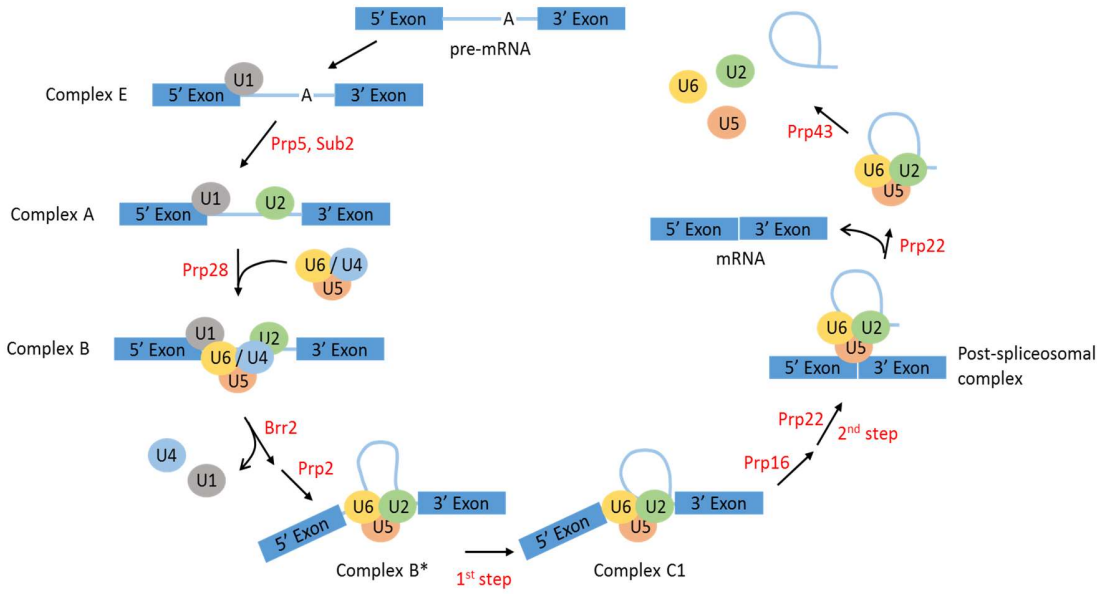

Figure 1.2 Splicing cycle

The spliceosome is assembled on the substrate pre-mRNA by stepwise binding of snRNPs. The U1 snRNP binds to the 5' splice site and the U2 snRNP binds to the branchpoint site. Subsequently, the U4/U6•U5 tri-snRNP complex joins to form complex B. By the action of Brr2 and Prp2, the spliceosome is catalytically activated (complex B*). The first catalytic step results in the formation of free 5' exon and lariat intermediate. Next, the spliceosome undergoes Prp16-induced conformational change and proceeds the second catalytic step, yielding ligated exon and the lariat intron. Disassembly of the spliceosome is catalyzed by Prp22 and Prp43.

proposed to be recruited to the branchpoint site by Msl5 and Mud2 and then ATP hydrolysis by Sub2 releases Msl5, allowing U2 snRNP to access to the branchpoint site (Kistler and Guthrie, 2001; Shen et al., 2008). The base-pairing interaction between the U2 snRNA and the branchpoint site is stabilized by the sequence-independent binding of SF3a and SF3b subunits of U2 snRNP to the upstream of the branchpoint site (Gozani et al., 1996) The SF3a subunit consists of Prp9, Prp11 and Prp21 (Brosi et al., 1993; Ruby et al., 1993) and the SF3b subunit is composed of six factors, Cus1p, Hsh49p, Hsh155p, Rse1p, Ysf3p, and Rds1p (Wang et al., 2005).

Complex A transitions into complex B by addition of the U4/U6•U5 snRNPs in the form of pre-formed tri-snRNP. In the tri-snRNP, U4 and U6 snRNAs are extensively base-paired through two regions of sequence complementarity. Catalytic activation of spliceosome requires several rearrangements (Makarov et al., 2002) including the displacement of U1 and U4 snRNPs. The U1 snRNA is unwound from 5' splice site by the ATP-dependent action of Prp28 (Staley and Guthrie, 1999). Also, U4 snRNA is dissociated from U6 snRNA by the action of Brr2 in an ATP-dependent manner (Raghunathan and Guthrie, 1998). Together, these rearrangements switch the base-pairing interaction between the U1 snRNA and the 5' splice site to an exclusive interaction between 5' splice site and U6 snRNA (Staley and Guthrie, 1999). At this time, a heteromeric Prp19-complex (NTC) also associates the spliceosome, leading to the formation of complex B^{act} by stabilizing the interaction between the U6 snRNA and 5' splice site

(Chan and Cheng, 2005). This complex is almost complete but yet catalytically inactive.

Formation of catalytically active complex B* is achieved by the RNA helicase Prp2 (Chen and Lin, 1990; Kim et al., 1992; King and Beggs, 1990). Prp2 induces a conformational change within the complex B^{act} in an ATP-independent manner and also displaces SF3 subunits of U2 snRNP in an ATP-dependent manner, exposing the branchpoint adenosine (Lardelli et al., 2010). These rearrangements allow the first transesterification reaction to proceed in which the 2’ hydroxyl group from the branchpoint adenosine nucleophilically attacks the 5’ splice site, forming the complex C1 containing the 5' free exon and the lariat intron-3' exon intermediate.

For the second transesterification reaction to occur, the RNA helicase, Prp16, induces additional conformational changes within the spliceosome to place the 5' splice site and the 3' splice site in close proximity (Schwer and Guthrie, 1991, 1992). The transition from the complex C1 to complex C2 involves the destabilization of the interaction between the U6 snRNA and the 5' splice site and the actions of the splicing factors such as Prp18, Slu7, and Prp17 to promote the second catalytic splicing step (James et al., 2002; Sapra et al., 2008). The RNA helicase Prp22 is also involved in the second catalytic step in an ATP-independent manner (Schwer and Gross, 1998).

Now the second transesterification reaction proceeds by the 5' exon attacking the 3' splice site, ligating the exons and releasing the excised lariat-

structured intron. The ligated exons are released from the spliceosome by the ATP-dependent action of Prp22 (Schwer and Gross, 1998; Wagner et al., 1998).

Finally, the RNA helicase Prp43 disassembles the post-splicing spliceosome to recycle the snRNPs for another round of splicing (Martin et al., 2002) and the released intron is debranched by a debranching enzyme, Dbr1, and eventually degraded (Chapman and Boeke, 1991).

1.2.3.2 Penta-snRNP model of spliceosomal assembly

The penta-snRNP model of spliceosome assembly suggests that the snRNPs are also able to pre-assemble prior to associating with the substrate pre-mRNA. In 1988, Konarska and Sharp showed that U2, U4, U5, and U6 snRNPs form a stable interaction in the absence of the precursor RNA using HeLa nuclear extracts. The formation of tetra-snRNP supported the model that the structure of the spliceosome is basically defined by snRNP-snRNP interactions (Konarska and Sharp, 1988). The tetra-snRNP was also identified by Stevens and Abelson in their characterization of a 25S tri-snRNP (Stevens and Abelson, 1999). Under the salt concentration of 150 mM, U2•U4/U6•U5 tetra-snRNP was observed and when the salt concentration was lowered to 50 mM, U1•U2•U4/U6•U5 penta-snRNP was isolated (Stevens et al., 2002). Given that the pre-mRNA splicing in yeast extracts occurs at 50-120 mM monovalent salt condition (Lin et al., 1985), the observation of the penta-snRNP and the tetra-snRNP in the absence of the pre-mRNA proposed an alternative way of the spliceosome assembly in which either the all

five snRNPs are pre-assembled prior to interacting with the pre-mRNA or the tetra-snRNP is recruited to the U1 snRNP initially associated with the pre-mRNA (Stevens et al., 2002). This model was supported by that isolated penta-snRNP includes almost all known yeast splicing factors and also exhibits splicing activity *in vitro* (Stevens et al., 2002).

1.3 Origin and evolution of spliceosomal introns

While extensive studies have been performed to define the mechanism of pre-mRNA splicing and address the splicing machinery since the finding of spliceosomal introns (Matera and Wang, 2014), several evolutionary mysteries remain regarding introns: how did spliceosomal introns invade and persist in eukaryotic genomes? How are they removed from the genomes of organisms undergoing intron loss? Have introns been added over evolutionary time and if so, how does that occur?

The origin of spliceosomal introns has been a matter of debating for nearly 30 years as presented in two opposite hypotheses (Irimia and Roy, 2014). The introns-first hypothesis supports that the spliceosomal introns are extremely ancient and existed even at the earliest stages of evolution (Darnell, 1978; Gilbert, 1987). The introns-late hypothesis holds that the spliceosomal introns arose with eukaryogenesis and have persisted in eukaryotic genomes across the evolution of eukaryotes (Cavalier-Smith, 1985, 1991; Dibb and Newman, 1989; Logsdon et al., 1995). Though both scenarios are competing with the scarcity of direct data, the

introns-late hypothesis is more appealing given that there is no evidence of any prokaryote possessing a spliceosomal intron or any of the spliceosome components.

Also, there is accumulating evidence that spliceosomal introns are evolutionarily related to the group II self-splicing introns and the last eukaryotic common ancestor (LECA) had an intron-rich genome and a complex spliceosome (Collins and Penny, 2005), suggesting that the spliceosomal introns would have evolved from the group II self-splicing introns at the time of eukaryogenesis (Irimia and Roy, 2014; Koonin, 2006; Rogozin et al., 2012). The following introduction will describe the current knowledge on the origin and evolution of the spliceosomal introns and the spliceosome system.

1.3.1 Group II self-splicing introns

Group II self-splicing introns are predicted to predate the eukaryogenesis and may even have preceded the origin of cellular life (Koonin, 2006). They have been shown to be evolutionarily related to the spliceosomal introns, the spliceosome, and the retrotransposons in eukaryotes (Lambowitz and Zimmerly, 2011).

Group II introns are mobile genetic elements found in bacteria and organellar genomes such as mitochondrial and chloroplast genomes of some eukaryotes including fungi and plants (Lambowitz and Zimmerly, 2011). Interestingly, group II introns have not been found in eukaryotic nuclear genomes.

The evolutionary relationship between group II introns and spliceosomal introns has been suggested and seems plausible considering the analogous mechanism of splicing and also numerous structural similarities (Cavalier-Smith, 1991; Cech, 1986; Rogers, 1989; Sharp, 1991).

Group II introns consist of an intron RNA and often, an intron-encoded protein (IEP) (Lambowitz and Zimmerly, 2011). As a ribozyme, group II intron RNA catalyzes self-splicing through the identical mechanism by which spliceosome catalyzes pre-mRNA splicing (Fig. 1.3). Also group II intron RNAs have conserved 5' and 3' junction sequences, GUGYG and AY, similar to the sequences conserved at 5' and 3' splice sites in spliceosomal introns. Group II intron RNAs have a conserved secondary structure composed of six domains, DI-DVI, forming a distinct tertiary structure to form an active site. Among these domains, DIV loop encodes the IEP which is a multifunctional reverse transcriptase (RT). In addition to synthesizing a cDNA copy of the intron RNA using its RT activity, the IEP also plays a role in stabilizing catalytically active intron RNA structure for splicing. After the group II splicing reaction, the IEP stays bound to the lariat intron RNA for DNA invasion which involves reverse splicing and subsequent reverse transcription. Since transesterification reactions are reversible, the excised intron RNA can be directly reverse spliced into a DNA strand using its own ribozyme activity and then the RT activity of IEP catalyzes reverse transcription of intron RNA into cDNA. By a series of splicing, reverse splicing, and reverse transcription, group II introns are capable of being propagated within genomes.

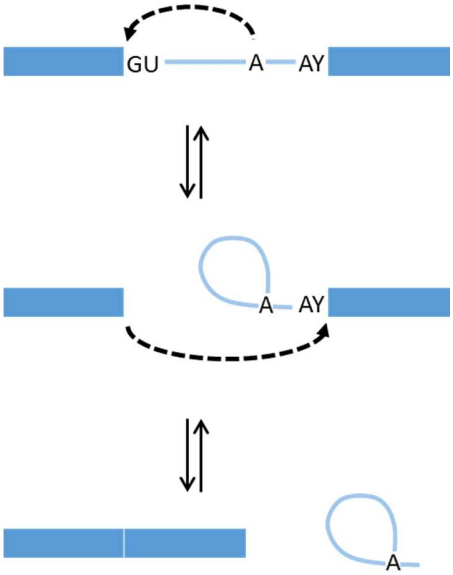

Figure 1.3 Group II self-splicing mechanism

Group II introns catalyze self-splicing *via* two sequential transesterification reactions. In the first step, the 2′ OH of the branchpoint adenosine in DVI nucleophilically attacks the 5′-splice site, yielding a lariat intron and 3′-exon intermediate. In the second step, the 3′ OH of the 5′ exon attacks the 3′-splice site, completing exon ligation and excision of a lariat intron.

There are also structural similarities between group II introns and spliceosomal RNAs (Lambowitz and Zimmerly, 2011). U5 snRNA stem loop in recognition of 5' and 3' exons, U2 snRNA base-paired with branchpoint region of the spliceosomal intron, and U6 snRNA base-pairing with U2 snRNA correspond to the DI-exon interactions, DVI, and the active site helix DV of group II intron RNAs respectively.

In addition to the mechanistic and structural similarities, there have been recent reports that extracts of snRNAs can catalyze both splicing reactions in the absence of proteins *in vitro* indicating the *in vivo* catalytic potential of the snRNAs similar to the group II intron RNAs (Dayie and Padgett, 2008; Valadkhan and Manley, 2009; Valadkhan et al., 2007).

1.3.2 Establishment of the spliceosome system

Given the features that the group II self-splicing introns and the spliceosomal system share, it is widely accepted that the spliceosomal introns originated from the group II self-splicing introns and appeared at the time of eukaryogenesis (Irimia and Roy, 2014; Rogozin et al., 2012). And this transition is expected to have occurred before the appearance of LECA as the spliceosomal introns are present in all extant eukaryotic supergroups. The most favored hypothetical scenario is that the α-proteobacterial ancestor of mitochondria possessing the group II introns were taken inside of an archaea-like eukaryotic ancestor and subsequently the group II introns were transferred to the host genome (Cavalier-Smith, 1991; Irimia and Roy, 2014; Rogozin et al., 2012; Woese

et al., 1990). After the intron transfer event, it is likely that the proto-spliceosomal introns highly resembled the group II introns and that for some period of time retained their mobility and self-splicing ability considering that LECA is very likely to have had an intron-rich genome (Irimia and Roy, 2014; Koonin et al., 2013; Lambowitz and Zimmerly, 2011). This may have enabled and promoted considerable proliferation of these introns at the time of eukaryogenesis (Irimia and Roy, 2014; Rogozin et al., 2012). The group II intron-derived features including self-splicing activity and mobility are likely to have degenerated over time, leading to a necessity for the development of the pre-mRNA splicing machinery composed of *trans*-acting snRNAs and *trans*-acting protein factors with the increasing number of introns.

There are recent studies to support the hypothesis that the LECA had a spliceosomal system nearly as complex as the modern spliceosome. Comparative analyses of the differences between spliceosomal proteins from many basal eukaryotic lineages were performed to infer properties of the ancient spliceosomal system. Distribution of spliceosomal components between different eukaryotic lineages revealed that the spliceosome in the LECA must have contained most of the key factors present in contemporary eukaryotes including snRNPs-specific proteins and also many splicing-related proteins (Anantharaman et al., 2002; Collins and Penny, 2005). This ancestral spliceosomal system may have evolved in a way that adds more complexity for higher efficiency fulfilled by the modern

spliceosome, which is one of the largest and the most complex macromolecular machines in the cell.

1.3.3 Evolution of the spliceosomal introns

Comparative analyses of the intron positions have shown that the homologous genes from different intron-rich eukaryotic genomes have significant coincidence in their intron positions, suggesting that their ancestors already contained the corresponding introns (Fedorov et al., 2002; Rogozin et al., 2003). A recent statistical study using 99 eukaryotic genomes reconstructed the history of intron gain and loss throughout the evolution of eukaryotes and inferred that the ancestors of each major eukaryotic group were intron-rich with 53-74% of the modern human intron density (Csuros et al., 2011). It is consistent with the previous studies suggesting that the genome of LECA was intron-rich (Csuros et al., 2008; Rogozin et al., 2003; Roy, 2006; Roy and Gilbert, 2005a), indicating that there was at least one massive intron proliferation event since the group II intron transfer to the genome of the eukaryotic ancestor.

The intron densities of the modern eukaryotic genomes are notably different: 8.1 introns per gene in humans, 4.4 introns per gene in *Arabidopsis thaliana*, 3.4 introns per gene in *Drosophila melanogaster*, 0.05 introns per gene in *Saccharomyces cerevisiae* (Consortium, 2004; Drysdale and Crosby, 2005; Haas et al., 2005; Hirschman et al., 2006). Even unicellular organisms show great variation in intron densities from few to high numbers comparable to the intron-rich

multicellular lineages, indicating that the intron density does not necessarily reflect biological complexity (Collen et al., 2013; Irimia and Roy, 2014; Merchant et al., 2007). Given that the LECA was intron-rich with 53-74% of the modern human intron density, 4.4-6 introns per gene, the evolutionary processes clearly appear to have experienced extensive lineage-specific intron loss and intron gain (Csuros et al., 2011; Rogozin et al., 2003) although the mechanisms by which spliceosomal introns are obtained and lost largely remain elusive. The currently proposed models for the mechanisms of intron gain and loss will be further discussed in 1.3.3.1 and 1.3.3.2.

It is also notable that the size of spliceosomal introns greatly differ between eukaryotic genomes. Taking the hypothesis that the spliceosomal introns originated from the group II introns, the earliest form of spliceosomal introns are expected to be similar in the length to the group II introns, which span 400-800 nucleotides long (Irimia and Roy, 2014; Lambowitz and Zimmerly, 2011). Most eukaryotic genomes investigated so far contain approximately 150 nucleotide-long spliceosomal introns on average but the distribution of intron length across eukaryotes shows the extremely wide range from the smallest recorded introns, 18-21 nucleotides long, found in the chlorarachniophyte *Bigelowiella natans* nucleomorph (Gilson et al., 2006) to some introns reaching one megabase (Irimia and Roy, 2014). The intron size may have been deviated during eukaryotic evolution to optimize gene expression since the length of intron affects the splicing efficiency. The appearance of large introns is likely due to obtaining transposable

elements or *cis*-acting regulatory elements within the intron sequence as well as limited selective pressure (Irimia et al., 2011; Irimia and Roy, 2014; Jiang and Goertzen, 2011).

With few exceptions (Lane et al., 2007), the spliceosomal introns have been maintained within eukaryotic genomes throughout the evolution, suggesting their functional importance in regulating gene expression. For instance, in budding yeast with very low intron density, the majority of introns exist in the ribosomal protein-coding mRNAs that are highly translated, regulating the biogenesis and functions of ribosomes. Deletion of the introns from the ribosomal protein-coding genes caused reduced fitness and growth under stress conditions, highlighting their functional importance (Parenteau et al., 2011; Rogozin et al., 2012).

1.3.3.1 Mechanisms of intron loss

Given that the LECA was intron-dense, the modern intron-sparse eukaryotes have involved dominant intron loss over intron gain. For instance, comparative analysis showed that *Saccharomyces cerevisiae* contains only 14% of the introns shared between animals and *Arabidopsis thaliana* (Roy and Gilbert, 2006), suggesting that budding yeast is likely to have undergone widespread intron loss during its evolution.

Several models exist for how introns might be lost (Boeke et al., 1985; Fink, 1987; Mourier and Jeffares, 2003) and experimental intron loss has been demonstrated in at least one organism (Derr, 1998; Derr et al., 1991). Two

definitive molecular mechanisms are reverse transcriptase-mediated intron loss and genomic deletion (Roy and Gilbert, 2006; Yenerall and Zhou, 2012).

In the mechanism of reverse transcriptase-mediated intron loss, the spliced transcript of an intron-containing gene is reverse transcribed into the cDNA copy which no longer contains the intronic sequence. Homologous recombination between the cDNA copy and the chromosomal allele deletes the intron from the genomic locus (Fig. 1.4) (Derr, 1998; Derr and Strathern, 1993; Derr et al., 1991). This mechanism was demonstrated *in vivo* in *Saccharomyces cerevisiae*. This model is likely to be correct, as it reflects the genomic reality that budding yeast introns generally exist close to the 5’ end of intron-containing genes as would be expected in a reaction mediated by reverse transcriptase which begins copying the mRNA from the 3’ end (Mourier and Jeffares, 2003; Roy and Gilbert, 2005b; Sakurai et al., 2002; Sverdlov et al., 2004). The operative reverse transcriptase is presumably from active retrotransposons (Baltimore, 1985; Derr and Strathern, 1993).

The second mechanism of intron loss, genomic deletions, suggests that introns are lost by simple in-frame genomic deletions of intron sequences (Cho et al., 2004; Kent and Zahler, 2000). There is growing evidence that non-homologous recombination relying on the microhomology between the 5' and 3' splice sites (consensus motif AG|GT of the 5' and 3' splice sites) leads to the precise loss of intron (Robertson, 1998). More recently proposed models of genomic deletions are

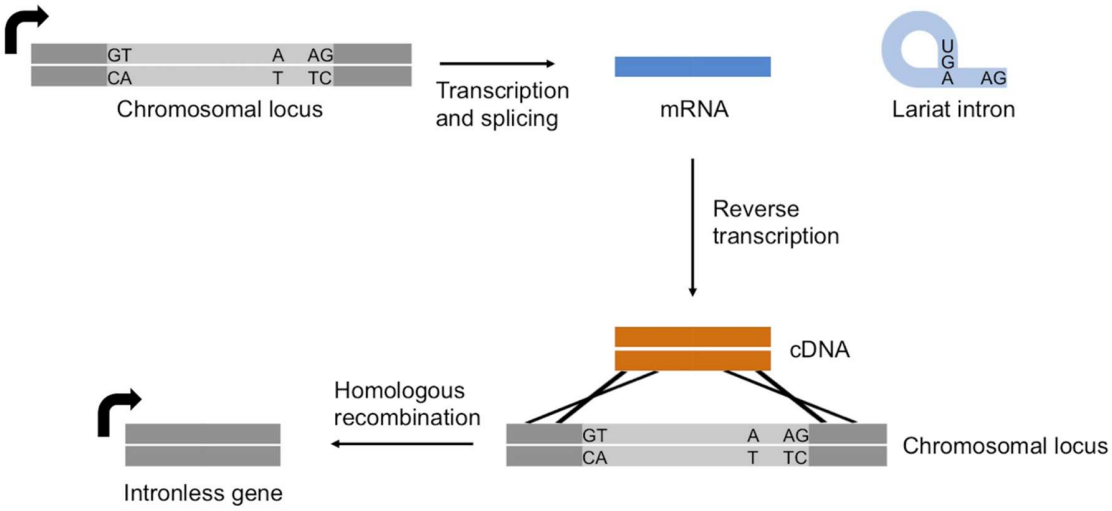

Figure 1.4 Reverse transcriptase-mediated intron loss

After an intron-containing gene is transcribed and spliced, the resulting mRNA is reverse transcribed to the cDNA by cellular reverse transcriptase. The cDNA is then incorporated into the genome by homologous recombination, leading to the generation of a new, intronless allele.

implicated with DNA double-strand break (DSB) and non-homologous end joining repair (Farlow et al., 2011; Hu, 2006). In these models, the DSB occurred within intron sequence is repaired by microhomology pairing between 5' and 3' splice sites, leading to the precise removal of the intron (Farlow et al., 2011). The genomic deletions may not remove the entire intron sequence but still result in intron loss by deleting splice signals or limiting physical space to load the spliceosome (Robertson, 1998).

1.3.3.2 Mechanisms of intron gain

It remains unknown and debatable how spliceosomal introns might be gained. Models of intron gain are numerous and all of these models could potentially lead to intronogenesis, none have yet been experimentally validated (Roy and Irimia, 2009; Yenerall and Zhou, 2012). The most favored models will be reviewed here.

Intron transposition involving reverse splicing is the most parsimonious and favored model. In this model, a spliced intron RNA retained in the residual spliceosome is incorporated into an intron-naïve mRNA which has encountered and stably interacted with this species by reverse splicing (Roy and Irimia, 2009; Sharp, 1985; Yenerall and Zhou, 2012). Subsequently the mRNA containing the intron is reverse transcribed into a cDNA copy by a cellular reverse transcriptase. Homologous recombination between the cDNA copy and the chromosomal allele that does not have intron leads to the formation of newly inserted intron (Fig. 1.5). This model is the reversal of reverse transcriptase-mediated intron loss mechanism except for the additional process of reverse splicing (Roy and Irimia, 2009). Reverse splicing was shown *in vitro* using a mutant RNA helicase, Prp22, with a defect in mRNA release from the spliceosome under non-physiological

condition (Tseng and Cheng, 2008). Despite this *in vitro* demonstration, the extremely rare nature of reverse splicing restricts the experimental validation of intron transposition model (Roy and Irimia, 2009). However, the presence of highly similar introns in the genome and their intraspecific presence-absence of polymorphisms support the intron transposition model (Torriani et al., 2011). The intron transfer model proposes that a newly gained intron is transferred to a paralogous intronless gene by homologous recombination (Yenerall and Zhou, 2012). Intron transfer is also indirectly supported by evidence that introns with similar sequences are found (Torriani et al., 2011) and also that intron gain appears more frequently in the paralogous genes (Babenko et al., 2004; Castillo-Davis et al., 2004).

Intron gain *via* double-strand break repair model hypothesizes that non-homologous end joining between the overhangs generated from double-strand break and an exogenous DNA molecule leads to intron gain (Farlow et al., 2011; Li et al., 2009; Yenerall and Zhou, 2012). The presence of short direct repeats flanking the introns suggests that these introns may result from the double-strand break and non-homologous end joining repair (Farlow et al., 2011; Lieber, 2010).

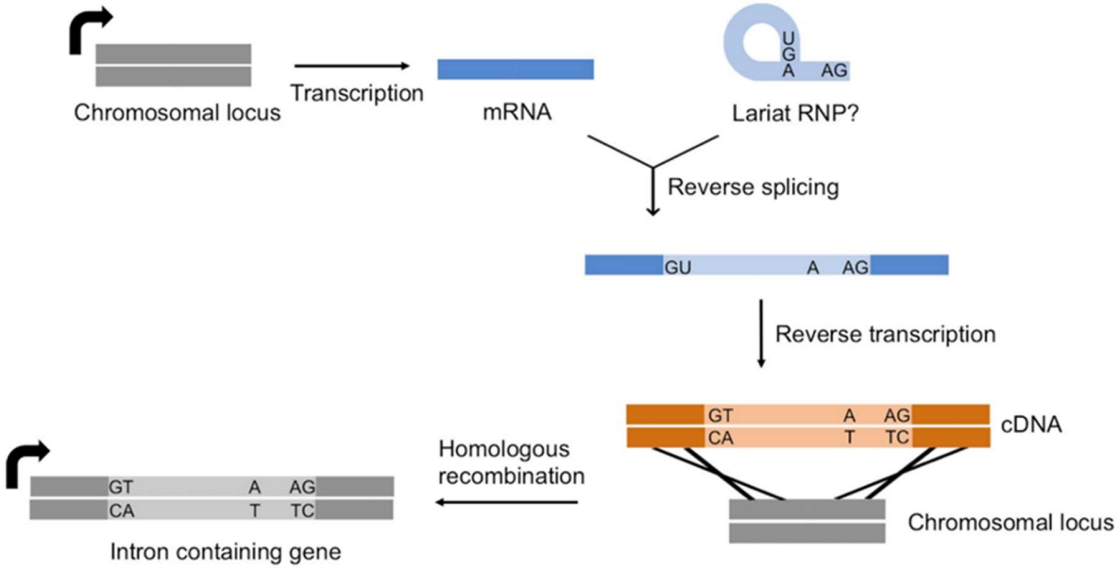

Figure 1.5 Intron transposition involving reverse splicing

The model involving reverse splicing employs a spliced-out lariat RNA as the source of a new intron gain. This lariat RNA remains engaged in the residual spliceosome after mRNA release. Reverse splicing and subsequent reverse transcription generate the cDNA copy containing the newly inserted intron. Homologous recombination between the genome and cDNA copy yields a new intron gain allele.

In addition to the models reviewed above, there are alternative models including tandem genomic duplication, insertion of transposable element, insertion of a group II intron, and intronization (Chalamcharla et al., 2010; Irimia et al., 2008; Rogers, 1989; Sharp, 1985; Yenerall and Zhou, 2012). Although these models also have indirect evidence to support them, they are less likely or solely depend on random mutations (Yenerall and Zhou, 2012).

1.4 Dissertation objectives

Despite extensive studies on spliceosomal intron gain and loss, the molecular mechanisms remain largely unknown with no direct experimental evidence. In this dissertation, the second chapter describes how a novel reporter system to detect intron gain and loss events in *Saccharomyces cerevisiae* was developed and most importantly, reports the first experimental demonstration of *in vivo* intron gain *via* intron transposition mechanism. The third chapter describes how known cellular genes related to pre-mRNA splicing and recombination affect intron mobilization.

Chapter 2. Detection of the first intron transposition-induced intron gain using a novel reporter system

2.1 Background

Most of the published work on the phenomena of intron gain and loss has employed phylogenetic comparisons of intron presence and position across intron-containing genes, made possible by the existence of extensive genome sequence databases (Yenerall and Zhou, 2012). Definitive conclusions of intron gain or loss are difficult to make by these analyses, however it is clear that introns massively infiltrated the genome of the last eukaryotic common ancestor (LECA) and that introns have continued to have been gained and lost over evolutionary time (Irimia and Roy, 2014; Rogozin et al., 2012).

The critical limitation of proposed intron gain models is the lack of experimental validation. Although the existing models are supported by indirect evidence based on comparative analyses, none of them has been proven to be an authentic mechanism leading to intronogenesis.

We developed a novel reporter system to detect intron gain and loss events in *S. cerevisiae* using the *S. pombe his5+* gene interrupted by an artificial intron as a selectable marker to experimentally verity the intron transposition model mechanism in which a spliced out intronic RNA is a source of intron gain by reverse splicing-mediated RNA transposition.

2.1.1. Reversibility of pre-mRNA splicing

While the splicing of group II introns is readily reversible (Lambowitz and Zimmerly, 2011) , the pre-mRNA splicing reaction was known to proceed in the forward direction only. However, in 2008, Tseng and Cheng experimentally demonstrated the reversibility of the first and the second catalytic splicing steps using a mutant of the RNA helicase Prp22 *in vitro* (Tseng and Cheng, 2008). Prp22 is required for the second catalytic step of splicing in an ATP-independent manner (Schwer and Gross, 1998) and for the release of mRNA from the spliceosome in an ATP-dependent manner (Schwer and Gross, 1998; Wagner et al., 1998). The mutant they employed, *prp22S635A,* is able to catalyze the second transesterification reaction but cannot release the mRNA from the spliceosome after the splicing reaction is completed (Schwer and Meszaros, 2000). Therefore, the spliceosome retains both the mRNA and the excised lariat intron. In *in vitro* splicing assays with the affinity-purified spliceosome associated with this mutant RNA helicase, they showed that the second catalytic step of splicing is reversible in the absence of monovalent ions and also the reversed splicing intermediates proceed through the second transesterification reaction again when the monovalent ions are added back, suggesting that the spliceosome might be interchangeably rearranged to prefer the forward or the reverse reaction depending on the concentration of monovalent ions (Tseng and Cheng, 2008). In addition, the first catalytic step was also shown to be efficiently reversed when the forward reaction is inhibited. Notably, no errors in the selection of the splice sites through the forward or reverse reactions were detected, indicating that the spliceosome is highly faithful in the reverse reactions of splicing (Tseng and Cheng, 2008).

Still, experimental limitations exist in proving that pre-mRNA splicing can occur in the reverse direction *in vivo*. Reverse splicing was only shown *in vitro* under non-physiological condition. *In vitro* demonstration indicates the potentiality of reverse splicing, however, it remains unclear if this is indeed the case *in vivo*. Also, in the case of the reverse splicing of intron RNA into the original position within the corresponding transcript, as demonstrated experimentally (Tseng and Cheng, 2008), it is unlikely to be detected unless the location of the reverse splicing reaction changes.

2.2 Materials and methods

2.2.1 Construction of a reporter system to detect intron gain and loss

The intron reporter was generated by modifying the splicing reporter used previously in this laboratory (Sorenson and Stevens, 2014). The two-color fluorescence reporter containing the *RPL28* intron fused to eGFP and mCherry in the intron (*URA3*-marked, pRS316 backbone) was restriction enzyme digested with SphI and AflIII to remove the mCherry open reading frame. A synthetic construct (Genscript) containing the *S. pombe tef1+* promoter driving the *S. pombe his5+* gene and terminating with the *S. pombe tef1+* termination sequence was cloned into those same restriction sites.

2.2.2 Yeast strains and DNA oligonucleotides

SS4056, BY4733, and SS4019 were used for intron transposition assays. SS5230 was obtained by an intron gain event and SS5231, SS5232, SS5233 and SS5235 were generated to further analyze the *RPL8Bint* strain. *RPL8B-TAP* tagged strain from the Yeast-TAP tagged ORF library (GE Dharmacon), *rpl8a*Δ strain and *rpl8b*Δ strain

from the Yeast MATa Knock Out collection (GE Dharmacon) were employed to quantitatively analyze Rpl8 expressed in the *RPL8Bint* strains. Refer to Table 2.1 for the full genotype of yeast strains.

The oligonucleotides used in this chapter and their sequence information are listed in Table 2.2.

2.2.3 Yeast transformation with the reporter plasmid

Strains were grown at 31°C until OD_{600} reaches 1.0-1.5 for optimal transformation efficiency. 50 ml of liquid culture was pelleted and washed twice with 100 mM lithium acetate, then mixed with 33.3% PEG 8000, 100 mM lithium acetate, 100 ug of salmon sperm carrier DNA and 200 ng of pRS416 (negative control), pRS316HIS5AI (reporter), and pRS316HIS5AIΔbp (branch site deletion control), individually. The mix was heat shocked at 42°C for 45 minutes and pelleted, washed with ddH_2O, resuspended in 1 ml of ddH_2O. 100 ul of each reaction was plated on SD-ura and grown for 72 hours at 31°C.

Table 2.1 Yeast strains used in Chapter 2

Strain	Mating type	Genotype
SS4056	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0*
BY4733	a	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0*
SS4019	a	*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, trp1Δ*
SS5230	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0, RPL8B::his5+::int*
SS5231	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0, RPL8B::his5+::int::TAP::HYG*
SS5232	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0, URA3::GPD-RPL8B::his5+::int*
SS5233	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0, URA3::GPD-RPL8B::his5+::int::TAP::HYG*
SS5235	α	*his3::KanMX, leu2Δ0, lys2Δ0, ura3Δ0, rpl8aΔ, URA3::GPD-RPL8B::his5+::int*
RPL8B-TAP	a	*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RPL8B-TAP::HIS3MX6*
rpl8aΔ	a	*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, rpl8a::KanMX*
rpl8bΔ	a	*his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, rpl8b::KanMX*

Table 2.2 DNA oligonucleotides used in Chapter 2

Name	Sequence (5' to 3')
SpHIS5PCR5	CTCACATCACATCCGAACATAAAC
SpHIS5PCR3	GTTACTTTACAACACTCCCTTCGTGCTTGG
SpHIS5iPCR2F	CCAAGCACGAAGGGAGTGTTGTAAAGTAAC
SpHIS5iPCR2R	GTTTATGTTCGGATGTGATGTGAG
SpHIS5iPCR3F	CAAGTAATCCAAGTAGACAC
SpHIS5iPCR3R	ACATGCAAAGTAATTCCAGC
2014.8	GGAATTTGTATCTATCAATTACTATTCC
2014.7	GTATTGTGATGCGCACGTTGAAATTCAC
2014.5	CAGCGGTGTTTCTGTCCAAAGTG
RPL8BPCRA	TTCCATCCCAATGAAGATCCGTAC
RPL8BPCRB	AAAGGAAGTGAGAAGGAACAAGAG
RPL8BtapaHYG	AGCTAAGATGGACAAGAGAGCTAAGACTTCCGACTCCGCTATGGAAAAGAGAAGATGGAA
RPL8BtapBpRS41H	TACAAAATATAATTATATTACGATGTTCGAAATTCTATATACTGAGAGTGCAGCGACATG
RPL8B-GPD-URA-KIA	ATAGAACGCATTGAAACTTTTCCCATCTCAAAATCCAGGGACAATAGTATGGGATGCGCGTTTCGGTGATGACGGTG
RPL8BINT-GPD-URA-KIB	GGTTCTTTCATTCCCTCTTCCAAATGGAACTACATACTTTCTTACCTGGAGCCATTTTGTTTGTTTATGTGTGTTTATTCG

(Table 2.2 continued)

CYH2E1fwd	TAGAAAGCACAGAGGTCACGTC
GPD xhol	GGTATTGATTGTAATTCCG
GFP 5	ACTTGTGGCCGTTTACGTCG

2.2.4 Intron transposition assay

Individual uracil prototroph containing pRS316HIS5AI was grown in 10 ml SD-ura liquid medium overnight at 31°C. Pelleted cells were plated on a 150 mm SD-his plate and incubated at 31°C for 4 to 6 days. All *his+* colonies were patched onto YPD with G418 (200 ug/ml, Geneticin) plates, grown overnight at 31°C and then streaked on SD-his containing 1 mg/ml 5-fluoroorotic acid (5-FOA, Oakwood Chemical). The 5-FOA selects against cells expressing the *URA3* gene. After incubation at 31°C for 6 days, *his+* cells were scored for their resistance to 5-FOA.

2.2.5 Diagnostic PCR to determine genomic integration

Genomic DNA of all *his+* / 5-FOA^r colonies was isolated for diagnostic PCR to confirm the genomic integration of *his5+* gene. Primer sets of GPD xhoI / SpHIS5iPCR2F and SpHIS5iPCR2R / GFP 5 were used.

2.2.6 Confirmation of intron loss

For *his+* / 5-FOA-sensitive colonies, plasmid was isolated using Zymoprep™ Yeast Plasmid Miniprep II kit (Zymo Research) for sequencing to determine intron loss.

2.2.7 Inverse PCR analysis

Genomic DNA was isolated from *his+* / 5-FOA^r cells using YeaStar™ Genomic DNA Kit (Zymo Research) and cut sequentially with BstYI and BclI restriction endonucleases (New England Biolabs). The DNA was phenol extracted, ethanol precipitated, and then resuspended in ddH_2O to be ligated using T4 DNA ligase (New

England Biolabs) under the condition favoring self-ligation (Derr et al., 1991). After overnight ligation reaction at 16°C, the DNA was extracted, precipitated, and resuspended in 50 ul of ddH_2O for use in inverse PCR. DNA was amplified with the primers hybridizing within the reporter intron region using Taq DNA polymerase with ThermoPol buffer (New England Biolabs). Primer pairs of SpHIS5iPCR2F / SpHIS5iPCR3R and SpHIS5iPCR3F / SpHIS5iPCR2R were used. PCR products were sequenced using the same primers to analyze the chromosomal integration site.

2.2.8 *RPL8Bint* identification and confirmation by PCR

The intron insertion in *RPL8B* gene was first identified by direct genomic DNA sequencing (Horecka and Jigami, 2000) and confirmed by PCR with primers hybridizing upstream and downstream of the *RPL8B* gene. A primer pair of 2014.7 / 2014.5 was used.

2.2.9 TAP tagging of *RPL8Bint*

The TAP tag was introduced at the C-terminus of the genomic *RPL8Bint* locus by homologous recombination with hygromycin B resistant selectable marker (HYG) adjacent to the TAP cassette (Puig et al., 2001). Oligonucleotide primers contain a 40 bp-long region identical to the yeast genome for homologous recombination. Primers, RPL8BtapA-HYG and RPL8BtapBpRS41H, hybridize at the 5' end of the CBP coding sequence and at the downstream of the coding sequence of hygromycin B phosphotransferase (HPH), respectively.

2.2.10 *RPL8Bint* promoter switch

The GPD (*TDH3*) promoter was introduced to replace the promoter of *RPL8Bint*. Primers contain a 55 nucleotide-long region identical to regions upstream and downstream of the *RPL8Bint* start codon for homologous recombination. Oligonucleotides, RPL8B-GPD-URA-KIA and RPL8Bint-GPD-URA-KIB, were designed to amplify a DNA fragment containing the GPD promoter with the *URA3* marker for selection purpose.

2.2.11 Yeast sporulation and tetrad dissection

GPD-RPL8Bint (SS5232) and *rpl8a*Δ strains were mixed on an YPD plate within a drop of ddH_2O, incubated overnight at 31°C, and then diploids selected by streaking on SD-his-lys media. The diploid cells were inoculated into 2 ml of SD-his, incubated 30 hours at 31°C, harvested and washed with water. The cells were resuspended in 2 ml of SPM (100 mM Potassium acetate, 400 uM Raffinose) and incubated on a roller drum at room temperature for 5 days to sporulate. 100 µl of the sporulated culture was treated with 15 µg of Zymolyase T100 in 20 mM β-mercaptoethanol at 30°C for 20 min and dissected on a tetrad dissection microscope.

2.2.12 Protein isolation and western blot analysis

Whole cell extract was prepared as previously described with minor modifications (Yaffe and Schatz, 1984). Yeast cells were grown in YPD at 31°C and 1 OD_{600} unit of the cells was harvested by centrifugation, resuspended in 1 ml of YPD, chilled on ice for 10 min and then mixed with 150 µl of ice cold 2 N NaOH, 8% β-mercaptoethanol. After an

additional 10 min incubation on ice, the cells were mixed with 150 µl of ice cold 50% TCA, chilled on ice for another 10 min, and pelleted by centrifuging at 4°C at 13,000 RPM for 2 min. The pellets were washed with 1 ml of ice cold acetone and resuspended in 100 µl of 1X NuPAGE® LDS sample buffer containing 100 mM DTT. Whole cell extract was separated by SDS-PAGE and transferred to nitrocellulose membrane (Bio-Rad) in transfer buffer (20% methanol, 0.001% SDS, 386 mM glycine, 48 mM Tris base). Western blot analysis to detect TAP-tagged protein, GFP, Rpl8, and Rps8 was performed separately as follows. The membrane was blocked for 1 hour at room temperature in BLOTTO (5% dry milk, 1x PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH_2PO4, 0.2% Tween-20)), washed twice for 5 minutes each in 1X PBST, incubated for 1 hour in BLOTTO containing a primary antibody at appropriate dilution: 1:10000 for rabbit Peroxidase Anti-Peroxidase (PAP) (Rockland), 1:5000 anti-GFP (GenScript) conjugated to HRP using Innova Biosciences Lightning-Link HRP conjugation kit, 1:30000 for rabbit anti-Rpl8 (A. Johnson, UT Austin), 1:5000 for rabbit anti-Rps8 (A. Johnson, UT Austin). After four times of 15 minutes washing in 1X PBST, the membrane was incubated in BLOTTO containing 1:10000 dilution of anti-rabbit IgG (H&L) (mouse) antibody peroxidase conjugated for 1 hour except for HRP-conjugated anti-GFP. After washing, the blots were visualized on X-ray film by chemiluminescence using Western Lightning Plus ECL substrate (Perkin Elmer) or chemifluorescence using Pierce™ ECL Plus Western Blotting Substrate (Thermo Scientific). Quantitation of protein signals in western blots was performed using the Quantity One analysis software (Bio-Rad) and ImageJ (National Institutes of Health).

2.2.13 Total RNA extraction

Total RNA was extracted by lysing cells in 1:1 (v:v) of acid phenol and AES buffer (50 mM NaOAc pH 5.3, 10 mM EDTA, and 1% SDS). Cells were vortexed, incubated at 65 ºC for 5 minutes with vortexing every minute, and then chilled on ice for another 5 minutes (Wise, 1991). RNA was purified using 2 mL Phase Lock Gel tubes (5 PRIME) and precipitated with isopropanol at -80 ºC overnight. The pellets were washed with ice cold 70% ethanol and resuspended in nuclease-free ddH_2O.

2.2.14 RT-PCR analysis to detect recursive splicing products

Total RNA (5 µg) was RQ1 DNase treated (Promega) for 20 minutes at 37°C and purified using RNA Clean and Concentrator kit (Zymo Research). After adding 1 µL 10X RT Buffer (0.5 M Tris-HCl pH 8.5), 1 µL random nonamer primers (25 µM, Integrated DNA Technologies) and 2 µL DEPC H_2O to 5ul of purified RNA, the RNA / primer mix was incubated at 60°C for 5 minutes and then chilled on ice for 5 minutes. 11 µL of an RT master mix (1X RT Buffer, 3 mM $MgCl_2$, 10 mM DTT, 0.5 mM dNTPs and 5 U/µL MultiScribe RT enzyme (Applied Biosystems)) was added. The RT reaction was conducted by incubating overnight at 42°C. The resulting cDNA was subject to PCR to amplify the recursive splicing products using primers, CYH2E1fwd and GPD 5.

2.2.15 Serial dilutions and spotting test

Strains were grown in YPD, adjusted to OD_{600} of 1.0 and serially diluted by 10-fold with the same media. 3 uL of each dilution was spotted on YPD plate and incubated

at different temperatures. Growth at 25 °C, 31 °C, and 37 °C was monitored for 3 days and at 16°C up to 10 days.

2.3 Results

2.3.1 Construction and validation of the reporter system

The design of the reporter to detect intron gain and loss is outlined in Fig. 2.1(A). Short stretches of exons 1 and 2 and the intron of the *RPL28* gene have been fused to eGFP. We have inserted a synthetic construct containing the heterologous *S. pombe tef1+* promoter driving transcription of the *S. pombe his5+* gene and the *S. pombe tef1+* terminator within the intron. The expression of *S. pombe his5+* gene confers histidine prototrophy in *S. cerevisiae his3* mutants, but only after splicing of the artificial intron within *S. pombe his5+* gene and RNA transposition. This reporter is conceptually related to ones used previously to detect *de novo* Ty1 retrotransposition events (Curcio and Garfinkel, 1991) and RNA-mediated intron loss (Derr et al., 1991), however the selection signal for the other reporters was contained on the resulting mRNA whereas this reporter selection signal is contained within the excised intron. Intron mobilization leading to histidine prototrophy can be a result of a few different events including intron loss, intron gain and pseudogene formation.

The key feature of this reporter is that the transcriptional direction of *S. pombe his5+* gene is in the opposite orientation to that of the eGFP construct. Within the *his5+* gene, we have inserted an artificial intron (modeled on those described in Curcio and Garfinkel, 1991; Derr et al., 1991) containing pre-mRNA splicing signals that are only capable of being spliced from the eGFP transcript, but not from the *his5+* transcript. Only

after splicing of the artificial intron followed by RNA-mediated recombination, transcription of the *his5+* gene driven by *tef1+* promoter allows histidine prototrophy in *S. cerevisiae his3* mutants. As a splicing disabled control, we also designed a HIS5AIΔbp construct which lacks the branchpoint sequence within the artificial intron.

The proper splicing of this reporter was verified by virtue of eGFP protein production as shown in Fig. 2.1(B). Although the size of eGFP intron is 1339 nucleotides with the artificial intron, which is much longer than the natural introns in *S. cerevisiae*, we were able to detect eGFP splicing and expression in both HIS5AI and HIS5AIΔbp constructs by Western blot analysis.

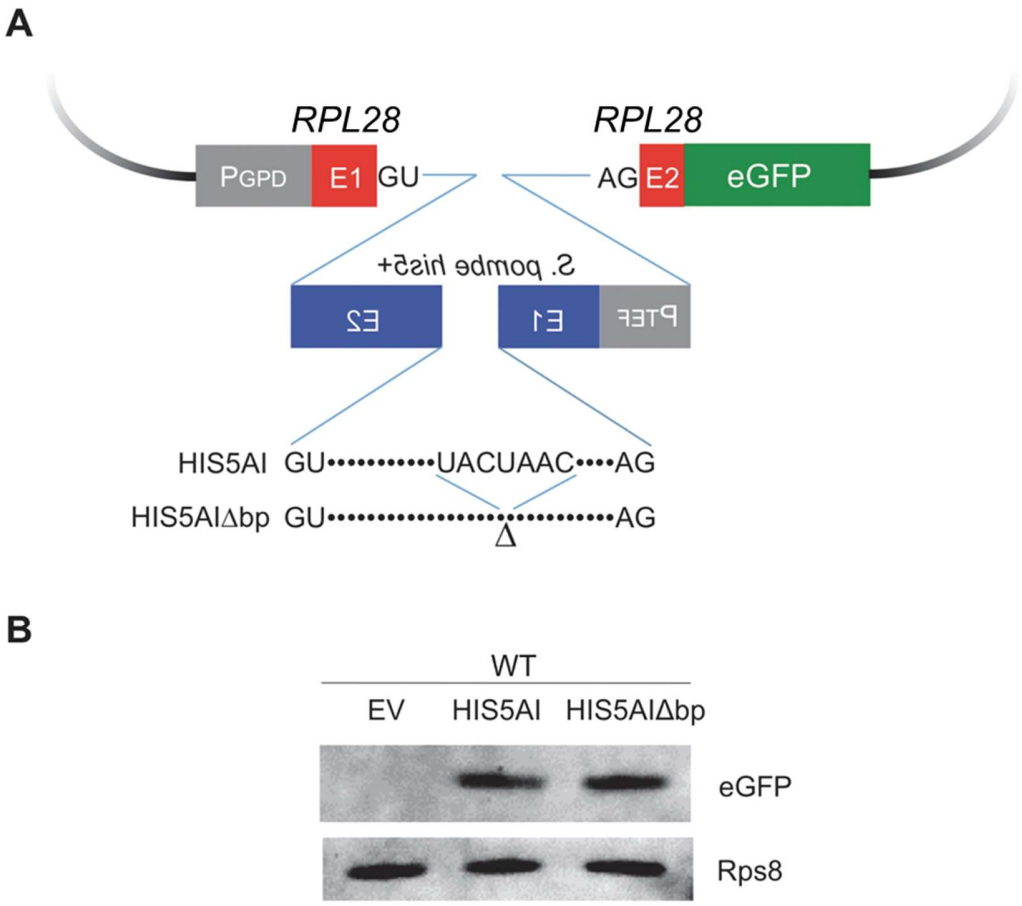

Figure 2.1 Design and validation of the intron gain and loss reporter

(A) Model of the reporter construct. *RPL28* gene exon 1 and exon 2 with portions of the intron were fused to eGFP using pRS316 as a backbone construct. The *S. pombe his5+* gene was inserted into the *RPL28* intron (eGFP gene). The his5+ gene was interrupted by an artificial intron (AI) containing splice sites which are only spliced from the eGFP transcript. The AI lacking the branchpoint sequence (AIΔbp) cannot be spliced from the eGFP transcript. (B) Demonstration of eGFP production. Splicing and expression of eGFP from the reporter was demonstrated by Western blot analysis using an antibody to GFP. Rps8 was used as a loading control.

2.3.2 Detection of recursive splicing products

In our reporter system, we hypothesized that the artificial intron spliced out from the *his5*+ gene is the substrate for intron transposition. As the eGFP intron and the artificial intron are oriented in the same direction and both can be spliced in accordance with the transcription driven by GPD promoter (Fig. 2.2(A)), the formation of the artificial intron splicing product needed to be verified.

The array of recursive splicing products was tested by RT-PCR and sequence analysis of the PCR products (Fig. 2.2). Oligonucleotide primers hybridizing to the exon 1 of *RPL28* gene and the eGFP gene were used for PCR to capture all the predicted transcripts as displayed in Fig. 2.2(B-F). For both HIS5AI and HIS5AIΔbp, the unspliced pre-mRNA (Fig. 2.2(B)) due to inefficient splicing and the splicing of the eGFP intron (Fig. 2.2(F)) as confirmed by Western blot analysis were easily detected. Splicing also occurred between the 5' splice site of the artificial intron and the 3' splice site of the eGFP intron (Fig. 2.2(E)). Although we did not detect the transcript (Fig. 2.2(D)) that can be produced only from the HIS5AI construct, not from the HIS5AIΔbp due to the lack of branchpoint sequence, we were able to observe the transcript (Fig. 2.2(C)) from the HIS5AI by further electrophoresis, indicating the formation of the spliced out artificial intron (Fig. 2.2(G)).

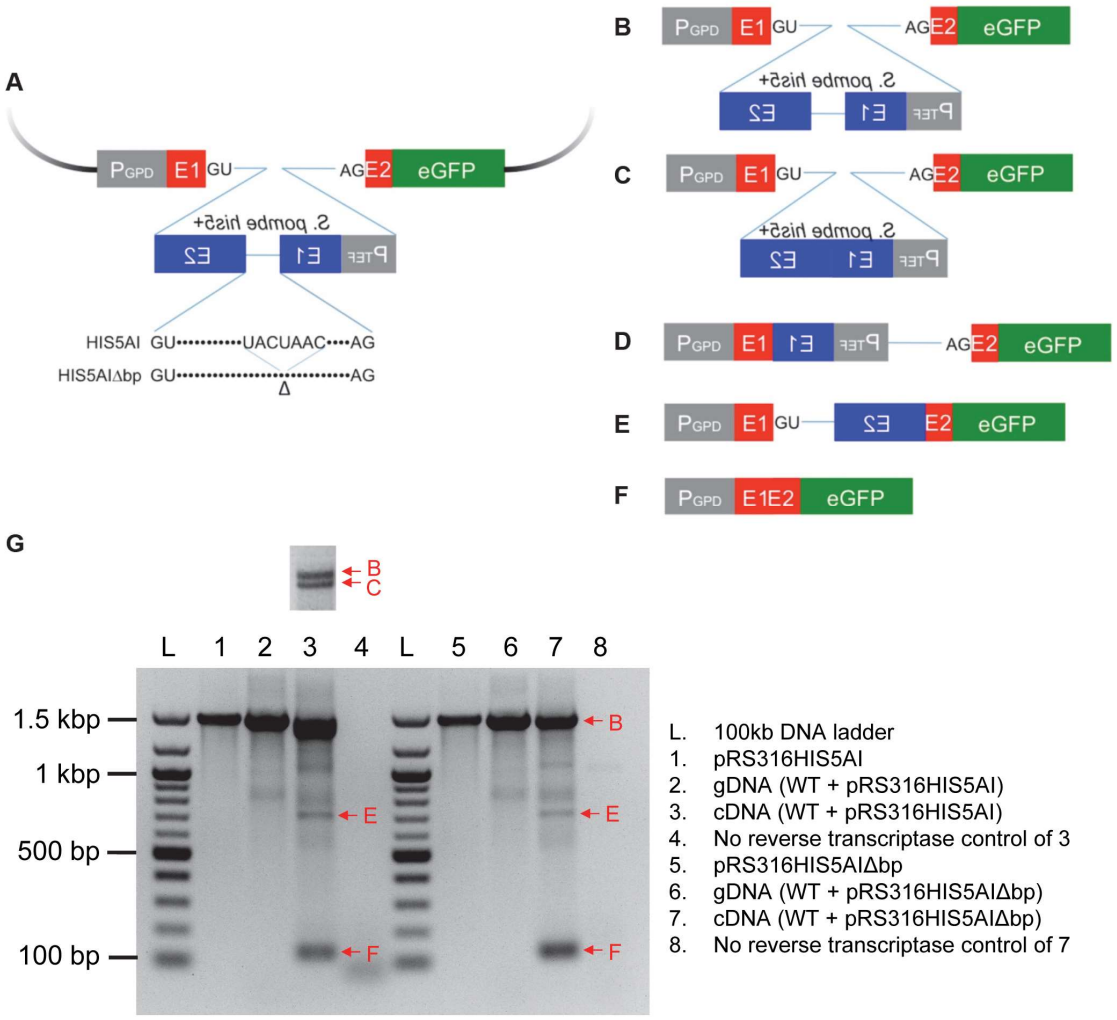

Figure 2.2 Detection of recursive intron splicing (A) The intron reporter constructs of both HIS5AI and HIS5AIΔbp. (B-F) Predicted and detected recursive intron splicing products. (B) The reporter pre-mRNA transcribed without being processed by splicing. (C) Only his5+ intron is spliced. (D) Splicing occurs at 5' splice site of the eGFP intron and 3' splice site of the his5+ intron. (E) Splicing occurs at 5' splice site of the his5+ intron and 3' splice site of the eGFP intron. (F) The entire eGFP intron is spliced. Note that the HIS5AIΔbp construct is not able to produce C or D. (G) Detection of the transcripts from the reporter

(Fig. 2.2 continued)

construct by RT-PCR. For comparison, PCR was performed using the reporter plasmids (lanes 1 and 5) and the corresponding genomic DNA (lanes 2 and 6) in addition to the no reverse transcriptase control (lanes 4 and 8). The expected sizes of each transcript are as follows: B (1473 bp), C (1370 bp), D (787 bp), E (717 bp), and F (134 bp)

2.3.3 Formation of *his+* prototroph

Our screen was conducted on 1.6 x 10^{11} wild-type cells containing the reporter plasmid and 7076 *his+* prototrophs were generated by potential intron mobilization events. We were able to distinguish the plasmid-borne events from the chromosomal integration events by whether the cells need to maintain the reporter plasmid. As the *HIS* expression is essential for growth, the plasmid-borne *his+* prototroph kept the reporter plasmid, leading to lethality in the 5-FOA selection. However, the *his+* prototrophs that resulted from the chromosomal integration of *his5+* gene naturally lost the reporter plasmid with histidine supplemented, exhibiting resistance to 5-FOA selection. Diagnostic PCR was also performed to categorize each *his+* prototroph as a plasmid-related event or a chromosomal event.

2.3.4 Detection of intron loss events

The most common event in our screen is plasmid-borne intron loss, related to the RNA-mediated recombination events shown previously (Curcio and Garfinkel, 1991; Derr et al., 1991) (Fig. 1.4 and Fig. 2.3). In this, the intron within the *his5+* gene has been spliced out, and a species of the eGFP intron RNA has been reverse transcribed and recombined back into the reporter plasmid. From these data, it is not possible to determine if this is a singly spliced, intron-containing eGFP pre-mRNA or the liberated eGFP intron that gave rise to these events, however sequencing of the *his5+* gene in the plasmid reveals that the *his5+* artificial intron has been precisely removed (data not shown). As a control, a screen with the HIS5AIΔbp reporter construct lacking the branchpoint sequence within the artificial intron produced no histidine prototrophs

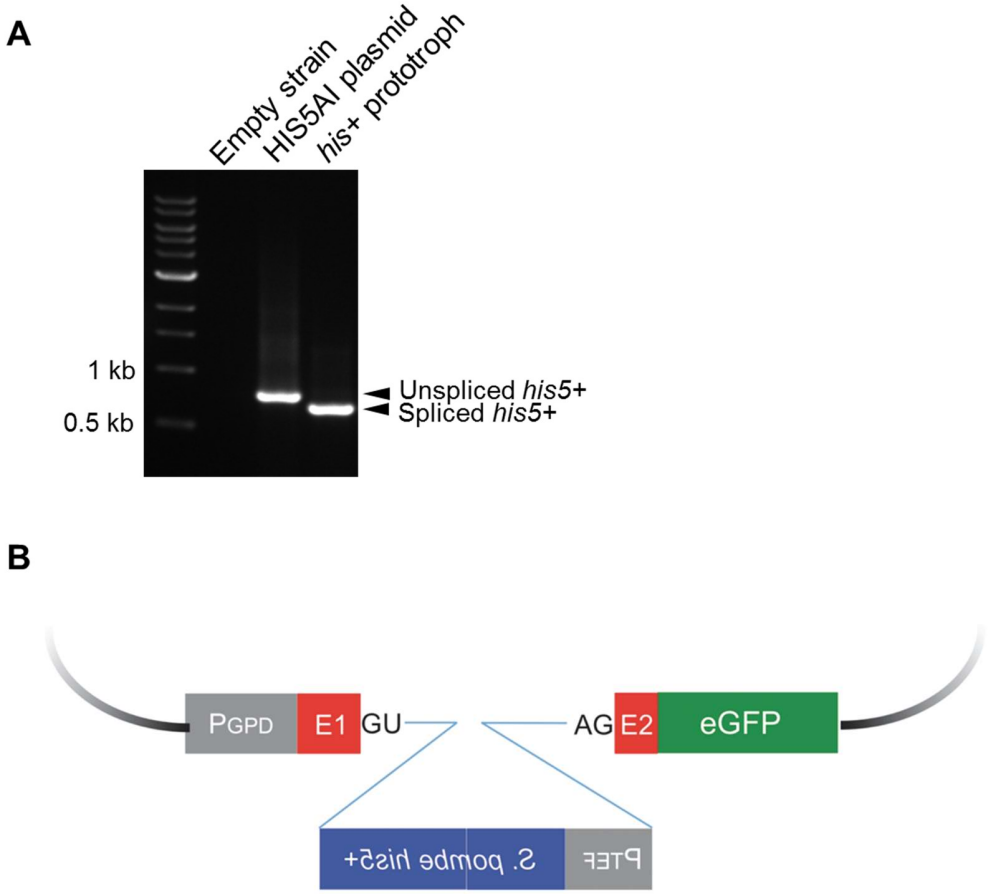

Figure 2.3 Plasmid-borne intron loss events

(A) PCR products of the *S. pombe his5+* gene from the HIS5AI reporter plasmid and the genomic DNA of the plasmid-borne *his+* prototroph. The size difference in the PCR products indicates the loss of the artificial intron (103 bp) in *his5+* gene. (B) The plasmids rescued from *his+* cells have lost the artificial intron within *his5+* gene.

(Fig. 2.2(A)), indicating that removal of the artificial intron from the eGFP intron by splicing followed by RNA transposition of the eGFP intron is required for histidine prototroph formation.

2.3.5 Detection of an intron gain event

We detected a single intron gain event that resulted from a chromosomal addition of the reporter intron into an mRNA-encoding gene. Direct genomic sequencing (Horecka and Jigami, 2000) and inverse PCR analysis from genomic DNA (Fig. 2.4 and Fig. 2.5) in that *his*+ / 5-FOA-resistant strain revealed that a single intronogenesis event occurred in the *RPL8B* gene. Inverse PCR was performed as described in Fig. 2.4(A) to analyze the insertion site within the genome and the resulting PCR products showed the size difference compared to those of the plasmid-born *his*+ prototroph (Fig. 2.4(B)).

Sequencing analysis confirmed that the transposed intron was inserted 18 nucleotides downstream from the start codon of the *RPL8B* gene (Fig. 2.5). This new allele was named *RPL8Bint.* The eGFP intron maintained intact splicing signals and the artificial intron had been precisely removed, suggesting that the transposed intron can be processed by the splicing machinery and the resulting mRNA is competent for translation into a proper Rpl8 protein.

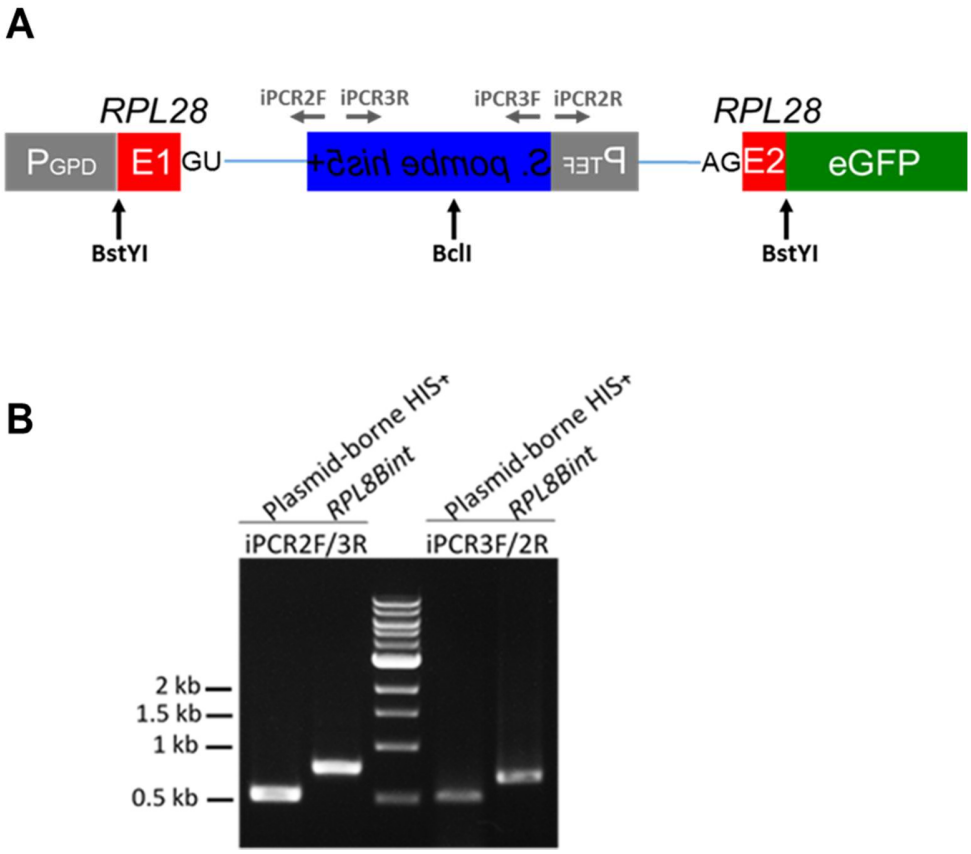

Figure 2.4 Inverse PCR analysis

(A) 5-FOA-resistant histidine prototrophs were subjected to the restriction digestion and inverse PCR to analyze the insertion sites. The cartoon displays the location of the BstYI and BclI cleavage sites within the reporter construct after splicing of the artificial intron. Two pairs of primers, iPCR2F / iPCR3R and iPCR3F / iPCR2R, were designed to amplify the sequences containing the partial *S. pombe his5*+ gene. (B) For the plasmid-borne intron loss event, 529 bp and 535 bp PCR products are expected to be amplified with the two sets of primers respectively, while the intron insertion into the chromosome gives larger PCR products. Inverse PCR analysis of *RPL8Bint* produced 780 bp and 680 bp products.

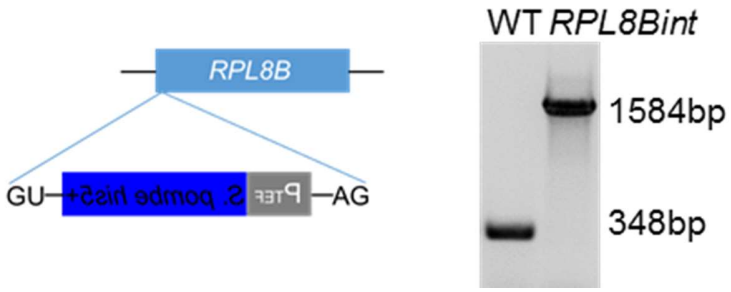

Figure 2.5 Transposition of the eGFP intron into the *RPL8B* locus

The *RPL8Bint* locus: (L) The entirety of eGFP intron (lacking the spliced-out artificial intron, 1236 bp) was transposed 18 nucleotides downstream from the start codon of *RPL8B* (771 bp). It contains all of the pre-mRNA splicing signals required to splice the *RPL8Bint* transcript. (R) PCR products of the entire *RPL8B* locus from genomic DNA of wild-type and *RPL8Bint* strains. The size difference in the PCR products of *RPL8Bint* indicates the acquisition of the reporter intron in *RPL8Bint*.

2.3.6 *RPL8Bint* is capable of splicing and expression.

To test if the reporter intron within the new allele *RPL8Bint* is spliced and expressed, we conducted western blot analysis of a strain in which *RPL8Bint* has been TAP tagged by homologous recombination (Fig. 2.6(A)). Quantitative comparisons of the levels of the *RPL8B-TAP* (Ghaemmaghami et al., 2003) to that of *RPL8Bint-TAP*, shows that Rpl8 protein produced from the *RPL8Bint-TAP* locus is only ~2% of that from the *RPL8B-TAP* locus (Fig. 2.6(B)). The reduced expression is most likely due to the inefficient splicing of the very large intron containing the *his5+* gene (Sorenson and Stevens, 2014). *S. cerevisiae* introns are typically much shorter than in other species and as size increases, splicing efficiency dramatically decreases (Klinz and Gallwitz, 1985). Indeed, at 1236 nucleotides, this intron is 235 nucleotides longer than the largest natural intron in *S. cerevisiae* (Spingola et al., 1999). Thus it is not surprising that knocking out the Rpl8 paralog *RPL8A* in *RPL8Bint* proved to be impossible either by direct removal or by crossing and dissection of tetrads (data not shown).

2.3.7 Overexpression of *RPL8Bint* restores the expression level.

As the levels of Rpl8 produced from *RPL8Bint* were too low to support growth, we modified both *RPL8Bint* and *RPL8Bint-TAP* to be driven from a strong heterologous promoter derived from the *TDH3* gene (Bitter and Egan, 1984) (Fig. 2.7(A)). Western blot analysis was performed to compare the expression levels of Rpl8-TAP between the *RPL8B-TAP* strain and the *GPD-RPL8Bint-TAP* strain. Quantitation of the signals verified that this modification restored the levels of Rpl8 to the wild-type levels in the *GPD-*

RPL8Bint-TAP strain (Fig. 2.7(B)), indicating that expression from *RPL8Bint* may suffice for production of all of Rpl8.

2.3.8 Rpl8 expressed from *GPD-RPL8Bint* strain is sufficient enough to overcome deletion of *RPL8A*.

As stated in 2.3.7, knocking out the Rpl8 paralog *RPL8A* in *RPL8Bint* could not be accomplished probably because the total amount of Rpl8 expressed is not sufficient for survival. Instead, *RPL8A* was deleted from the *GPD-RPL8Bint* strain by crossing *GPD-RPL8Bint* with the *rpl8A*Δ strain followed by sporulation and tetrad dissection. The strain was tested by PCR to confirm the genotype (Fig. 2.8). The *rpl8a*Δ / *GPD-RPL8Bint* strain was then subjected to Western blot analysis. Indeed, total levels of Rpl8 are very similar in wild-type, *GPD-RPL8Bint,* and *rpl8a*Δ / *GPD-RPL8Bint* strains (Fig. 2.9), indicating that the Rpl8 expressed from *GPD-RPL8Bint* locus is sufficient in these cells.

2.3.9 *GPD-RPL8Bint* strain has no growth defect.

Not only is the expression level comparable to the wild type, Rpl8 in *rpl8a*Δ / *GPD-RPL8Bint* strain is also fully functional as it exhibits no detectable growth defects (Fig. 2.10). Serially diluted cultures were grown to determine the relative growth rate between the strains, demonstrating that the overexpressed *RPL8Bint* allele is functionally comparable to the wild-type strain. *RPL8Bint* showed only a slight growth defect at 25 °C as observed in *rpl8a*Δ and *rpl8b*Δ strains.

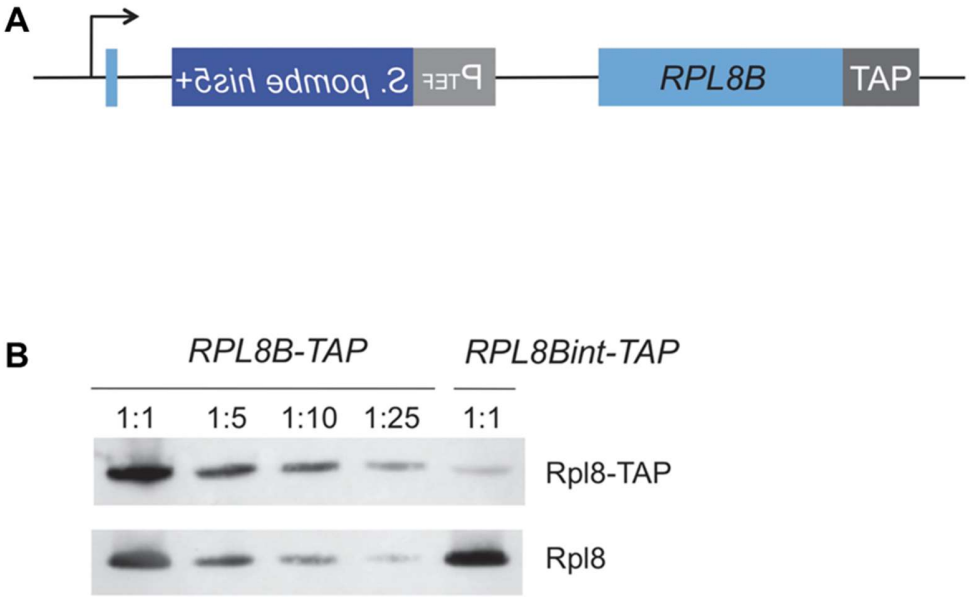

Figure 2.6 *RPL8Bint* expression

(A) TAP tagging of *RPL8Bint*. The TAP tag was inserted into the C-terminus of the *RPL8Bint* gene. (B) Quantitation of Rpl8 expressed from *RPL8Bint-TAP*. To quantitatively compare Rpl8-TAP protein levels in the indicated strains, *RPL8B-TAP* was serially diluted as shown in the figure. Rpl8 and Rpl8-TAP were detected by an antibody to Rpl8 and quantitated by chemifluorescence.

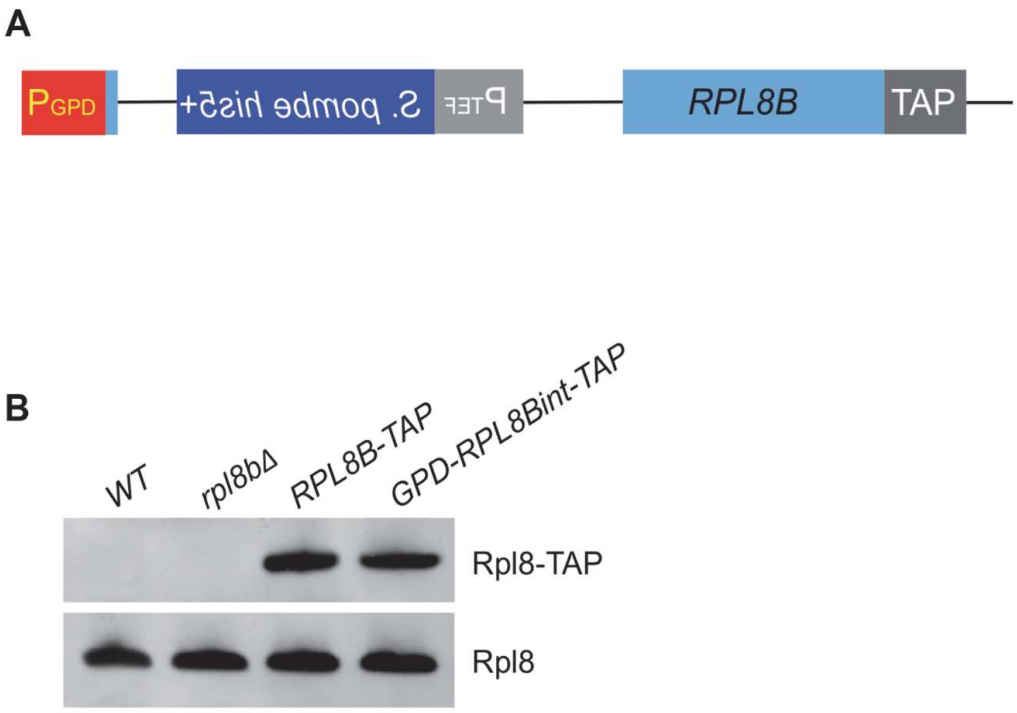

Figure 2.7 *RPL8Bint* expression with the GPD promoter

(A) *GPD-RPL8Bint* construction. The original promoter of the *RPL8Bint* (not shown) and *RPL8Bint-TAP* was replaced with the GPD promoter to increase the expression level. (B) Quantitation of Rpl8 expressed from *RPL8B-TAP* to *GPD-RPL8Bint-TAP*. To determine the effect of GPD promoter, the amount of Rpl8 expressed from *RPL8B-TAP* and *GPD-RPL8Bint-TAP* strains was determined by Western blotting. Rpl8 and Rpl8-TAP were detected by an antibody to Rpl8.

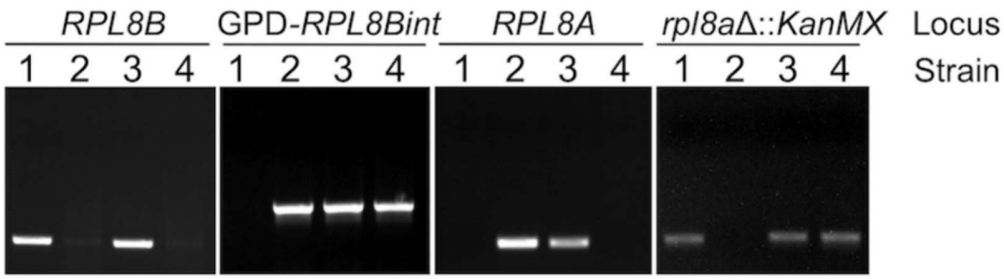

Figure 2.8 PCR confirmation of *rpl8a*Δ */ GPD-RPL8Bint* strain

Genomic DNA amplification of a portion of *RPL8B*, *GPD-RPL8Bint*, *RPL8A*, and *rpl8a*Δ*::KanMX* loci from each strain labeled 1 to 4 (1. *rpl8a*Δ, 2. *GPD-RPL8Bint*, 3. Diploid of *rpl8a*Δ and *GPD-RPL8Bint*, 4. Haploid *rpl8a*Δ / *GPD-RPL8Bint*).

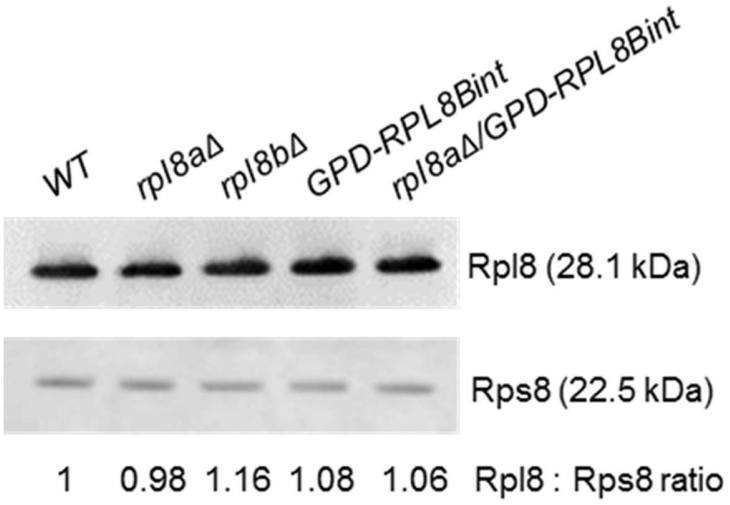

Figure 2.9 Quantitative Western blot

Total levels of Rpl8 were determined in wild-type, *rpl8a*Δ, *rpl8b*Δ, *GPD-RPL8Bint*, and *rpl8a*Δ / *GPD-RPL8Bint* strains by western blotting using chemifluorescence. Rps8 was detected as a loading control by an anti-Rps8.

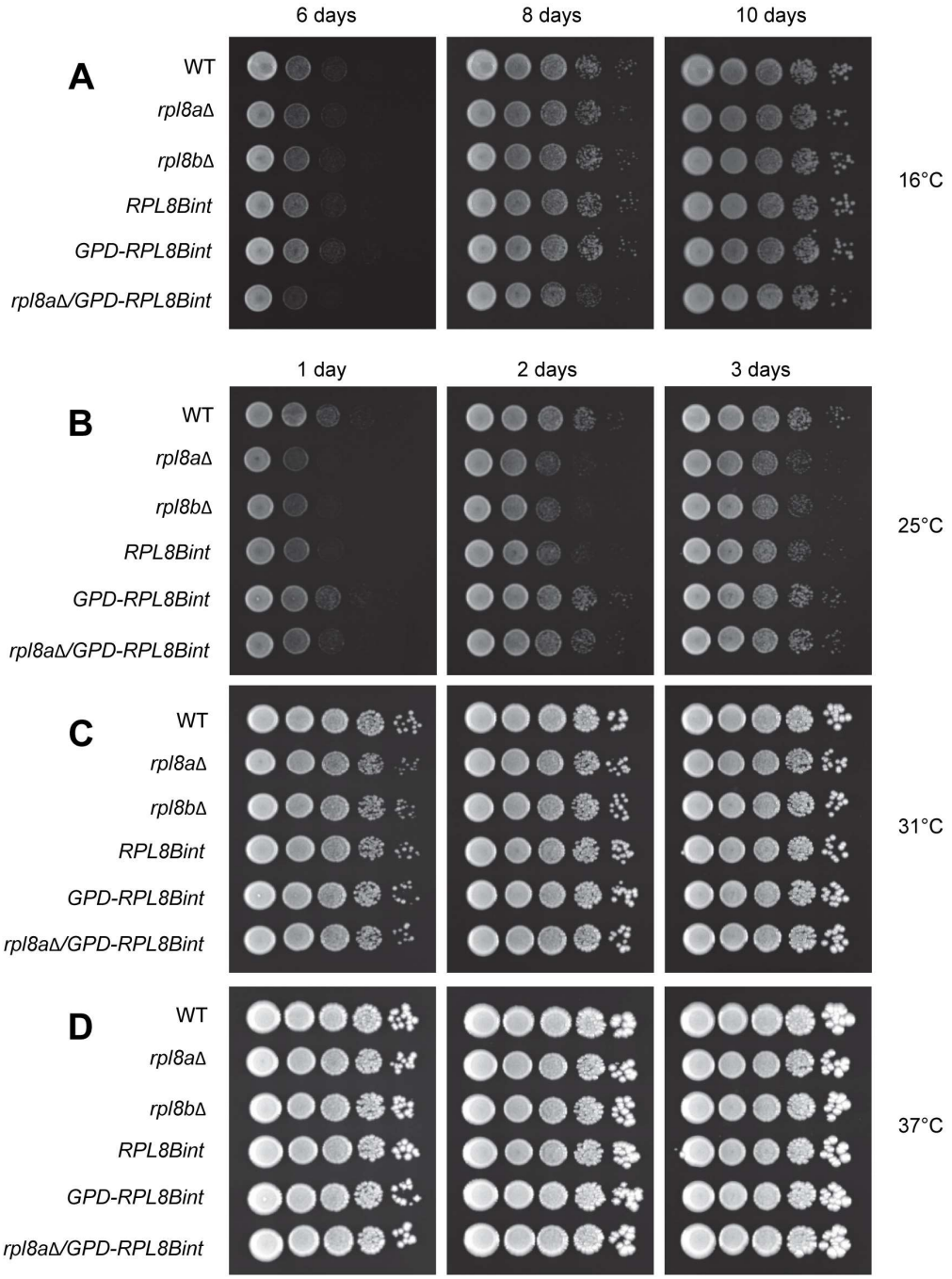

Figure 2.10 Overexpressed *RPL8Bint* allele is functionally comparable to the wild-type strain.

Relative growth rate of *RPL8Bint* was determined by serial dilution analysis. *RPL8Bint*

(Fig. 2.10 continued)

shows a slight growth defect as seen in *rpl8a*Δ and *rpl8b*Δ strains. When *RPL8Bint* is overexpressed, no significant growth defect is detected even in the absence of *RPL8A*. For serial dilutions, each strain was grown in YPD liquid culture, adjusted to OD_{600} of 1.0 and serially diluted by 10-fold with the same media. Growth on YPD plates was monitored at (A) 16°C for 10 days and at (B) 25°C, (C) 31°C, and (D) 37°C up to 3 days.

2.4 Discussion

We developed a novel reporter system to detect *in vivo* intron gain and loss in *S. cerevisiae* and conducted the intron mobilization assay. It may not be surprising that in ~ 2 x 10^{11} cells we captured only a single intron gain event while observing nearly 7,000 plasmid-borne intron loss events considering that intron gain involves an additional reverse splicing step which is expected to be an exceptionally rare event *in vivo* (Roy and Irimia, 2009). Previous work showed the capacity of the spliceosome to achieve the reversal of the two catalytic steps of splicing *in vitro* (Tseng and Cheng, 2008) and our intron gain mutant is likely to have been the result of an *in vivo* reverse splicing reaction. We would like to point out that the sequences of our reporter construct and the location of insertion in *RPL8B* do not have even short regions of homology which could explain this intron gain as a microhomology-based recombination event (Truong et al., 2013) (Fig. 2.11). Interestingly, 3 of 4 nucleotides in the *RPL8B* transcript at the site of intron insertion (the -2, -1, +1 and +2 positions relative to the intron) and the nucleotide at the -3 position possess base-pairing potential with the invariant U5 loop at the proper positions (Newman and Norman, 1992; Newman et al., 1995) (Fig. 2.12). This provides further evidence that reverse splicing may have occurred and suggests that in this mode of intron gain, new mRNAs may be selected for intron addition by the spliceosome at sites at which splicing may be more efficient in the forward reaction (Fig. 2.13).

Here we report the first experimental demonstration of spliceosomal intron transposition in any eukaryotic organism. We detected the insertion of the reporter intron into the genomic *RPL8B* using a novel reporter construct. In our model of *RPL8Bint* formation, the reporter intron RNA lacking the *his5+* artificial intron was retained in the

residual spliceosome after Prp22-mediated mRNA release (Fig. 2.13). This complex then encountered an *RPL8B* mRNA and inserted the reporter intron by reverse splicing (Curcio and Garfinkel, 1991; Tseng and Cheng, 2008). *RPL8Bint* pre-mRNA was then reverse transcribed by a cellular reverse transcriptase (likely derived from the Ty1 retroelement, data not shown) and the resulting *RPL8Bint* cDNA was incorporated into the *RPL8B* genomic locus by homologous recombination.

Although how the spliceosomal introns emerged and why only eukaryotes possess this class of introns remain unclear, it is widely accepted that the spliceosomal introns originated from the group II self-splicing introns and appeared at the time of eukaryogenesis (Irimia and Roy, 2014; Rogozin et al., 2012). It is likely that the proto-spliceosomal introns resembled the group II introns and that for some period of time retained their mobility and self-splicing ability (Lambowitz and Zimmerly, 2011). This may have enabled and promoted considerable proliferation of these introns at the time of eukaryogenesis (Irimia and Roy, 2014; Rogozin et al., 2012). The group II intron-derived features are likely to have degenerated over time, leading to a necessity for the development of the pre-mRNA splicing machinery composed of *trans*-acting snRNAs and *trans*-acting protein factors.

Demonstration of intron gain events *in vivo* has been a major obstacle to test existing models of intron gain. The *in vivo* intron gain presented in this work strengthens the intron gain model involving reverse splicing by providing the first evidence that it can occur on a laboratory time scale. At this point in the arc of evolution, for an intron gain event to become fixed in a population, either the intron insertion needs to result in a selective advantage to that organism or individuals possessing a new intron insertion

event need to find themselves in ecological niches in which they can further speciate. Indeed in recent reports, some intron gain events have been suggested to have occurred by means other than reverse splicing, and are highly mosaic depending on the individuals that have been analyzed (Li et al., 2009; van der Burgt et al., 2012; Worden et al., 2009) indicating that it may be that there are multiple mechanisms of intron gain which may change over time and genetic circumstance.

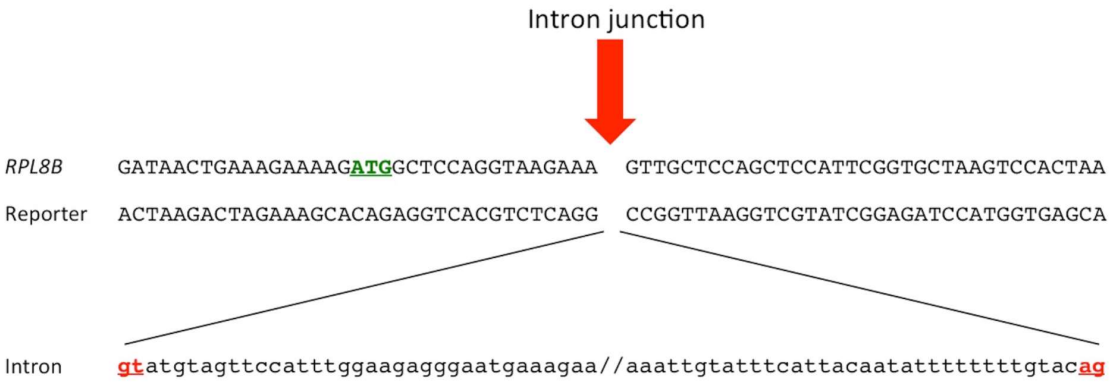

Figure 2.11 Microhomology-mediated end joning rules preclude recombination as a means of the intron acquisition.

The coding strand sequences of the reporter construct and *RPL8B* gene at both ends of the inserted intron were juxtaposed to examine the possibility of very short sequence homology. Note the start codon of *RPL8B* highlighted in green. The upstream and downstream sequences of *RPL8B* at the intron junctions do not show potential microhomology to that of the reporter construct.

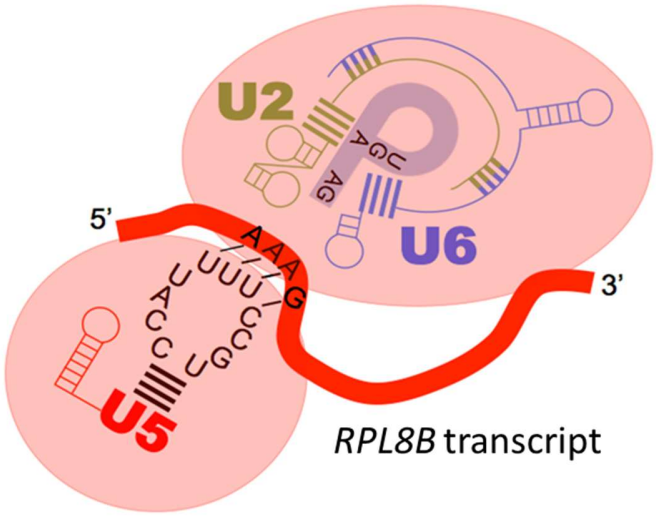

Figure 2.12 Predicted base-pairing interactions between the *RPL8B* transcript and the U5 snRNA at the new splice site.

3 of 4 nucleotides in the *RPL8B* transcript at the site of intron insertion (the -2, -1, +1 and +2 positions relative to the intron) and the nucleotide at the -3 position possess base-pairing potential with the invariant U5 loop at the proper positions.

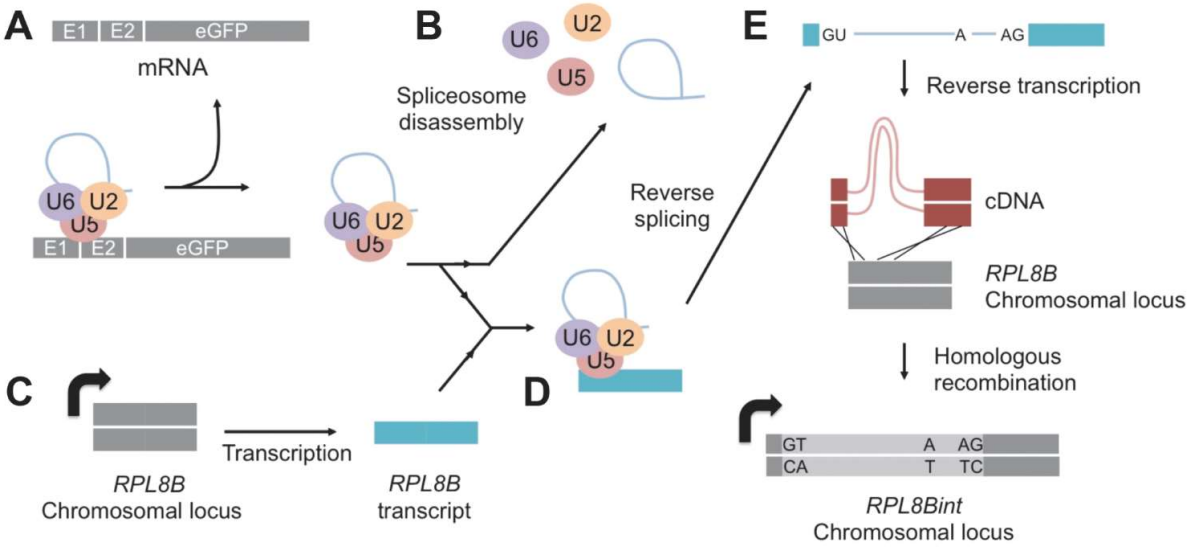

Figure 2.13 Proposed model for the mechanism of *RPL8B* intron gain (A) The last step of the pre-mRNA splicing reaction. The spliced mRNA (top) is disengaged from the residual spliceosome composed of the lariat intron, the U2, U5 and U6 snRNPs and other splicing proteins. (B) Disassembly of the residual spliceosome under normal conditions. The U2, U5 and U6 snRNPs and the lariat intron are dissociated from each other at the end of the splicing reaction. (C) Transcription of the *RPL8B* mRNA, the proposed substrate for the intron transposition. (D) Model of the transposition of the reporter intron contained within the intact residual spliceosome into the *RPL8B* transcript. The *RPL8B* transcript was captured by an intact, lariat intron containing spliceosome prior to going through both catalytic steps of pre-mRNA splicing, in reverse (Tseng and Cheng, 2008). (E) Mechanism of transposed intron mobilization into the *RPL8B* locus. The transposed reporter intron within the *RPL8B* mRNA underwent a conversion to double-stranded cDNA by a reverse transcriptase followed by homologous recombination into the genome at the *RPL8B* locus.

Chapter 3. The involvement of splicing and recombination genes in intron mobilization

3.1 Background

With the detection of an intron gain event in the wild-type strain, the effect of the genes implicated in our intron gain model on intron mobilization was examined to understand the nature of the mechanism. For this study, we performed the intron mobilization assay using the same reporter system as in Chapter 2 in a series of *S. cerevisiae* knock out strains. We particularly focused on spliceosome-related genes, splicing RNA helicase genes, and genes involved in DNA recombination.

3.1.1 Spliceosome-related genes: *DBR1*, *SNU66*, and *BUD13*

Dbr1 is an enzyme that catalyzes debranching of lariat intron RNA in pre-mRNA splicing (Chapman and Boeke, 1991). Dbr1 hydrolyzes the 2'-5' phosphodiester linkage at the branchpoint adenosine of an intron RNA released from the postspliceosomal complex. The lariat-structured intron RNA is converted to a linear RNA that is subject to rapid degradation (Chapman and Boeke, 1991; Nam et al., 1994). Strains deleted of *DBR1* were shown to accumulate the intron lariat RNAs as the branched portion of RNA is protected from 3' and 5' exonucleases. Although deletion of *DBR1* showed little growth defects in *S. cerevisiae*, it resulted in severe growth defect and abnormal cell morphology in *S. pombe*, suggesting that debranching and subsequent degradation of intron RNA might play a role in cell growth (Nam et al., 1997).

Snu66 was first isolated in the affinity purified yeast U4/U6•U5 tri-snRNP complex (Gottschalk et al., 1999; Stevens and Abelson, 1999). The human orthologue of Snu66

was also shown to more stably associate with the tri-snRNP than with individual U4/U6 di-snRNP or U5 snRNP, suggesting that Snu66 might play a role in bridging U4/U6 di-snRNP and U5 snRNP (Liu et al., 2006; Makarova et al., 2001).

Yeast two-hybrid assays demonstrated that human Snu66 interacts with Prp3, a U4/U6 di-snRNP-specific protein, and Brr2 and Prp6, the U5 snRNP-specific proteins (Liu et al., 2006). Although *SNU66* is not essential, deletion of *SNU66* resulted in severe growth defect at 16°C and this cold-sensitivity correlated to a pre-mRNA splicing defect (Stevens et al., 2001).

Bud13 is a subunit of the trimeric complex, pre-mRNA Retention and Splicing (RES), along with Snu17 and Pml1 (Dziembowski et al., 2004). *BUD13* is not essential and its deletion caused slight growth defect at 30°C and the growth defect became exacerbated at 37°C. It was shown that RES is not associated with snRNPs but interacts with the pre-mRNA, suggesting that RES is transiently associated with pre-mRNA splicing prior to the first catalytic step. The RES complex was demonstrated to be participate in *in vitro* and *in vivo* pre-mRNA splicing and also for pre-mRNA retention in nucleus (Dziembowski et al., 2004). Recently, is was shown that deletion of *BUD13* specifically caused a splicing defect of the *MATa1* pre-mRNA encoding a repressor that inhibits expression of haploid-specific genes in diploid cells (Tuo et al., 2012).

3.1.2 Homologous recombination and repair genes

Homologous recombination is the exchange or transfer of DNA strands and occurs between DNA molecules with perfect or nearly identical sequence homology (San Filippo et al., 2008; Sung and Klein, 2006). Homologous recombination is a crucial process for

mitotic and meiotic recombination. It is required for the repair of DNA double-strand breaks (DSBs) and the creation of sequence variation in gametes during meiosis (San Filippo et al., 2008; Symington, 2002).

Several genes involved in homologous recombination were identified in the mutant strains which are defective in the repair of DNA damage caused by ionizing radiation. The genes involved in the repair of DSBs in the yeast are *RAD50*, *RAD51*, *RAD52*, *RAD54*, *RAD55* and *RAD57*, which belong to the *RAD52* epistasis group (Malone et al., 1988). Mutations in any gene of the *RAD52* epistasis group result in DNA-damage sensitivity and defective homologous recombination (Symington, 2002).

RAD51 is required for the genetic recombination and the repair of X-ray-induced DNA damage (Symington, 2002). It was shown that *S. cerevisiae RAD51* has significant amino acid sequence homology to the bacterial recombinase protein, RecA (Basile et al., 1992). The conserved sequences are crucial for the recombinase activity, which mediates the homologous DNA paring and strand exchange during homologous recombination (Symington, 2002). Like RecA, Rad51 polymerizes onto the single-stranded DNA (ssDNA) at DSBs to form a helical protein filament, the presynaptic filament. The synaptic filament is formed when the presynaptic filament binds to the double-stranded DNA to seek sequence homology for recombination (San Filippo et al., 2008).

Rad52, Rad55 and Rad57 function as a recombination mediator (Sung and Klein, 2006). Rad52 is implicated in all known homologous recombination pathways including both Rad51-dependent reactions and Rad51-independent reactions (Symington, 2002). Therefore, the *rad52* mutants exhibit the most severe defects to the DNA-damaging agents. Physical interaction between Rad52 and Rad51 was experimentally

demonstrated and the recruitment of Rad51 to DSBs was shown to be dependent on Rad52 in both mitotic and meiotic cells, suggesting that Rad52 plays a role in loading Rad51 to the ssDNA substrate during homologous recombination (Milne and Weaver, 1993; Shinohara et al., 1992).

Rad55 and Rad57 form a heterodimeric complex (Sung, 1997). Like Rad52, the Rad55–Rad57 complex also helps the assembly of presynaptic filament through the physical interaction with Rad51 (Hays et al., 1995; Sung, 1997). Additionally, it was suggested that the Rad55-Rad57 complex might contribute to stabilizing the assembled presynaptic filament (Fortin and Symington, 2002). Rad55 and Rad57 proteins also have sequence homology to RecA and Rad51 but do not possess the recombinase activity (Sung, 1997).

In addition to the *RAD52* epistasis group, *RAD1* is also involved in DSB-induced recombination (Ivanov and Haber, 1995). *RAD1* belongs to the *RAD3* epistasis group that is required for nucleotide excision repair (NER) of UV-damaged DNA in *S. cerevisiae* (Ivanov and Haber, 1995). Rad1 exhibits single-strand DNA endonuclease activity as a complex with Rad10 in the nucleotide excision repair mechanism (Tomkinson et al., 1993). The Rad1/Rad10 complex cleaves 3' non-homologous DNA tail from the recombination intermediates in homologous recombination (Davies et al., 1995; Li et al., 2013).

3.1.3 RNA helicase genes: *PRP16*, *PRP22*, and *PRP43*

RNA helicases are the enzymes that unwind double-stranded RNA (dsRNA) utilizing the energy released from NTP hydrolysis (de la Cruz et al., 1999). In addition to dsRNA, RNA helicases also regulate RNA-protein complexes by stabilizing or

destabilizing their interactions in numerous gene expression processes including transcription, pre-mRNA splicing, nuclear export, and translation (Cordin et al., 2006; Tanner and Linder, 2001).

For pre-mRNA splicing, eight RNA helicases are required to rearrange RNA-protein and RNA-RNA interactions in formation, activation, and disassociation of the spliceosome (Fig. 1.2) (Cordin and Beggs, 2013). All these RNA helicases are members of the superfamily 2 (SF2) of helicases. They share two RecA-like domains and conserved motifs for NTP binding and hydrolysis. The RNA helicases involved in splicing belong to three families: Prp5, Sub2, Prp28 (DEAD-box family), Brr2 (Ski-2 family), and Prp2, Prp16, Prp22, and Prp43 (DEAH-box family) (Cordin and Beggs, 2013). Here we focused on the three RNA helicases that act in the later stages of splicing (Fig. 1.2), Prp16, Prp22, and Prp43 for the intron mobilization assay.

Prp16 induces conformational rearrangements for the second step of splicing using its ATPase activity (Cordin and Beggs, 2013). It was shown that Prp16 mediates the release of Cwc25 and Yju2 from the spliceosome to allow the access of the splicing factors required for the second step of splicing including Prp22, Slu7, and Prp18. At this point, Prp22 also contributes to facilitate the second step by stabilizing the binding of Slu7 in an ATP-independent manner (James et al., 2002).

Prp22 was first characterized to mediate ATP-dependent mRNA release from the spliceosome (Company et al., 1991), however, Prp22 is required only when the branchpoint sequence and the 3' splice site are distant longer than 21 nucleotides (Schwer and Gross, 1998). When the distance is closer, the deletion of *PRP22* did not block the second catalytic step (Schwer and Gross, 1998). As described previously, the

prp22S635A mutant catalyzes the second transesterification reaction but fails to release the mRNA and was employed to demonstrate the reversibility of pre-mRNA splicing (Tseng and Cheng, 2008), suggesting that *prp22* mutant strains are promising targets for this intron mobilization assay.

Prp43 is required for the release of the excised intron and the disassembly of the post-spliceosomal complex (Arenas and Abelson, 1997; Martin et al., 2002). Prp43 also conducts the proofreading function by dissociating the stalled spliceosomal complex and discarding the suboptimal pre-mRNAs (Koodathingal et al., 2010; Mayas et al., 2010). Like Prp22, mutants of Prp43 are considered to be encouraging candidates for enhancing the reverse splicing reaction as it remains associated with the excised lariat intron, the proposed substrate for intron transposition. The fact that the spliceosomal complex associated with Prp43 has released the mRNA makes the *prp43* mutant a better target for intron mobilization as the probability of reinsertion into the corresponding mRNA is expected to be reduced (Fig. 3.1).

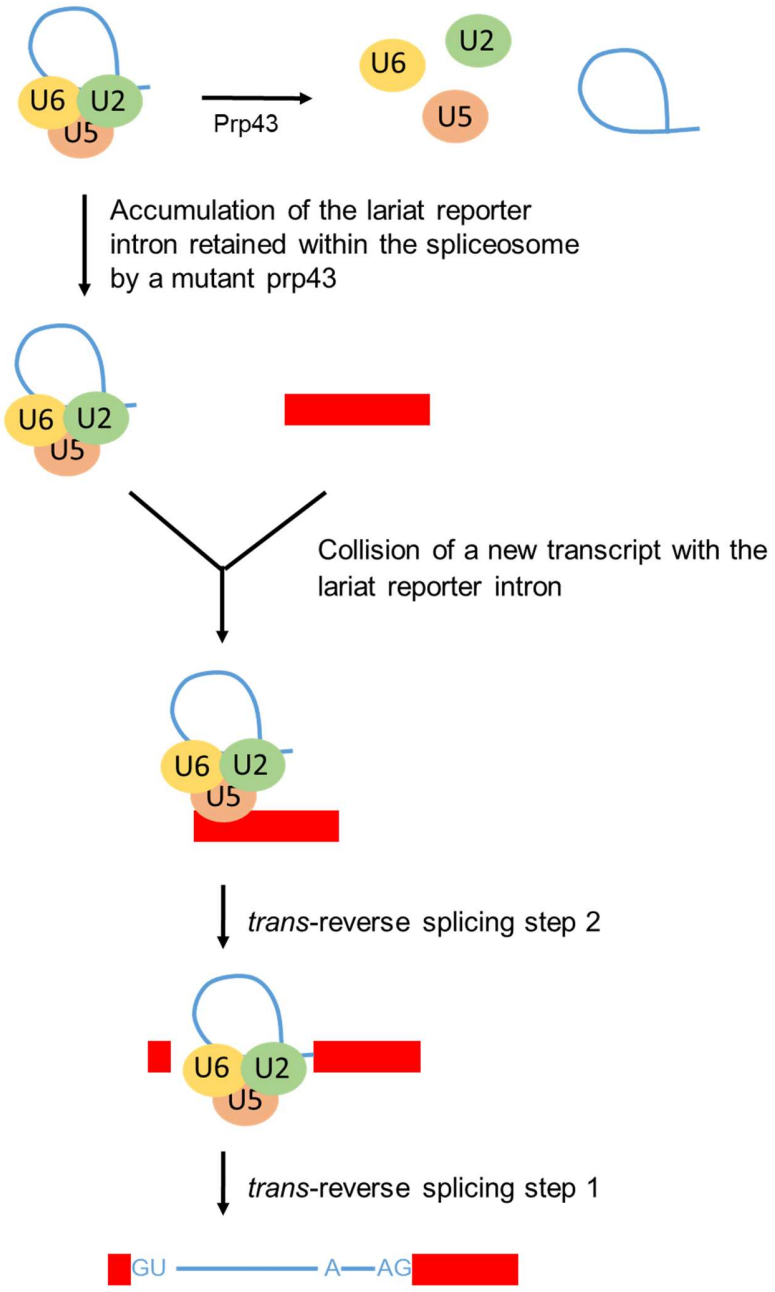

Figure 3.1 Proposed model of *trans*-reverse splicing facilitated by a mutant prp43

The lariat reporter intron associating with the residual spliceosome after Prp22-mediated mRNA release was accumulated by a mutant prp43, providing a higher chance of colliding with a new transcript leading to *trans*-reverse splicing.

3.2 Materials and methods

3.2.1 Plasmids

The pMPY-URA3 (pMPY-ZAP) was used to knock out *TRP1* gene *via* removable *URA3* (Schneider et al., 1996). The pFA6a-KanMX6 and pRS41H-TAP were used for the amplification of KAN^R and HYG^R gene. *prp43S247A* was amplified from pSE358-prp43S247A and subcloned into pRS415 with the use of SacI-HF and XmaI.

3.2.2 Yeast strains

Yeast strains used in this chapter are listed in Table 3.1. For the essential genes, *PRP16*, *PRP22*, and *PRP43*, the cold sensitive mutant strains were employed: *prp16R456K*, *prp22D613N*, *prp43S247A*, and *prp43Q423N*. These mutants were shown to arrest the spliceosome at the specific stage each gene product acts during splicing cycle (Fig. 1.2) (Lardelli et al., 2010). Deletion strains of *BUD13*, *RAD1*, *RAD51*, *RAD52*, *RAD55*, and *RAD57* are from the Yeast MATα Knock Out collection (GE Dharmacon).

3.2.3 Construction of strains

3.2.3.1 *prp16 and prp22* cold-sensitive mutant strains

Isogenic strains of *prp16R456K* and *prp22D613N* were generated by modifying BY4733 strain. *URA3* gene in SS1316 strain harboring *prp16R456K-TAP::URA3* was first replaced with KAN^R gene (*prp16R456K-TAP::KAN*) for screening purpose using a primer pair of KlURA3->KANMX6F and KlURA3->KANMX6R. The DNA fragments containing each mutation, *prp16R456K-TAP::KAN* and *prp22D613N-TAP::KAN* by PCR using

Phusion® High-Fidelity DNA Polymerase (New England Biolabs), respectively. The primer pairs of PRP16-928FWD / PRP16-TAP-KANR and PRP22-698-FWD / PRP22-3'-REV were used for amplification. BY4733 strain was subsequently transformed with these DNA fragments, separately, to replace the wild-type *PRP16* and *PRP22* with the corresponding mutant by homologous recombination. Transformation was conducted as described in 2.2.3 except the selection was achieved by plating on YPD containing 200 ug/ml of G418. Strains were confirmed by cold-sensitive phenotype and DNA sequencing.

3.2.3.2 *prp43* cold-sensitive mutant strains

SS5247 strain harboring *prp43Q423N* mutant was generated by knocking out the *TRP1* gene from SS5246 (JPS575, J. Staley, University of Chicago) for screening purpose. *TRP1* was deleted using the MPY-URA3 replacement method (Schneider et al., 1996). The MPY-URA3 DNA fragment was amplified with the primers targeting *TRP1* (TRP-MPYup and TRP-MPYdown) and introduced into SS5246 strain. *Ura+* transformants were patched on YPD and then streaked onto synthetic complete plates with 5-FOA to select for cells that had looped out the incorporated *URA3* in place of *TRP1* gene.

prp43S247A was subcloned into pRS415 for screening purpose. BY4734 strain was transformed with pRS415-prp43S247A and then the chromosomal *PRP43* was replaced with KAN^R gene by homologous recombination.

3.2.3.3 *prp43* cold-sensitive diploid strain

The *LYS2* gene was knocked out from SS5247 strain with HYG^R gene for diploid

selection purpose. *Prp43* cold-sensitive diploid strain was obtained by mating *prp43S247A* and *prp43Q423N* strains and selecting diploid on SD-lys with 300 ug/ml of Hygromycin (Invitrogen) and 200 ug/ml of G418. For the double selection media, ammonium sulfate is omitted and glutamic acid (1 g/L) is added instead.

3.2.3.4 Wild-type diploid strain

The *Lys2* gene in BY4733 and BY4734 strains was knocked out with HYG^R gene and KAN^R gene respectively by homologous recombination for diploid selection purpose. BY4733 and BY4734 strains with a different selectable marker gene were mated and the diploid was selected by streaking on YPD containing 300 ug/ml of Hygromycin and 200 ug/ml of G418.

Table 3.1 Yeast strains used in Chapter 3.

Strain	Mating type	Genotype
BY4733	a	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0*
BY4734	α	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0*
SS1316	a	*ade2Δ, his3Δ, leu2Δ, lys2Δ, trp1Δ, ura3Δ, prp16::LYS2* pSE358-*prp16R456K-TAP::URA3*
SS1330	a	*gal-, mal-, his4-619, prp22D613Q-TAP::KAN*
SS5241	a	*his3Δ200, trp1Δ63, leu2Δ0, met15Δ0,ura3Δ0, prp16R456K-TAP::KAN*
SS5240	a	*his3Δ200, trp1Δ63, leu2Δ0, met15Δ0, ura3Δ0, prp22D613N-TAP::KAN*
SS3098	a	*his3Δ,leu2Δ,met15Δ,ura3Δ, prp43::HIS3* pJPS1163-*prp43Q423N-TAP::KANMX Brr2-cHP::LEU*
SS5242	a	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, lys2Δ::HYG*
SS5243	a	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, lys2Δ::KANMX6*
SS5244	α	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, lys2Δ::HYG*
SS5245	α	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, lys2Δ::KANMX*
SS5246	a	*his3Δ200, leu2Δ0, met15Δ0, ura3Δ0, prp43::KANMX4* + pJPS643-*prp43Q423N*
SS5247	a	*his3Δ200, leu2Δ0, met15Δ0, ura3Δ0, trp1Δ, prp43::KANMX4* + pJPS643-*prp43Q423N*

(Table 3.1 continued)

SS5248	a / α	SS5242 X SS5245
SS5249	a / α	SS5243 X SS5244
SS5250	α	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, prp43::KANMX* + pRS415-*prp43S247A*
SS5251	a	*his3Δ200, leu2Δ0, met15Δ0, ura3Δ0, trp1Δ, lys2Δ::HYG, prp43::KANMX4* + pJPS643-*prp43Q423N*
SS5252	a / α	SS5250 X SS5251
SS1036	α	*his3Δ200, leu2Δ0, met15Δ0, trp1Δ63, ura3Δ0, snu66Δ*
SS4058	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, dbr1ΔKanMX*
bud13Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, bud13ΔKanMX*
rad1Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, rad1ΔKanMX*
rad51Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, rad51ΔKanMX*
rad52Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, rad52ΔKanMX*
rad55Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, rad55ΔKanMX*
rad57Δ	α	*his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, rad57ΔKanMX*

3.2.4 Oligonucleotides

Table 3.2 DNA Oligonucleotides used in Chapter 3.

Name	Sequence (5' to 3')
PRP16-928FWD	GGGGAAGTACTAGGCTTGGAG
PRP16-TAP-KANR	ATGCATATAACTATATAATAACATATATGAATATTT TGCCGCGGCGTTAGTATCGAATCG
PRP43S247AFWD	TTGAAGATAATTATTATGGCTGCTACTCTGGAGCA GAG
PRP22-698-FWD	AAAGGCGTGCTTTGACTTCACCAG
PRP22FWD1735	ATGACTGATGGTATGTTACAGAG
PRP22-3'-REV	CATAGTCCAGAAACGCCACC
PRP43-513-FWD	GTTGGTTATTCTATCAGATTCG
PRP43FWD1085	AGTCACATAATGGTAGACCAGG
JPS-PRP43-TAPR	TTTATATAAATCTATTTTTTTTTTTTTTTCGACACAAA ATGTTACGACTCACTATAGGGA
KlURA3->KANMX6F	TTATCGGGGTATATGATAATTAATCTTACCAATTGG ATGATTTAGCTTGCCTCGTCCCCG
KlURA3->KANMX6R	CCCGGAGACAATCATATGGGAGAAGCAATTGGAA GATAGATGGATGGCGGCGTTAGTATC
TRP-MPYup	TATTGAGCACGTGAGTATACGTGATTAAGCACACA AAGGCAGCTTGGAGTAGGGAACAAAAGCTGG
TRP-MPYdown	TGCAGGCAAGTGCACAAACAATACTTAAATAAATAC TACTCAGTAATAACCTATAGGGCGAATTGG

(Table 3.2 continued)

newPRP43_PCR5	CGAACAGGTCCCAATAAGGATGC
newPRP43_PCR3	TGGGAAGCCAACATCCTCG
LYS2::KAN-F1	AACTGCTAATTATAGAGAGATATCACAGAGTTACTC ACTACGGATCCCCGGGTTAATTAA
LYS2::KAN-R1	TAATTATTGTACATGGACATATCATACGTAATGCTC AACCGAATTCGAGCTCGTTTAAAC
LYS2::HYG-F1	AACTGCTAATTATAGAGAGATATCACAGAGTTACTC ACTAGACATGGAGGCCCAGAATAC
LYS2::HYG-R1	TAATTATTGTACATGGACATATCATACGTAATGCTC AAC CACGCATCTGTGCGGTATTTC

3.2.5 Yeast transformation with the reporter plasmid

Each mutant strain was transformed with the pRS316HIS5AI reporter as described in 2.2.3.

3.2.6 Intron transposition assay

Intron transposition assay was performed as described in 2.2.4 with modifications for the cold-sensitive mutant strains. For all deletion strains, individual *ura+* prototrophs containing the reporter pRS316HIS5AI were grown in 10 ml SD-ura liquid medium overnight at 31°C. For *prp16*, *prp22*, *prp43*, *snu66*Δ, and *rad52*Δ mutant strains, the temperature for the liquid culture was adjusted to 20-25 °C to induce the cold-sensitive phenotype.

3.2.7 Rate determination of *his+* prototroph formation

Rate of *his+* prototroph formation in each mutant strain was determined by normalizing the number of *his+* prototrophs to the number of total cells tested.

3.2.8 Diagnostic PCR and inverse PCR

Diagnostic PCR to determine the chromosomal integration and inverse PCR to analyze the integration sites were conducted as described in 2.2.5 and 2.2.7.

3.3 Results

3.3.1 Formation of *his*+ prototroph is rare in *dbr1*Δ.

It was previously mentioned that the *dbr1* deletion strain is viable and exhibits accumulation of excised lariat introns at high levels with little growth defect in *S. cerevisiae* (Chapman and Boeke, 1991). Thus, if the free intron lariats are the substrates for intron mobilization, this would be the ideal strain in which to perform these experiments. The strains harboring pRS416, pRS316HIS5AI, and pRS316HIS5AIΔbp reporter plasmid respectively were grown overnight at 31°C to induce intron transposition. As expected, no *his*+ prototroph arose with the negative control reporter, HIS5AIΔbp, as well as the empty vector (Table 3.3). With the HIS5AI reporter, the *his*+ prototroph formation rate was decreased by 18 times compared to the wild-type, disproving the model of free intron lariat being the substrate for intron mobilization including RNA-mediated intron loss event (Fig. 3.2).

3.3.2 *his*+ prototroph formation increases in the splicing defective mutants, *bud13*Δ and *snu66*Δ.

As the formation of *his*+ prototroph using our reporter in the wild-type cells is dependent on the splicing of the indicator gene, HIS5AI, we wanted to examine how the defects in splicing affect the intron mobilization. For this goal, we employed two splicing defective mutants, *bud13*Δ and *snu66*Δ. The intron transposition assay was performed at 31°C which induces a slight growth defect only in *bud13*Δ strain. We observed an increase of *his*+ formation in both splicing defective mutants (Table 3.4, Fig. 3.2). *his*+

prototroph formation rates in *bud13*Δ and *snu66*Δ are 4 and 1.9 times higher than the rate of the wild-type strain tested at the same temperature, respectively (Fig. 3.2). These results agree with our hypothesis that the splicing intermediates detained and accumulated by splicing defects would provide a greater opportunity for intron mobilization although more splicing defective mutants need to be tested to make a clearer conclusion.

3.3.3 The effect of defective recombination on *his*+ prototroph formation

As most of the existing models for intron gain and loss employ homologous recombination to explain the mechanism, we examined how known recombination genes of different classes affect intron mobilization in our study. We performed the intron transposition assay in the deletion strains of *RAD1*, *RAD51*, *RAD52*, *RAD55*, and *RAD57* and investigated how the formation of *his*+ is affected. The statistical information of the recombination defective mutant strains is in Table 3.5 and Fig. 3.2. We observed overall decrease in the rate of appearance of *his*+ prototroph by deletion of each recombination gene as predicted. While the deletion of *RAD1* and *RAD51* showed only slight decrease, the formation of *his*+ prototrophs was reduced by 4 to 5 times in *rad57*Δ and *rad55*Δ strains compared to the wild-type. Interestingly, the deletion of *RAD52* completely eliminated *his*+ colony formation. As the *RAD52* gene is involved in most recombination events in *S. cerevisiae* (Malone et al., 1988), this result supports that the *his*+ formation, either *via* plasmid event or chromosomal event, would rely on *RAD52*-dependent recombination.

Table 3.3 *his+* prototroph formation in *dbr1Δ.*

Construct	Total cells tested	*his+*
pRS416	0.93 X 10^{10}	0
pRS316HIS5AI	1.02 X 10^{10}	26
pRS316HIS5AIΔbp	0.89 X 10^{10}	0

Table 3.4 *his+* prototroph formation in the splicing defective mutants.

Strain	Construct	Total cells tested	*his+*
bud13Δ	pRS416	4.27 X 10^{9}	0
	pRS316HIS5AI	4.53 X 10^{9}	802
	pRS316HIS5AIΔbp	5.01 X 10^{9}	0
snu66Δ	pRS416	2.91 X 10^{9}	0
	pRS316HIS5AI	3.59 X 10^{9}	298
	pRS316HIS5AIΔbp	3.47 X 10^{9}	0

Table 3.5 *his+* prototroph formation in the recombination defective mutants.

Strain	Construct	Total cells tested	*his+*
*rad1*Δ	pRS416	3.04 X 10^9	0
	pRS316HIS5AI	3.68 X 10^9	145
	pRS316HIS5AIΔbp	3.47 X 10^9	0
*rad51*Δ	pRS416	3.55 X 10^9	0
	pRS316HIS5AI	5.66 X 10^9	185
	pRS316HIS5AIΔbp	6.23 X 10^9	0
*rad52*Δ	pRS416	5.07 X 10^9	0
	pRS316HIS5AI	5.65 X 10^9	0
	pRS316HIS5AIΔbp	6.16 X 10^9	0
*rad55*Δ	pRS416	3.06 X 10^9	0
	pRS316HIS5AI	3.81 X 10^9	32
	pRS316HIS5AIΔbp	3.94 X 10^9	0
*rad57*Δ	pRS416	3.80 X 10^9	0
	pRS316HIS5AI	2.74 X 10^9	28
	pRS316HIS5AIΔbp	2.56 X 10^9	0

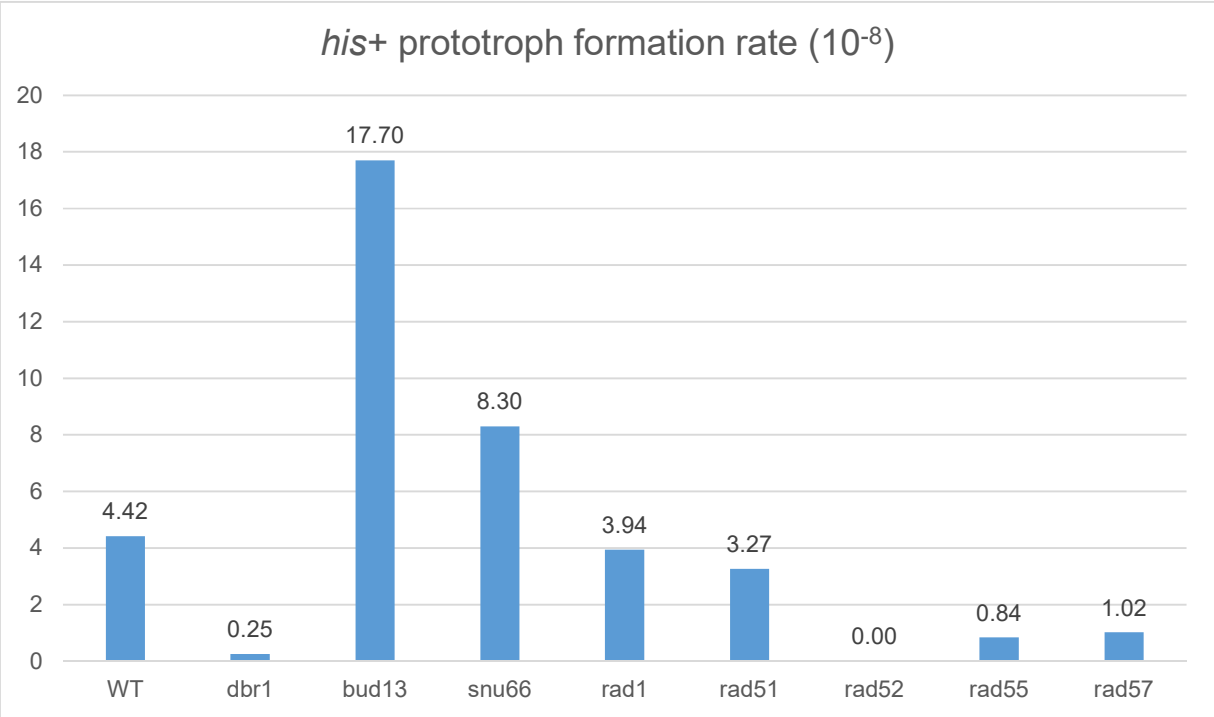

Figure 3.2 Rate of *his*+ prototroph formation in the wild-type, the splicing defective mutants, and the recombination defective mutants

WT: 4.42, *dbr1*Δ: 0.25, *bud13*Δ: 17.70, *snu66*Δ: 8.30, *rad1*Δ: 3.94, *rad51*Δ: 3.27, *rad52*Δ: 0, *rad55*Δ: 0.84, *rad57*Δ: 1.02 ($X10^{-8}$)

3.3.4 The effect of splicing helicase mutants in *his+* prototroph formation

As our intron gain model involves the reversal of splicing reaction, we wanted to investigate how splicing helicase mutants affect the rate of the overall *his+* formation and particularly the frequency of chromosomal integration of intron RNA. We focused on three RNA helicases acting in the two catalytic steps and the post-catalytic step during splicing, Prp16, Prp22, and Prp43. As these are essential, we employed cold-sensitive conditional mutants and performed the intron transposition assay at lowered temperature ranging from 20-25°C (Table 3.6, Fig. 3.3). The mutant *prp16R456K* did not show significant difference in the rate of *his+* formation compared to the wild-type as predicted because Prp16 acts before the second catalytic step of splicing. At this point, the lariat intron RNA is still attached to the 3' exon. Thus, intron mobilization is less likely to occur. The *prp22D613N* strain exhibited 2.2-fold increase in the rate of *his+* formation relative to the wild-type, indicating that the intron mobilization occurs more actively with the spliceosome associating with both the mRNA and the lariat intron. Although, hypothetically, this mutant is most likely to experience the reverse splicing of intron RNA back into the original position, the exact mechanism is not understood yet (Fig. 3.4).

We employed two *prp43* mutant strains, *prp43S247A* and *prp43Q423N*. There was difficulty generating isogenic strains with chromosomal mutations, thus we introduced the plasmid harboring each mutant into the wild-type strain and then knocked out the genomic copy of *PRP43*. The *prp43S247A* mutant showed a highly decreased rate of *his+* formation compared to the wild-type strain while the *prp43Q423N* mutant did not show any change (Fig 3.3). The reduction in *prp43S247A* mutant was more obvious relative to *prp22D613N* (Fig 3.3).

Table 3.6 *his+* prototroph formation in the splicing helicase mutants.

Strain	Temperature	Total cells	*his+*	*his+* FOA^r
WT	25 °C	0.86 X 10^{10}	200	69
prp16R456K	23 °C	3.84 X 10^{10}	1068	17
prp22D613N	23 °C	2.29 X 10^{10}	1172	201
prp43S247A	23 °C	4.31 X 10^{10}	13	1
prp43Q423N	20 °C	0.21 X 10^{10}	23	4
prp43Q423N	25 °C	4.22 X 10^{10}	984	25

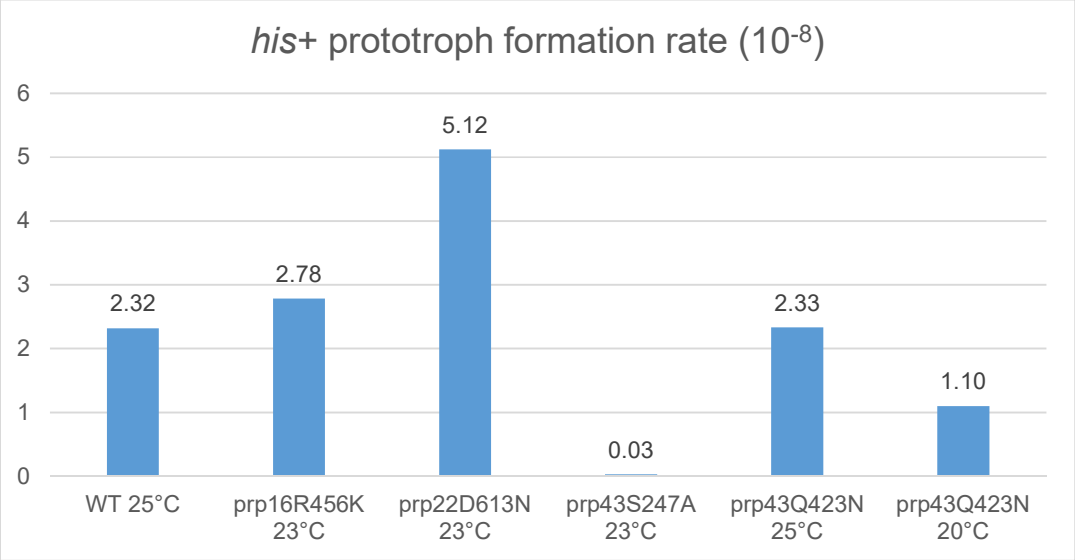

Figure 3.3 Rate of *his+* prototroph formation in the splicing helicase mutants

WT: 2.32, *prp16R456K*: 2.78, *prp22D613N*: 5.12, *prp43S247A*: 0.03, *prp43Q423N* at 25 °C: 2.33, *prp43Q423N* at 20 °C: 1.10 (X 10^{-8})

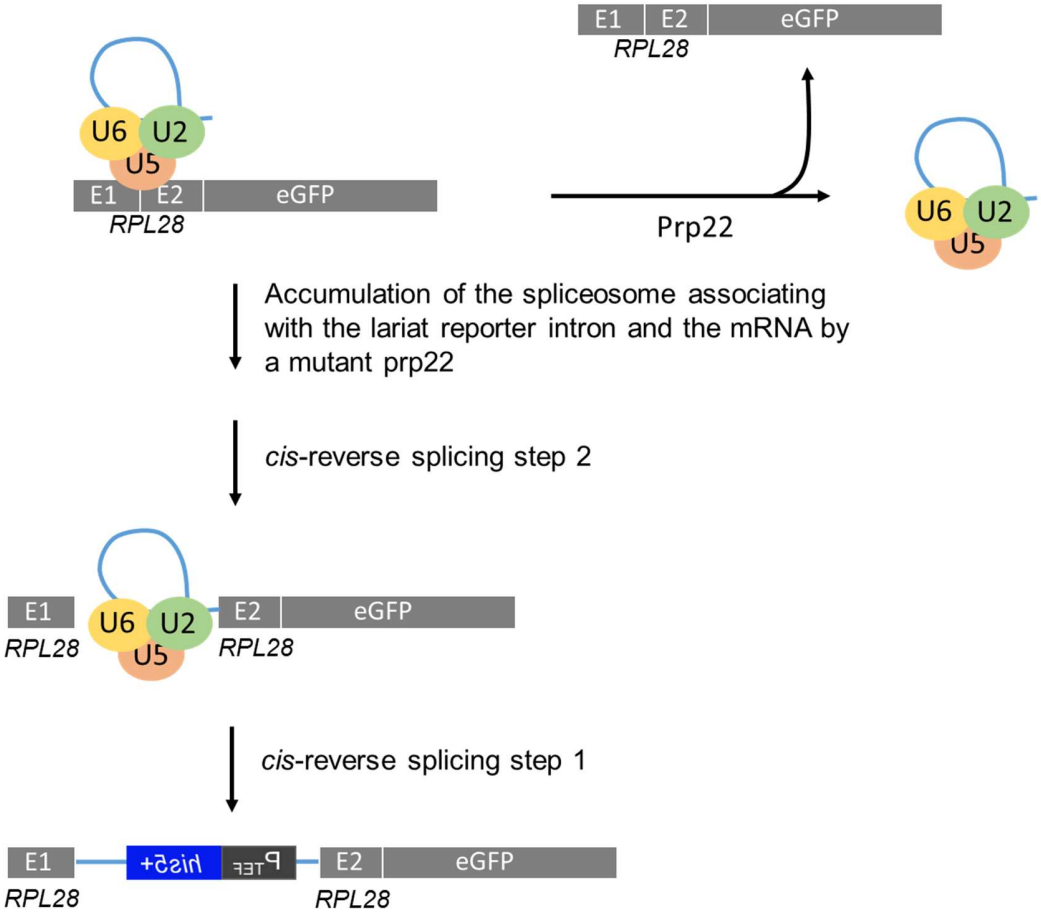

Figure 3.4 Proposed model of *cis*-reverse splicing facilitated by a mutant prp22

After 2nd splicing step, the spliceosome was associated with the mRNA and the lariat intron. This spliceosome intermediate was accumulated by a mutant prp22, leading to *cis*-reverse splicing by which the lariat intron RNA was reintegrated into the mRNA.

3.3.5 Formation of *his+* is inhibited in the diploid strains.

Diploid strains of the wild-type and *prp43* mutant were generated individually to enhance the possibility of a chromosomal integration of intron RNA within essential genes. This event would cause death in haploid cells as the splicing efficiency of the intron is too low due to its large size (discussed in Section 2.3.6). The *his+* formation in both diploid strains was unexpectedly very rare (Table 3.7), particularly in case of the wild-type compared to the corresponding haploid strain, suggesting that the mechanism leading to the intron mobilization is inhibited by one or more factors specific to the diploid strain.

3.3.6 Chromosomal events result in pseudogene formation, not intron gain.

Although a number of *his+* 5-FOA^r colonies were obtained in the mutant strains, most of them turned out to be plasmid-related events mediated by non-homologous recombination (Fig. 3.5). We performed PCR analysis to distinguish the chromosomal events from the plasmid-related events, and only a small number of chromosomal events were detected in the mutant strains. Interestingly, nearly all chromosomal events were associated with Ty transposable elements and only a part of the reporter intron was integrated, that is, the splicing signals at 5' splice site and 3' splice site were not retained. This resulted in pseudogene formation, but not intron gain. These results suggest that Ty transposable elements are implicated in intron mobilization in addition to providing the function of reverse transcriptase (Derr and Strathern, 1993; Derr et al., 1991).

As expected from the *his+* prototroph phenotype, the chromosomally inserted intron portions kept the minimal *S. pombe tef1+* promoter for *his5+* transcription, the *his5+* gene that had lost the artificial intron, and the *tef1+* terminator, however, the reporter

intron sequences flanking these elements were variably maintained. The schematic diagrams of chromosomal insertion events detected are in Fig. 3.6. All of these events led to pseudogene formation.

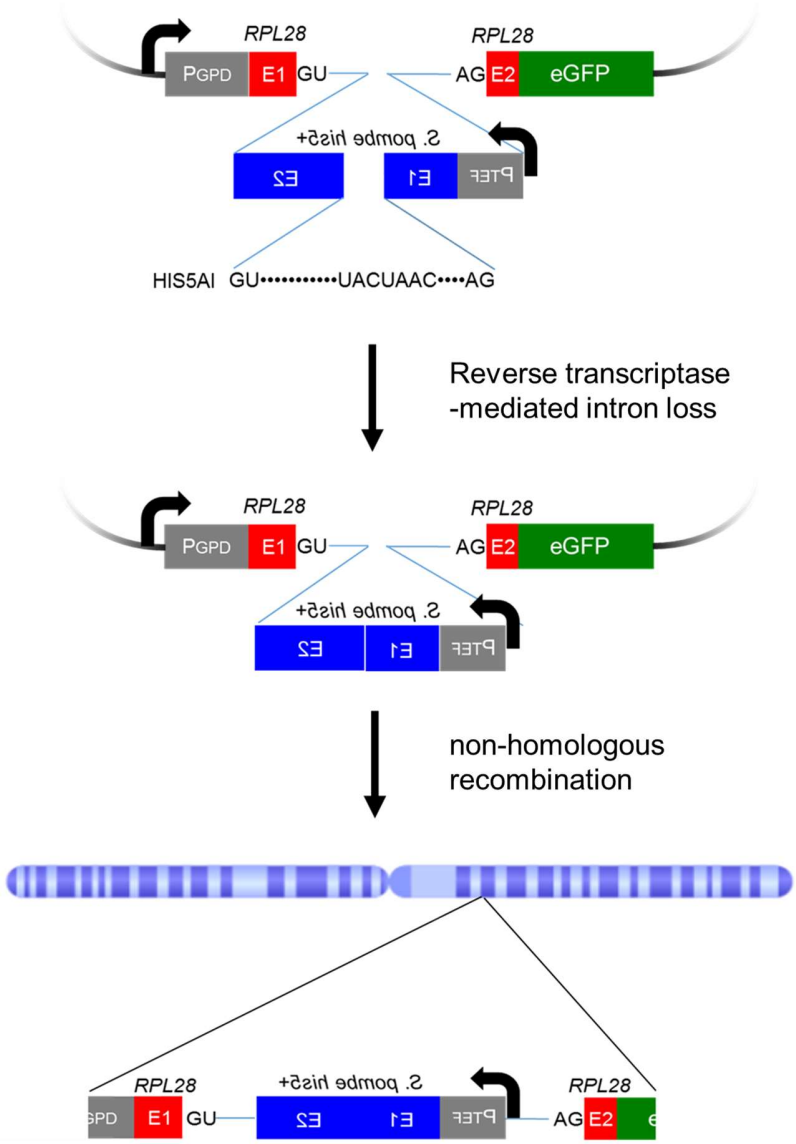

Figure 3.5 Proposed model of the plasmid-related events leading to the pseudogene additions

The reporter plasmid that has undergone an intron loss event was targeted for non-homologous recombination with the genomic DNA, resulting in that most genomic events were pseudogene additions.

Table 3.7 *his*+ prototroph formation in the diploid strains.

Strain	Temperature	Total cells	*his*+	*his*+ FOA^r
WT diploid	31 °C	4.63 X 10^{10}	3	0
prp43 diploid	23 °C	3.16 X 10^{10}	7	2

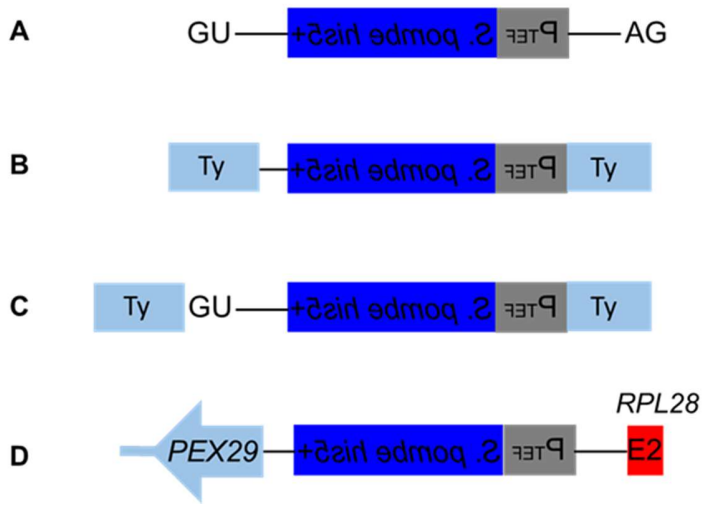

Figure 3.6 Chromosomal integration of the reporter intron resulted in pseudogene formation.

(A) The entire reporter intron RNA (B) Chromosomal integration of the partial reporter intron attached to Ty transposable elements at both junctions. The splicing signals are not maintained. (C) Same as (B) with the 5' splice site conserved. (D) Disruption of *PEX29* by chromosomal integration of the partial reporter intron. The insertion occurred at 224 nt from the start codon of *PEX29* (total length 1665 bp).

3.4 Discussion

We investigated the involvement of known splicing and recombination genes in the intron mobilization using the reporter detecting intron gain and loss. A series of mutant strains that are defective in the splicing and recombination were tested as our model of intron gain involves intron transposition *via* reverse splicing and homologous recombination. In our model of intron gain, the excised intron RNA needs to be reverse transcribed after reverse splicing. Reverse transcription is also required for the RNA-mediated intron loss mechanism. The most promising source of reverse transcriptase activity is derived from Ty1 transposable elements. Interestingly, our mutant analyses suggest that Ty transposable elements might be involved in the overall intron mobilization process, although we did not detect any additional intron gain event.

The rare appearance of *his+* prototroph in the *dbr1* deletion strain even with accumulated intron RNA population suggests that free intron RNA disassociated from the spliceosome may not be a suitable substrate for intron mobilization and that Ty transposable elements could be implicated in the intron mobilization leading to *his+* prototroph formation. It was shown that the frequency of Ty transposition is reduced in a *dbr1* mutant (Chapman and Boeke, 1991) and the branching and debranching might function in Ty reverse transcription (Cheng and Menees, 2004) although the involvement of Dbr1 in Ty transposition is not clearly understood yet. *S. cerevisiae* Ty1 replicates through reverse transcription of an RNA intermediate and the integration of the resulting cDNA into the genome. The chromosomal integration occurs by either homologous recombination or Ty integrase (Garfinkel et al., 2005).

The rarity of *his*+ prototroph formation in the wild-type diploid strains also supports the idea that Ty transposable elements might mediate the intron mobilization. We reasoned that the intron transposition is more easily detected in the diploid strains because the integration of the reporter intron could lead to death in cases of insertion into essential genes due to the low splicing efficiency of this reporter. However, our results indicated that *his*+ formation was strongly inhibited in the wild-type diploid strain. This is likely due to the inhibition of Ty RNA transcription in diploid strains (Elder et al., 1981), further evidence for a role of Ty transposable elements in intron mobilization.

The results from the recombination defective mutants also allowed us to consider the connection between Ty transposable elements and intron mobilization. Recombination defective mutants of the *RAD52* epistasis group including *RAD51*, *RAD52*, *RAD55*, and *RAD57* were shown to cause increased levels of Ty1 transposition (Curcio and Garfinkel, 1994; Derr, 1998; Rattray et al., 2000). It was also demonstrated that the deletion of *RAD52* does not change the overall rate of *his*+ formation in RNA-mediated gene conversion (Derr and Strathern, 1993). In our study, the rate of *his*+ prototroph formation was increased by deletion of *RAD1* and *RAD51* but it was decreased with the deletion of *RAD55* and *RAD57.* The most interesting result was that *his*+ formation is highly inhibited in *rad52* deletion strain as opposed to the previous study (Curcio and Garfinkel, 1994; Rattray et al., 2000). Together these results suggest a possibility that the intron mobilization might be mediated by *RAD52*-dependent recombination pathway which is independent from the Ty-mediated pathway and one mechanism is preferred under certain circumstances.

Additionally, we examined the effect of the defective splicing helicases on intron gain based on our hypothesis that reverse splicing is more likely to occur when the spliceosomes associating with the excised intron RNA are slowed and stalled. Although we did not detect any intron gain event in the splicing helicase mutants, it was interesting to observe the significant reduction of *his*+ formation in *prp43S247A* mutant relative to *prp22D613N* mutant. As nearly all the *his*+ isolates represented a plasmid-borne event that had undergone RNA-mediated intron loss, the increased level of *his*+ formation in *prp22* compared to the wild-type led us to consider a possibility that reverse splicing might be involved in this intron loss event. In this model, the reporter intron that has lost the artificial intron by recursive splicing remains within the spliceosome slowed by *prp22* and reverse splices back to the original position (Fig. 3.4). And the resulting pre-mRNA without the artificial intron is released from the spliceosome, reverse transcribed and then integrated into the plasmid by homologous recombination, leading to intron loss. This hypothesis is only partially and indirectly supported in that the smaller artificial intron is more efficiently spliced at lowered temperature, and that reverse splicing of intron RNA back into the corresponding mRNA was experimentally demonstrated *in vitro* using a *prp22* mutant (Tseng and Cheng, 2008).

Altogether, we detected intron loss and pseudogene formation involving Ty transposable elements in the splicing and recombination defective mutant strains using the intron gain and loss reporter system. Determination of *his*+ formation rate allowed us to understand how these cellular genes are implicated in the intron mobilization in *S. cerevisiae* and suggested that there might be independent mechanisms for intron mobilization mediated by *RAD52* and Ty transposable element.

Chapter 4. Significance and future directions

4.1 Significance

While our understanding of the structural and mechanistic features of spliceosomal introns and spliceosome have been under study since 1977, the most fundamental areas remain largely unproven. What is the origin of spliceosomal introns and how have they been propagated throughout evolution leading to the lineage-specific distribution of intron loss and gain? Most previous studies to answer these questions are based on phylogenetic analyses (Irimia and Roy, 2014; Yenerall and Zhou, 2012). Although intron loss has been demonstrated experimentally (Curcio and Garfinkel, 1991; Derr et al., 1991), several mysteries remain regarding the origin and propagation of introns.

We designed a novel reporter-based strategy that selects for both intron loss and intron gain events and experimentally verified the first demonstration of intron gain by intron transposition in any organism. The yeast gene *RPL8B* was shown to have a selected, perfectly transposed intron added into its coding sequence. We demonstrated Rpl8 expression from the novel *RPL8B* allele indicating that the newly gained intron is able to be correctly removed by spliceosome. We also showed that when overexpressed, this novel allele is functional in a strain lacking the Rpl8 paralogue *RPL8A* demonstrating that the gene targeted for intronogenesis is functional. The finding of intron transposition leading to intron gain provides evidence for the reversibility of pre-mRNA splicing and leads to a model for spliceosomal intron propagation and homeostasis.

We also examined the involvement of the known cellular genes in the intron transposition using the intron gain and loss reporter system to address the mechanism of

the intron gain and loss events we captured. Although we have not detected additional intron gain events yet, we observed a number of RNA mobilization events including the intron loss and the integration of transcripts derived from the reporter into the genome in a series of splicing and recombination defective mutant strains. These results suggest that this phenomenon is spliceosome-mediated and the RNA-mediated recombination is dependent on the *RAD52* gene product and Ty transposable elements.

Given the scarcity of direct evidence supporting current intron propagation models and the limited evidence for *in vivo* reverse splicing, our demonstration of the intron transposition proves ongoing spliceosomal intronogenesis and leads to greater confidence in the reversibility of pre-mRNA splicing *in vivo*.

4.2 Future directions

This dissertation highlights the development and validation of a reporter system as a tool of detecting intron gain and loss events *in vivo* in *S. cerevisiae* and helps to identify cellular genes implicated in the intron mobilization and integration into the genome, however, further studies are needed to address the mechanism of intron transposition.

4.2.1 Involvement of Ty transposable element

The mutant analyses in this study suggest that Ty transposable elements might be involved in the overall intron mobilization. In addition to being a promising source for the reverse transcriptase activity, Ty transposable elements are associated with the mobilized reporter intron, indicating that they might be implicated in genomic integration of the reporter intron. Given the predicted roles of Ty transposable elements, it would be

interesting to explore how intron mobilization is affected in various mutant strains of Ty transposable element including Ty1 reverse transcriptase active site mutants (Uzun and Gabriel, 2001), deletion and overexpression of Ty integrase strain (Friedl et al., 2010; Wilhelm and Wilhelm, 2006), Ty1-less strains (Garfinkel et al., 2003). This would help us to dissect the roles of Ty transposable elements in the intron mobilization. Additionally, as the *his+* formation occurs very infrequently in diploid strains possibly due to low transcription of Ty transposable elements, it would be also interesting to examine the effect of Ty overexpression, or heterologous reverse transcriptase overexpression in the diploid strains of the wild-type and conditional mutants of *prp22* and *prp43*.

4.2.2 Connection between *RAD52* and Ty transposable element

Considering that the transposition of Ty transposable element is likely to be involved in the intron mobilization in this study, the observation that the *his+* prototroph formation is highly inhibited in *rad52* deletion strain is not consistent with the previous study showing that Ty1 transposition is increased in *rad52*Δ strain (Rattray et al., 2000). The result that the diploid strain of the wild-type also showed strong inhibition of *his+* prototroph formation possibly due to the low transcription of Ty transposable element suggests that both *RAD52* and Ty transposable element are crucial factors for the intron mobilization. The specific role of *RAD52* and its potential connection to Ty transposable element in the intron mobilization need to be further examined. It would be interesting to determine how the overexpression of *RAD52* affects the rate of *his+* prototroph formation in the mutant strains of Ty transposable element (Dornfeld and Livingston, 1991).

4.2.3 Reversibility of pre-mRNA splicing

In this study, we showed that the rate of *his+* prototroph formation is increased in the *prp22* conditional mutant while it is inhibited in the *prp43* conditional mutants. This difference was notable as it may reflect the involvement of reverse splicing in the plasmid-borne event as discussed in Section 3.4. Further study is indispensable to understand the mechanism behind this observation.

Assuming the spliceosome is stable and does not undergo RNA helicase-mediated rearrangements for the reverse reaction as Tseng and Cheng suggested (Tseng and Cheng, 2008), it would be interesting to test how the mutation in the catalytic core affects the rate and nature of *his+* prototroph formation. *PRP8* would be an interesting target for this approach as it is associated with the 5' splice site, the branchpoint sequence, and the 3' splice site of the pre-mRNA as well as the U5 and U6 snRNAs within the catalytic center of the spliceosome (Grainger and Beggs, 2005). Additionally, the recent structural study of Prp8 has shown that it contains a reverse transcriptase-like domain as well as endonuclease and RNase H-like domains (Dlakic and Mushegian, 2011; Nguyen et al., 2015). The domain architecture of Prp8 is similar to that of the IEP of group II self-splicing intron, providing additional evidence that the spliceosome is evolutionarily related to the group II intron. Taken together, it would be interesting to examine the pattern of intron mobilization in the mutants of *PRP8* using our intron gain and loss reporter.

References

Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res *30*, 1427-1464.

Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc Natl Acad Sci U S A *94*, 11798-11802.

Babenko, V.N., Rogozin, I.B., Mekhedov, S.L., and Koonin, E.V. (2004). Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res *32*, 3724-3733.

Basile, G., Aker, M., and Mortimer, R.K. (1992). Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene RAD51. Mol Cell Biol *12*, 3235-3246.

Bitter, G.A., and Egan, K.M. (1984). Expression of heterologous genes in Saccharomyces cerevisiae from vectors utilizing the glyceraldehyde-3-phosphate dehydrogenase gene promoter. Gene *32*, 263-274.

Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491-500.

Brody, E., and Abelson, J. (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science *228*, 963-967.

Brosi, R., Hauri, H.P., and Kramer, A. (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. J Biol Chem *268*, 17640-17646.

Castillo-Davis, C.I., Bedford, T.B., and Hartl, D.L. (2004). Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites. Mol Biol Evol *21*, 1422-1427.

Cavalier-Smith, T. (1985). Selfish DNA and the origin of introns. Nature *315*, 283-284.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. Trends Genet *7*, 145-148.

Cech, T.R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell *44*, 207-210.

Chalamcharla, V.R., Curcio, M.J., and Belfort, M. (2010). Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. Genes Dev *24*, 827-836.

Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. J Biol Chem *280*, 31190-31199.

Chapman, K.B., and Boeke, J.D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. Cell *65*, 483-492.

Chen, J.H., and Lin, R.J. (1990). The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. Nucleic Acids Res *18*, 6447.

Cheng, Z., and Menees, T.M. (2004). RNA branching and debranching in the yeast retrovirus-like element Ty1. Science *303*, 240-243.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev *11*, 2755-2766.

Collen, J., Porcel, B., Carre, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K.*, et al.* (2013). Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proc Natl Acad Sci U S A *110*, 5247-5252.

Collins, L., and Penny, D. (2005). Complex spliceosomal organization ancestral to extant eukaryotes. Mol Biol Evol *22*, 1053-1066.

Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. Nature *349*, 487-493.

Consortium, I.H.G.S. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17-37.

Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. RNA Biol *10*, 83-95.

Csuros, M., Rogozin, I.B., and Koonin, E.V. (2008). Extremely intron-rich genes in the alveolate ancestors inferred with a flexible maximum-likelihood approach. Mol Biol Evol *25*, 903-911.

Csuros, M., Rogozin, I.B., and Koonin, E.V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS Comput Biol *7*, e1002150.

Curcio, M.J., and Garfinkel, D.J. (1991). Single-step selection for Ty1 element retrotransposition. Proc Natl Acad Sci U S A *88*, 936-940.

Curcio, M.J., and Garfinkel, D.J. (1994). Heterogeneous functional Ty1 elements are abundant in the Saccharomyces cerevisiae genome. Genetics *136*, 1245-1259.

Dalbadie-McFarland, G., and Abelson, J. (1990). PRP5: a helicase-like protein required for mRNA splicing in yeast. Proc Natl Acad Sci U S A *87*, 4236-4240.

Darnell, J.E., Jr. (1978). Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science *202*, 1257-1260.

Davies, A.A., Friedberg, E.C., Tomkinson, A.E., Wood, R.D., and West, S.C. (1995). Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination. J Biol Chem *270*, 24638-24641.

Dayie, K.T., and Padgett, R.A. (2008). A glimpse into the active site of a group II intron and maybe the spliceosome, too. RNA *14*, 1697-1703.

de la Cruz, J., Kressler, D., and Linder, P. (1999). Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci *24*, 192-198.

Derr, L.K. (1998). The involvement of cellular recombination and repair genes in RNA-mediated recombination in Saccharomyces cerevisiae. Genetics *148*, 937-945.

Derr, L.K., and Strathern, J.N. (1993). A role for reverse transcripts in gene conversion. Nature *361*, 170-173.

Derr, L.K., Strathern, J.N., and Garfinkel, D.J. (1991). RNA-mediated recombination in S. cerevisiae. Cell *67*, 355-364.

Dibb, N.J., and Newman, A.J. (1989). Evidence that introns arose at proto-splice sites. EMBO J *8*, 2015-2021.

Dlakic, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA *17*, 799-808.

Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E., and Abelson, J. (1984). Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. Cell *39*, 611-621.

Dornfeld, K.J., and Livingston, D.M. (1991). Effects of controlled RAD52 expression on repair and recombination in Saccharomyces cerevisiae. Mol Cell Biol *11*, 2013-2017.

Drysdale, R.A., and Crosby, M.A. (2005). FlyBase: genes and gene models. Nucleic Acids Res *33*, D390-395.

Dziembowski, A., Ventura, A.P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O., and Seraphin, B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO J *23*, 4847-4856.

Elder, R.T., St John, T.P., Stinchcomb, D.T., Davis, R.W., and Scherer, S. (1981). Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. II. Recombination and expression of Ty1 and adjacent sequences. Cold Spring Harb Symp Quant Biol *45 Pt 2*, 581-591.

Farlow, A., Meduri, E., and Schlotterer, C. (2011). DNA double-strand break repair and the evolution of intron density. Trends Genet *27*, 1-6.

Fedorov, A., Merican, A.F., and Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. Proc Natl Acad Sci U S A *99*, 16128-16133.

Fink, G.R. (1987). Pseudogenes in yeast? Cell *49*, 5-6.

Fortin, G.S., and Symington, L.S. (2002). Mutations in yeast Rad51 that partially bypass the requirement for Rad55 and Rad57 in DNA repair by increasing the stability of Rad51-DNA complexes. EMBO J *21*, 3160-3170.

Friedl, A.A., Kiechle, M., Maxeiner, H.G., Schiestl, R.H., and Eckardt-Schupp, F. (2010). Ty1 integrase overexpression leads to integration of non-Ty1 DNA fragments into the genome of Saccharomyces cerevisiae. Mol Genet Genomics *284*, 231-242.

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

Garfinkel, D.J., Nyswaner, K.M., Stefanisko, K.M., Chang, C., and Moore, S.P. (2005). Ty1 copy number dynamics in Saccharomyces. Genetics *169*, 1845-1857.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Gilbert, W. (1987). The exon theory of genes. Cold Spring Harb Symp Quant Biol *52*, 901-905.

Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J., and McFadden, G.I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A *103*, 9566-9571.

Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Luhrmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. EMBO J *18*, 4535-4548.

Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev *10*, 233-243.

Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. Cell *37*, 415-427.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. RNA *11*, 533-557.

Haas, B.J., Wortman, J.R., Ronning, C.M., Hannick, L.I., Smith, R.K., Jr., Maiti, R., Chan, A.P., Yu, C., Farzad, M., Wu, D.*, et al.* (2005). Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. BMC Biol *3*, 7.

Hays, S.L., Firmenich, A.A., and Berg, P. (1995). Complex formation in yeast double-strand break repair: participation of Rad51, Rad52, Rad55, and Rad57 proteins. Proc Natl Acad Sci U S A *92*, 6925-6929.

Hirschman, J.E., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hong, E.L., Livstone, M.S., Nash, R.*, et al.* (2006). Genome Snapshot: a new resource at the Saccharomyces Genome Database (SGD) presenting an overview of the Saccharomyces cerevisiae genome. Nucleic Acids Res *34*, D442-445.

Horecka, J., and Jigami, Y. (2000). Identifying tagged transposon insertion sites in yeast by direct genomic sequencing. Yeast *16*, 967-970.

Irimia, M., Maeso, I., Burguera, D., Hidalgo-Sanchez, M., Puelles, L., Roy, S.W., Garcia-Fernandez, J., and Ferran, J.L. (2011). Contrasting 5' and 3' evolutionary histories and frequent evolutionary convergence in Meis/hth gene structures. Genome Biol Evol *3*, 551-564.

Irimia, M., and Roy, S.W. (2014). Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol *6*.

Irimia, M., Rukov, J.L., Penny, D., Vinther, J., Garcia-Fernandez, J., and Roy, S.W. (2008). Origin of introns by 'intronization' of exonic sequences. Trends Genet *24*, 378-381.

Ivanov, E.L., and Haber, J.E. (1995). RAD1 and RAD10, but not other excision repair genes, are required for double-strand break-induced recombination in Saccharomyces cerevisiae. Mol Cell Biol *15*, 2245-2251.

James, S.A., Turner, W., and Schwer, B. (2002). How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. RNA *8*, 1068-1077.

Jiang, K., and Goertzen, L.R. (2011). Spliceosomal intron size expansion in domesticated grapevine (Vitis vinifera). BMC Res Notes *4*, 52.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kim, S.H., Smith, J., Claude, A., and Lin, R.J. (1992). The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. EMBO J *11*, 2319-2326.

King, D.S., and Beggs, J.D. (1990). Interactions of PRP2 protein with pre-mRNA splicing complexes in Saccharomyces cerevisiae. Nucleic Acids Res *18*, 6559-6564.

Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev *15*, 42-49.

Klinz, F.J., and Gallwitz, D. (1985). Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast Saccharomyces cerevisiae. Nucleic Acids Res *13*, 3791-3804.

Konarska, M.M., and Sharp, P.A. (1988). Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. Proc Natl Acad Sci U S A *85*, 5459-5462.

Koodathingal, P., Novak, T., Piccirilli, J.A., and Staley, J.P. (2010). The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing. Mol Cell *39*, 385-395.

Koonin, E.V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct *1*, 22.

Koonin, E.V., Csuros, M., and Rogozin, I.B. (2013). Whence genes in pieces: reconstruction of the exon-intron gene structures of the last eukaryotic common ancestor and other ancestral eukaryotes. Wiley Interdiscip Rev RNA *4*, 93-105.

Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol *3*, a003616.

Lane, C.E., van den Heuvel, K., Kozera, C., Curtis, B.A., Parsons, B.J., Bowman, S., and Archibald, J.M. (2007). Nucleomorph genome of Hemiselmis andersenii reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci U S A *104*, 19908-19913.

Lardelli, R.M., Thompson, J.X., Yates, J.R., 3rd, and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA *16*, 516-528.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet *34*, 77-137.

Legrain, P., Seraphin, B., and Rosbash, M. (1988). Early commitment of yeast pre-mRNA to the spliceosome pathway. Mol Cell Biol *8*, 3755-3760.

Li, F., Dong, J., Eichmiller, R., Holland, C., Minca, E., Prakash, R., Sung, P., Yong Shim, E., Surtees, J.A., and Eun Lee, S. (2013). Role of Saw1 in Rad1/Rad10 complex assembly at recombination intermediates in budding yeast. EMBO J *32*, 461-472.

Li, W., Tucker, A.E., Sung, W., Thomas, W.K., and Lynch, M. (2009). Extensive, recent intron gains in Daphnia populations. Science *326*, 1260-1262.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181-211.

Lin, R.J., Newman, A.J., Cheng, S.C., and Abelson, J. (1985). Yeast mRNA splicing in vitro. J Biol Chem *260*, 14780-14792.

Liu, S., Rauhut, R., Vornlocher, H.P., and Luhrmann, R. (2006). The network of protein-protein interactions within the human U4/U6.U5 tri-snRNP. RNA *12*, 1418-1430.

Logsdon, J.M., Jr., Tyshenko, M.G., Dixon, C., J, D.J., Walker, V.K., and Palmer, J.D. (1995). Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc Natl Acad Sci U S A *92*, 8507-8511.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205-2208.

Makarova, O.V., Makarov, E.M., and Luhrmann, R. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. EMBO J *20*, 2553-2563.

Malone, R.E., Montelone, B.A., Edwards, C., Carney, K., and Hoekstra, M.F. (1988). A reexamination of the role of the RAD52 gene in spontaneous mitotic recombination. Curr Genet *14*, 211-223.

Martin, A., Schneider, S., and Schwer, B. (2002). Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. J Biol Chem *277*, 17743-17750.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol *15*, 108-121.

Mayas, R.M., Maita, H., Semlow, D.R., and Staley, J.P. (2010). Spliceosome discards intermediates via the DEAH box ATPase Prp43p. Proc Natl Acad Sci U S A *107*, 10020-10025.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L.*, et al.* (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science *318*, 245-250.

Milne, G.T., and Weaver, D.T. (1993). Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. Genes Dev *7*, 1755-1765.

Mourier, T., and Jeffares, D.C. (2003). Eukaryotic intron loss. Science *300*, 1393.

Nam, K., Hudson, R.H., Chapman, K.B., Ganeshan, K., Damha, M.J., and Boeke, J.D. (1994). Yeast lariat debranching enzyme. Substrate and sequence specificity. J Biol Chem *269*, 20613-20621.

Nam, K., Lee, G., Trambley, J., Devine, S.E., and Boeke, J.D. (1997). Severe growth defect in a Schizosaccharomyces pombe mutant defective in intron lariat degradation. Mol Cell Biol *17*, 809-818.

Newman, A.J., and Norman, C. (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. Cell *68*, 743-754.

Newman, A.J., Teigelkamp, S., and Beggs, J.D. (1995). snRNA interactions at 5' and 3' splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA *1*, 968-980.

Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature *523*, 47-52.

O'Day, C.L., Dalbadie-McFarland, G., and Abelson, J. (1996). The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. J Biol Chem *271*, 33261-33267.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science *225*, 898-903.

Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.F., Wellinger, R.J., Chabot, B., and Elela, S.A. (2011). Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. Cell *147*, 320-331.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods *24*, 218-229.

Raghunathan, P.L., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol *8*, 847-855.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.

Rattray, A.J., Shafer, B.K., and Garfinkel, D.J. (2000). The Saccharomyces cerevisiae DNA recombination and repair functions of the RAD52 epistasis group inhibit Ty1 transposition. Genetics *154*, 543-556.

Rogers, J.H. (1989). How were introns inserted into nuclear genes? Trends Genet *5*, 213-216.

Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E.V. (2012). Origin and evolution of spliceosomal introns. Biol Direct *7*, 11.

Rogozin, I.B., Wolf, Y.I., Sorokin, A.V., Mirkin, B.G., and Koonin, E.V. (2003). Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol *13*, 1512-1517.

Roy, S.W. (2006). Intron-rich ancestors. Trends Genet *22*, 468-471.

Roy, S.W., and Gilbert, W. (2005a). Complex early genes. Proc Natl Acad Sci U S A *102*, 1986-1991.

Roy, S.W., and Gilbert, W. (2005b). The pattern of intron loss. Proc Natl Acad Sci U S A *102*, 713-718.

Roy, S.W., and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet *7*, 211-221.

Roy, S.W., and Irimia, M. (2009). Mystery of intron gain: new data and new models. Trends Genet *25*, 67-73.

Ruby, S.W., Chang, T.H., and Abelson, J. (1993). Four yeast spliceosomal proteins (PRP5, PRP9, PRP11, and PRP21) interact to promote U2 snRNP binding to pre-mRNA. Genes Dev *7*, 1909-1925.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317-331.

Sakurai, A., Fujimori, S., Kochiwa, H., Kitamura-Abe, S., Washio, T., Saito, R., Carninci, P., Hayashizaki, Y., and Tomita, M. (2002). On biased distribution of introns in various eukaryotes. Gene *300*, 89-95.

San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. Annu Rev Biochem *77*, 229-257.

Sapra, A.K., Khandelia, P., and Vijayraghavan, U. (2008). The splicing factor Prp17 interacts with the U2, U5 and U6 snRNPs and associates with the spliceosome pre- and post-catalysis. Biochem J *416*, 365-374.

Schneider, B.L., Steiner, B., Seufert, W., and Futcher, A.B. (1996). pMPY-ZAP: a reusable polymerase chain reaction-directed gene disruption cassette for Saccharomyces cerevisiae. Yeast *12*, 129-134.

Schwer, B., and Gross, C.H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J *17*, 2086-2094.

Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. Nature *349*, 494-499.

Schwer, B., and Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. EMBO J *11*, 5033-5039.

Schwer, B., and Meszaros, T. (2000). RNA helicase dynamics in pre-mRNA splicing. EMBO J *19*, 6582-6591.

Seraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. EMBO J *7*, 2533-2538.

Seraphin, B., and Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell *59*, 349-358.

Sharp, P.A. (1985). On the origin of RNA splicing and introns. Cell *42*, 397-400.

Sharp, P.A. (1991). "Five easy pieces". Science *254*, 663.

Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R., and Green, M.R. (2008). Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. Genes Dev *22*, 1796-1803.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell *69*, 457-470.

Sorenson, M.R., and Stevens, S.W. (2014). Rapid identification of mRNA processing defects with a novel single-cell yeast reporter. RNA *20*, 732-745.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221-234.

Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64.

Steitz, T.A. (2008). A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol *9*, 242-253.

Stevens, S.W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci U S A *96*, 7226-7231.

Stevens, S.W., Barta, I., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA *7*, 1543-1553.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

Sung, P. (1997). Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. Genes Dev *11*, 1111-1121.

Sung, P., and Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol *7*, 739-750.

Sverdlov, A.V., Babenko, V.N., Rogozin, I.B., and Koonin, E.V. (2004). Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion. Gene *338*, 85-91.

Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev *66*, 630-670, table of contents.

Tanner, N.K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell *8*, 251-262.

Tarn, W.Y., and Steitz, J.A. (1996a). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science *273*, 1824-1832.

Tarn, W.Y., and Steitz, J.A. (1996b). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell *84*, 801-811.

Tomkinson, A.E., Bardwell, A.J., Bardwell, L., Tappe, N.J., and Friedberg, E.C. (1993). Yeast DNA repair and recombination proteins Rad1 and Rad10 constitute a single-stranded-DNA endonuclease. Nature *362*, 860-862.

Torriani, S.F., Stukenbrock, E.H., Brunner, P.C., McDonald, B.A., and Croll, D. (2011). Evidence for extensive recent intron transposition in closely related fungi. Curr Biol *21*, 2017-2022.

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y., He, J., Wang, H., Razavian, N., Berns, M.W., and Wu, X. (2013). Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc Natl Acad Sci U S A *110*, 7720-7725.

Tseng, C.K., and Cheng, S.C. (2008). Both catalytic steps of nuclear pre-mRNA splicing are reversible. Science *320*, 1782-1784.

Tuo, S., Nakashima, K., and Pringle, J.R. (2012). Apparent defect in yeast bud-site selection due to a specific failure to splice the pre-mRNA of a regulator of cell-type-specific transcription. PLoS One *7*, e47621.

Turunen, J.J., Niemela, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA *4*, 61-76.

Uzun, O., and Gabriel, A. (2001). A Ty1 reverse transcriptase active-site aspartate mutation blocks transposition but not polymerization. J Virol *75*, 6337-6347.

Valadkhan, S., and Manley, J.L. (2009). The use of simple model systems to study spliceosomal catalysis. RNA *15*, 4-7.

Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J.L. (2007). Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. RNA *13*, 2300-2311.

van der Burgt, A., Severing, E., de Wit, P.J., and Collemare, J. (2012). Birth of new spliceosomal introns in fungi by multiplication of introner-like elements. Curr Biol *22*, 1260-1265.

Wagner, J.D., Jankowsky, E., Company, M., Pyle, A.M., and Abelson, J.N. (1998). The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. EMBO J *17*, 2926-2937.

Wang, Q., He, J., Lynn, B., and Rymond, B.C. (2005). Interactions of the yeast SF3b splicing factor. Mol Cell Biol *25*, 10745-10754.

Wilhelm, M., and Wilhelm, F.X. (2006). Cooperation between reverse transcriptase and integrase during reverse transcription and formation of the preintegrative complex of Ty1. Eukaryot Cell *5*, 1760-1769.

Wise, J.A. (1991). Preparation and analysis of low molecular weight RNAs and small ribonucleoproteins. Methods Enzymol *194*, 405-415.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.

Worden, A.Z., Lee, J.H., Mock, T., Rouze, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M.V.*, et al.* (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science *324*, 268-272.

Yaffe, M.P., and Schatz, G. (1984). Two nuclear mutations that block mitochondrial protein import in yeast. Proc Natl Acad Sci U S A *81*, 4819-4823.

Yenerall, P., and Zhou, L. (2012). Identifying the mechanisms of intron gain: progress and trends. Biol Direct *7*, 29.

VITA

Sujin Lee was born in 1982 in Chuncheon, South Korea. She earned her B.S. and M.S. in Science in Agriculture at Seoul National University, South Korea. She began her Ph.D. study at University of Texas at Austin.

sujinlee@utexas.edu

This dissertation was typed by Sujin Lee