5-2016

Effect of Group IIC Introns on Host Gene Expression in S. cerevisiae

Daniel P. Bollen
University at Albany, State University of New York

Effect of group IIC introns on host gene expression in *S. cerevisiae*

An honors thesis presented to the
Department of Biological Sciences
University at Albany, State University of New York
in partial fulfillment of the requirements for
graduation with Honors in Biological Sciences
and
graduation from the Honors College.

Daniel P. Bollen
May 2016

Research Mentor: Guosheng Qu, PhD
Research Advisor: Marlene Belfort, PhD

Abstract

Group II introns are self-splicing ribozymes believed to be the evolutionary ancestors to eukaryotic spliceosomal introns. They are found in bacteria, archaea, and eukaryotic organelles but not in eukaryotic nuclear genomes. Group IIA and IIB introns introduced to nuclei in *Saccharomyces cerevisiae* are able to transcribe and splice, but post-transcriptionally silence their host genes. It has been shown that both RNA-RNA interactions and pre-mRNA cytoplasmic mislocalization contribute to the silencing of the genes. RNA-RNA interactions include binding between the exon-binding sequences (EBS) in the intron and intron-binding sequences (IBS) in the ligated exons. Group IIC introns contain reduced EBS-IBS sequences compared to their group IIA and group IIB counterparts; thus it is worthwhile to investigate if IIC introns will function differently in yeast, either causing no silencing or using a distinct silencing mechanism. To this end, a group IIC-reporter cassette was prepared and transformed into yeast. Phenotype screening of the yeast was carried out to determine if silencing occurs. Mechanisms, including the presence of RNA-RNA interactions and their impact on the host genes were investigated. Primer extension assays conducted show that group IIC-reporter constructs can splice *in vivo* in *S. cerevisiae*, although exon ligation does not occur. This is a noteworthy distinction in silencing mechanism between the closely related group IIA, IIB, and IIC introns. This information will assist in establishing the evolutionary narrative of the expulsion of group II introns from nuclear genomes and the evolution of group II introns to ancestrally-related spliceosomal introns.

Acknowledgements

I would like to formally thank my research advisor, Dr. Marlene Belfort, for her guidance and mentorship, as well as the invaluable opportunity to work in her lab. I would like to extend sincerest thanks to Dr. Guosheng Qu for his mentorship and expertise regarding my honors thesis project. In addition, I would like to thank Carol Lyn Piazza for all of her fantastic help in troubleshooting and technical assistance. I would like to thank Dorie Smith for her technical assistance and guidance. I would also like to thank Eren Dong for his mentorship and assistance during my first year in the lab.

Within the Department of Biology, I would like to thank Dr. Richard Zitomer for his continued counsel and advisement over my undergraduate career. I would like to thank Dr. Robert Osuna for his advisement and support. Finally, I would like to thank Dr. Prashanth Rangan for his help on my thesis as a member of my Honors Committee.

Last but not least, I would like to thank my friends and family for their endless love and support before, during, and after my undergraduate career.

Table of Contents

Introduction

Group II introns are autocatalytic mobile genetic elements found in bacteria, archaea, and eukaryotic organelles (Lambowitz & Belfort, 2015). They are like spliceosomal introns in that they interrupt host genes and are spliced out post-transcriptionally via the same splicing pathway, but differ in that they catalyze their own splicing, instead of relying on cellular machinery (Beauregard, Curcio, & Belfort, 2008). They consist of a self-splicing ribozyme and a multi-functional intron-encoded protein (IEP) (Lambowitz & Zimmerly, 2011). The IEP assists in splicing and mobility of the ribozyme, allowing the intron to propagate efficiently through host genomes (Lambowitz & Belfort, 2015). In this way they are similar to retrotransposons (Beauregard, Curcio, & Belfort, 2008). Group II introns are found in all domains of life, (Lambowitz & Belfort, 2015) though are notably absent from eukaryotic nuclear genomes (Zimmerly & Semper, 2015; Lambowitz & Belfort, 2015). Group II introns are commonly believed to have played a pivotal role in eukaryotic evolution, and are believed to be ancestrally related to both retrotransposons and spliceosomal introns (Lambowitz & Zimmerly, 2011).

Group II introns splice themselves from their host mRNAs via one of two pathways: a lariat-producing, transesterification pathway similar to that of the splicing of spliceosomal introns (Fig. 1A) or a linear intron-producing, hydrolytic pathway (Fig. 1B) (Lambowitz & Zimmerly, 2011). Only some introns use the hydrolytic pathway, and some use it exclusively *in vivo*. The pathways consist of two consecutive transesterification reactions (Lambowitz & Zimmerly, 2011), In the lariat-producing pathway, first a 2'-OH of an adenosine near the 3' exon executes a nucleophilic attack on the 5' splice site, forming a 3' exon-lariat intermediate. Then, the 3'-OH of the 5' exon attacks the 3' splice site, ligating the two exons together (Lambowitz & Zimmerly, 2011). The group II intron is then released as a lariat, similar to spliceosomal introns (Lambowitz & Belfort, 2015). The hydrolytic pathway is similar, except water acts as the nucleophile in the first transesterification reaction, producing a linear intron-3' exon intermediate. This intermediate is then attacked by the 3'-OH of the 5' exon, ligating them together and excising linear intron (Lambowitz & Zimmerly, 2011). Group II introns using either pathway require the IEP to splice *in vivo* (Zimmerly & Semper, 2015). These reactions are reversible (Lambowitz &

Zimmerly, 2011; Lambowitz & Belfort, 2015), therefore an excised intron can insert itself back into ligated exons.

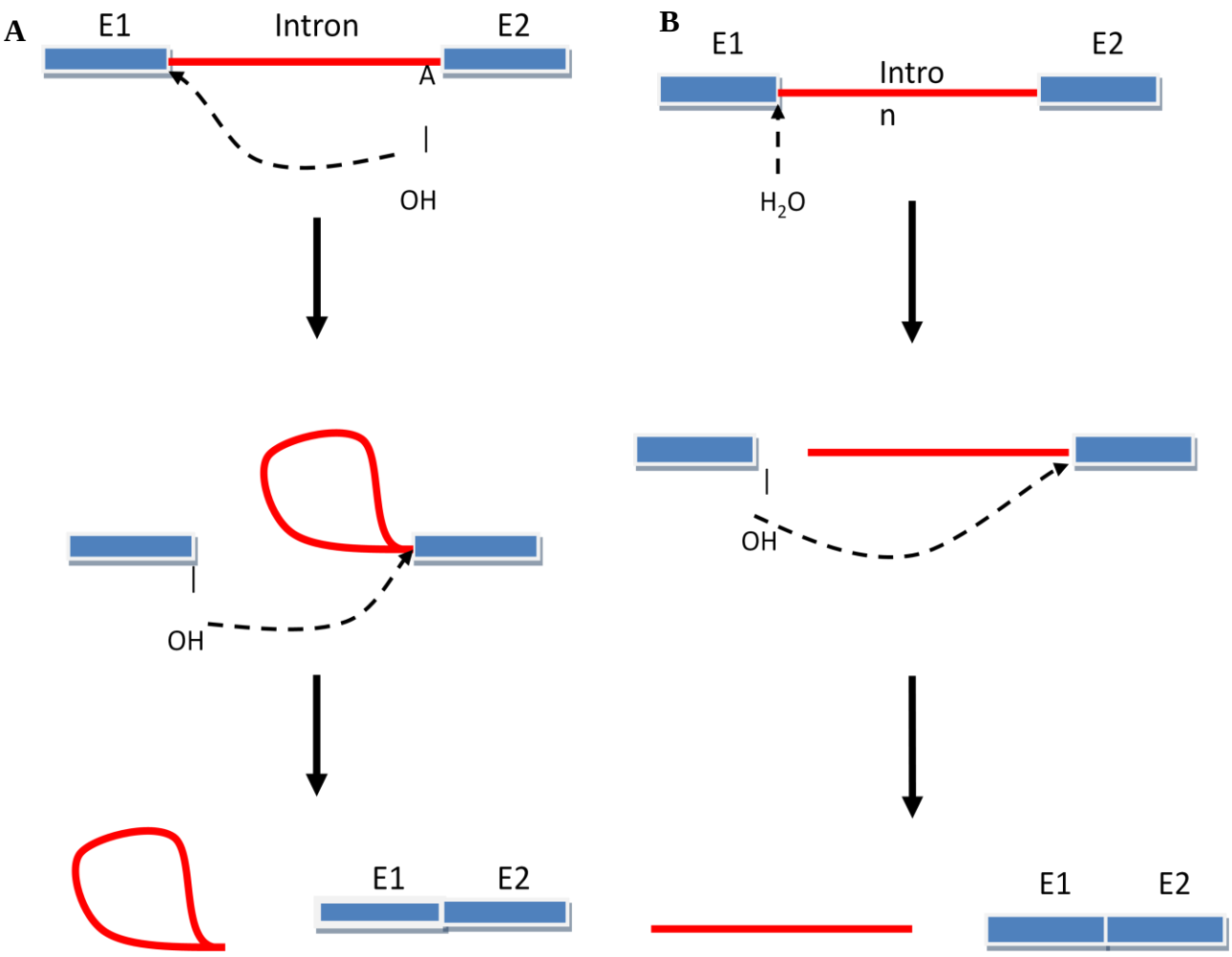

Figure 1—Splicing reactions of group II introns. The IEP is not shown. (A) The lariat-producing splicing mechanism of the group II intron. Two transesterification reactions result in the intron removed as a lariat. (B) The linear-intron producing splicing mechanism. Water executes the first nucleophilic attack on the 5' splice site in this pathway.

Group II introns achieve efficient splicing and mobility through use of their IEP (Zimmerly & Semper, 2015). The IEP is multi-functional, containing a maturase, reverse transcriptase, and sometimes endonuclease domain (Beauregard, Curcio & Belfort, 2008; Lambowitz & Belfort, 2015). The maturase domain of the IEP folds and stabilizes the intron's secondary and tertiary structure into the catalytic state. This promotes splicing *in vivo* (Lambowitz & Zimmerly, 2011; Lambowitz & Belfort, 2015). Once

spliced, the ribozyme associates with the IEP to form a ribonucleoprotein, or RNP. The IEP binds to the ribozyme at the maturase and reverse transcriptase domains (Zimmerly & Semper, 2015) just after translation of the IEP, and remains bound during splicing and intron mobility (Lambowitz & Belfort, 2015). Once the RNP is formed, the group II intron can insert into DNA sequences through reverse splicing (Lambowitz & Zimmerly, 2011). To do so, the ribozyme first reverse splices into a strand of DNA. If an endonuclease domain is present, it cleaves the other DNA strand, allowing the reverse transcriptase domain to reverse transcribe the intron into the DNA. If there is no endonuclease, the insertion happens in ssDNA, in the template for the lagging strand during DNA replication (Beauregard, Curcio & Belfort, 2008; Lambowitz & Zimmerly, 2011). The primer for reverse transcription in this case is the nascent lagging strand (Lambowitz & Zimmerly, 2011). Mobile group II introns can insert through two retromobility pathways: retrohoming and retrotransposition (Beauregard, Curcio & Belfort, 2008). In retrohoming, which occurs at frequencies near 100% in recipients acquiring the intron, insertion occurs at specific DNA sequences. Retrotransposition consists of insertion at ectopic DNA sites, and happens at least 10^4 less often than retrohoming (Lambowitz & Zimmerly 2011; Lambowitz & Belfort, 2015).

There are three subclasses of group II introns, known as group IIA, IIB, and IIC. They differ in size, secondary structure, mode of exon recognition, IEP capabilities, and intron-exon interactions. Group IIA intron IEPs have endonuclease domains, and thus insert into dsDNA, whereas IIB and IIC introns insert into ssDNA (Lambowitz & Belfort, 2015). Group IIC introns are found only in bacteria (Toor et al., 2006; Zimmerly & Semper, 2015), whereas group IIA and IIB introns are also found in organellar genomes of eukaryotes (Lambowitz & Belfort, 2015). IIC introns are also shorter than their IIA and IIB counterparts (Toor et al., 2006). Intron-exon interactions that differ between intron subclasses include interactions between intron-binding sequences in the exons (IBSs) and exon-binding sequences in the intron (EBSs) (Lambowitz & Belfort, 2015). These IBS and EBSs interact in a Watson-Crick base pairing mechanism (Zimmerly & Semper, 2015) to bind the intron to the ligated exons, and are key in splicing and target site recognition for intron retromobility. The size of these interactions differs for each subclass: ~13 bp for IIA, ~14 bp for IIB, and ~6 bp for IIC (Lambowitz & Zimmerly, 2011).

As stated previously, group II introns are found in all domains of life. They are most common in bacteria, where they are found in 25% of genomes. Very few are found in archaea, and those found in eukaryotic organellar genomes (mitochondrial and chloroplast genomes) are often degenerate, and thus no longer correctly code for mobility (Lambowitz & Zimmerly, 2011; Lambowitz & Belfort, 2015). Group II introns are notably absent from eukaryotic nuclear genomes, but are believed to be the ancestors of eukaryotic spliceosomal introns (Martin & Koonin, 2006). Spliceosomal introns and group II introns possess a number of mechanistic and structural similarities (Lambowitz & Belfort, 2015), including splicing mechanism and boundary sequences (Lambowitz & Zimmerly, 2011). It is hypothesized that group II introns began as bacterial retroelements (the retroelement ancestor hypothesis). It is also believed that they are the ancestors of non-LTR retroelements (Lambowitz & Semper, 2015).

Group II introns are said to have played a "pivotal role" (Lambowitz & Belfort, 2015) in eukaryotic evolution. Martin and Koonin (2006) argued that group II introns spurred the evolution of the nuclear envelope shortly after the mitochondrial endosymbiosis. Group II-filled α-proteobacteria may have been enveloped by an archaeon. Subsequent group II intron invasions of the archaeal genome may have occurred (Martin & Koonin, 2006). Since group II introns impose minimal impairments or silencing of gene expression in bacteria (Beauregard, Curcio & Belfort, 2008; Lambowitz & Zimmerly, 2011) this is not an issue. However, once group II introns proliferate and become degenerate, they can no longer splice, causing problems (Martin & Koonin, 2006). For example, degenerate group II introns in mitochondrial and chloroplast genomes rely on host-encoded proteins to splice (Lambowitz & Zimmerly, 2011). To prevent translation of unspliced transcripts, a nuclear envelope would have evolved to separate transcription from translation in both space and time (Martin & Koonin, 2006).

When group II introns are placed into eukaryotic genomes from which they are so notably absent, they subject their host genes to silencing (Chalamcharla, Curcio & Belfort, 2010). This is perplexing: since group II introns can self-splice, they could be seen as self-silencing (Beauregard, Curcio & Belfort, 2008). Chalamcharla, Curcio and Belfort found that although splicing still occurs, it is carried out in the cytoplasm instead of the nucleus, as is the case with spliceosomal introns. After splicing, the RNA is

subject to nonsense-mediated decay (NMD) and translational repression. Group I and artificial spliceosomal introns inserted into the same loci do not cause silencing of the host gene (Chalamcharla, Curcio & Belfort, 2010). Qu et al. (2014) showed that this silencing is due to pre-mRNA cytoplasmic mislocalization, as well as RNA-RNA interactions via IBS-EBS base pairing.

Silencing of group II introns in *S. cerevisiae* has been observed and studied with group IIA and IIB introns, but not IIC introns. Due to their reduced IBS-EBS interactions, group IIC introns in *S. cerevisiae* may not be subjected to silencing as group IIA and IIB introns are, or may be silenced via a different mechanism. Studying the effects of reduced EBS-IBS interactions in *S. cerevisiae* can help establish the precise nature of group II intron silencing in eukaryotes. This can further elucidate the evolutionary relationship between group II introns and spliceosomal introns.

Materials and Methods

Cloning

To determine if group IIC introns undergo the same silencing fate in *S. cerevisiae* as other group II introns, a group IIC-reporter cassette was devised. The group IIC intron known as BhI1 from *Bacillus halodurans* was placed upstream of a CUP1 reporter gene, which confers resistance against copper. This cassette is then ligated into the yeast vector pCSR and transformed into yeast.

To complete this cloning process, the BhI1 intron was isolated from *B. halodurans* genomic DNA using PCR. The intron was then ligated to the CUP1 reporter gene using overlap extension PCR, in which homologous sequences of DNA overlap during the PCR process and act as primers (see Table I for primer list), resulting in the two ligated DNA fragments (Fig. 2). This was then ligated into the pCSR yeast vector and transformed into yeast. Ligation into the yeast vector requires a BamHI and SalI sticky end on either end of the insert. Because the BhI1-CUP1 cassette made has SalI restriction sites at each end, a 50 bp BamHI-SalI linker was designed and ligated into the pCSR vector (Fig. 3).

Name	Locus	Use	Sequence
4493	Upstream of Intron	Cloning	GTCGACATGAGCACAATCCAATCGGGTTG TGCTTTTTTAG
4160	Downstream of CUP1	Cloning	GTCGACGTCAGTGAGCGAGGAAGCGGAAG A
4893	5’ end of intron	Primer Extension	CGGGACTTGAACCCTAGAGAA
4163	Exon 2	Primer Extension	CCTTCATTTTGGAAGTTAATTAATTCGCTG

Nine yeast strains were produced for use for the duration of the study (Fig. 4). Each strain was transformed with a plasmid containing either the BhI1-CUP1 cassette, a HS-CUP1 cassette, or a BhI1ΔORF-CUP1 cassette (Fig 5A). The HS-CUP1 construct contains the ligated BhI1 exons with no intron (known as the homing site, or HS), and acts as a positive control. BhI1ΔORF-CUP1 is an alternate

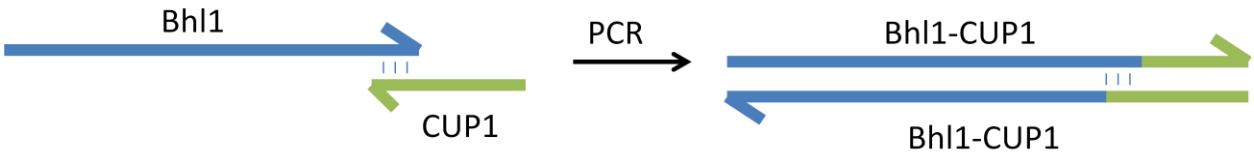

Figure 2—Overlap Extension PCR. Overlap between the BhI1 and CUP1 genes results in a ligated construct after undergoing the polymerase chain reaction. The annealed segment of the two fragments acts as a primer.

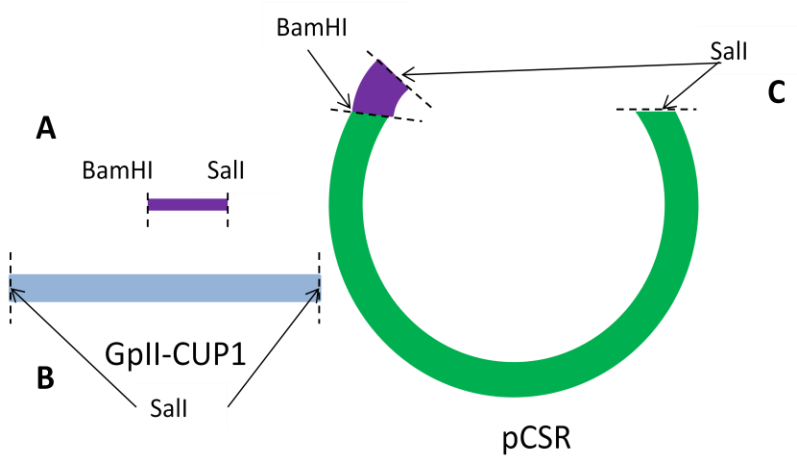

Figure 3—Cloning Strategy. (A) The BamHI-SalI linker. (B) The Group II-CUP1 cassette, with SalI sites at each end. (C) The pCSR yeast vector showing the BamHI and SalI sites, with linker ligated. The Group II-CUP1 cassette is ligated into the two SalI sites.

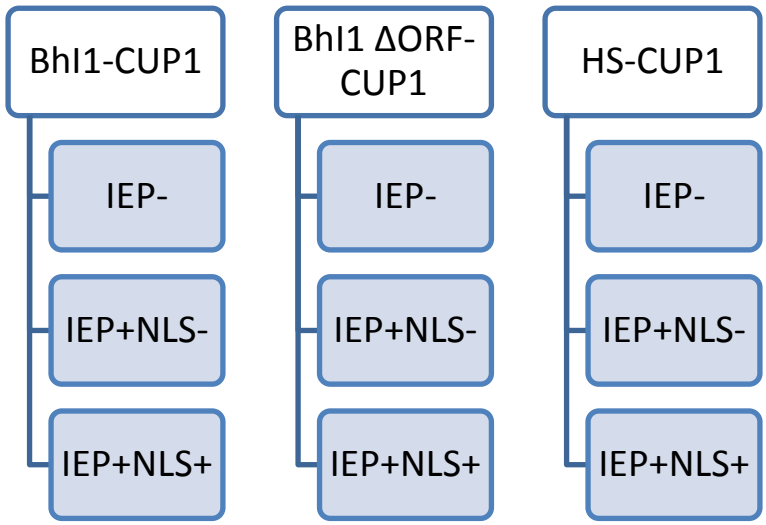

Figure 4—Intron-reporter cassette and expression vector combinations used. The IEP is the intron-encoded protein from BhI1. NLS is a nuclear localization signal. HS-CUP1 is the homing site of the LtrB intron with the intron itself removed. HS-CUP1 will express CUP1 and should experience no silencing. IEP- refers to no expression vector.

construct identical to the BhI1-CUP1 construct, though it contains a deletion of the open reading frame that codes for the IEP. Each cassette was inserted into the pCSR plasmid. HS-CUP1 and BhI1ΔORF-CUP1 cassettes were made in previous work (Miller, Qu, & Belfort (unpublished)). The pCSR plasmid contains a Trp+ selectable marker, allowing cells that contain them to grow on media lacking tryptophan.

Yeast containing each of these plasmids was either co-transformed with an IEP expression vector or contained no expression vector as a negative control. This expression vector consisted of a modified pESC plasmid containing a Leu+ selectable marker (which allows cells to grow on media lacking leucine), and a cassette containing the BhI1 IEP with a galactose-inducible promoter. Those co-transformed with an expression vector were of one of two varieties: an IEP vector including a nuclear localization signal (NLS) or one without (Fig. 5B).

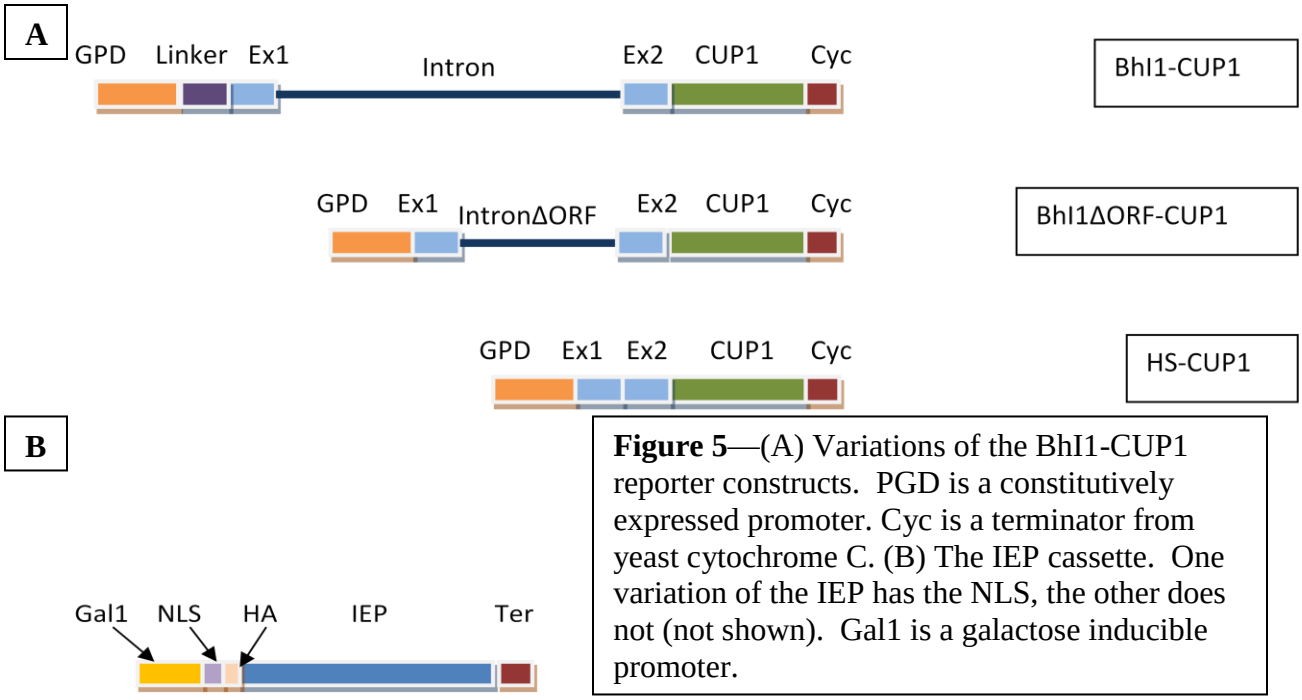

Figure 5—(A) Variations of the BhI1-CUP1 reporter constructs. PGD is a constitutively expressed promoter. Cyc is a terminator from yeast cytochrome C. (B) The IEP cassette. One variation of the IEP has the NLS, the other does not (not shown). Gal1 is a galactose inducible promoter.

Phenotype Assay

Once the nine yeast strains were produced, phenotypic assays were carried out. Four yeast colonies were plucked from a plate and resuspended in 30 µL water. 2 µL of this cell suspension was pipetted onto media containing $CuSO_4$ in concentrations of 0 mM, 0.01 mM, 0.05 mM, 0.1 mM, 0.5 mM, and 1.0 mM. Yeast without IEP expression vectors were plated on plates lacking tryptophan and containing glucose. Those with IEP expression vectors were plated on plates lacking tryptophan and leucine and containing galactose, which would induce expression of the IEP. Plates were incubated overnight at 30 ºC.

Primer Extension Assay

Once CUP1 phenotype was ascertained, investigations into intron splicing were conducted. Splicing of the intron-reporter construct *in vitro* and *in vivo* was assayed through the use of reverse transcriptase primer extension. In this process, a radiolabeled DNA primer is mixed with an RNA template and a reverse transcriptase. The reaction is incubated, and the reverse transcriptase uses the primer to make a radioactive cDNA. 1 µL of radiolabeled primer (see Table I for primers used) is mixed with 5 µg whole cell RNA or 250 ng of specific RNA template, along with 0.5 µL of 10 mM of each deoxyribonuclotide. Water is used to bring the reaction up to 6.5 µL. This mixture is incubated for 5 minutes at 65 ºC, then placed on ice for at least 1 minute. 3.5 µL of enzyme master mix consisting of 2 µL First Strand Buffer (ThermoFisher), 0.5 µL of 0.1 M DTT, 0.5 µL of RNaseIn, and 0.5 µL of SuperScript III enzyme (ThermoFisher) is added to the first solution. This 10 µL mixture is then incubated for 30 minutes at 50 ºC. The reaction mixture was mixed with a 2x loading dye, and 10 µL was loaded onto a 10% urea gel (National Diagnostics). The reaction products are visualized after being exposed on an autoradiograph phosphor screen. Length of the cDNA based on migration in the gel is compared to expected results for pre-mRNA, mature mRNA, excised intron or intron-exon intermediate. In addition, a reverse transcriptase primer extension was conducted using two different primers: an intron-specific primer and 3'-exon specific primer (Fig. 6A). When extended via the reverse transcriptase reactions, these transcripts are expected to produce cDNA products of sizes seen in Fig. 6B. These two primers were radiolabeled with P^{32} and were used in two separate primer extension reactions, extending both the negative control mRNA and the *in vitro* spliced mRNA.

***In Vitro* Splicing Assay**

To assay splicing of BhI1-CUP1 *in vitro*, BhI1-CUP1 constructs were first amplified using PCR and an *in vitro* transcription reaction was carried out. To do so, RNA polymerase was mixed with the PCR-produced DNA template along with dNTPs (MEGAscript T7 Transcription Kit, ThermoFisher). The reaction was made up of 125 ng DNA template, 2 µL each of each ribonucleotide at 75 mM, 2 µL of

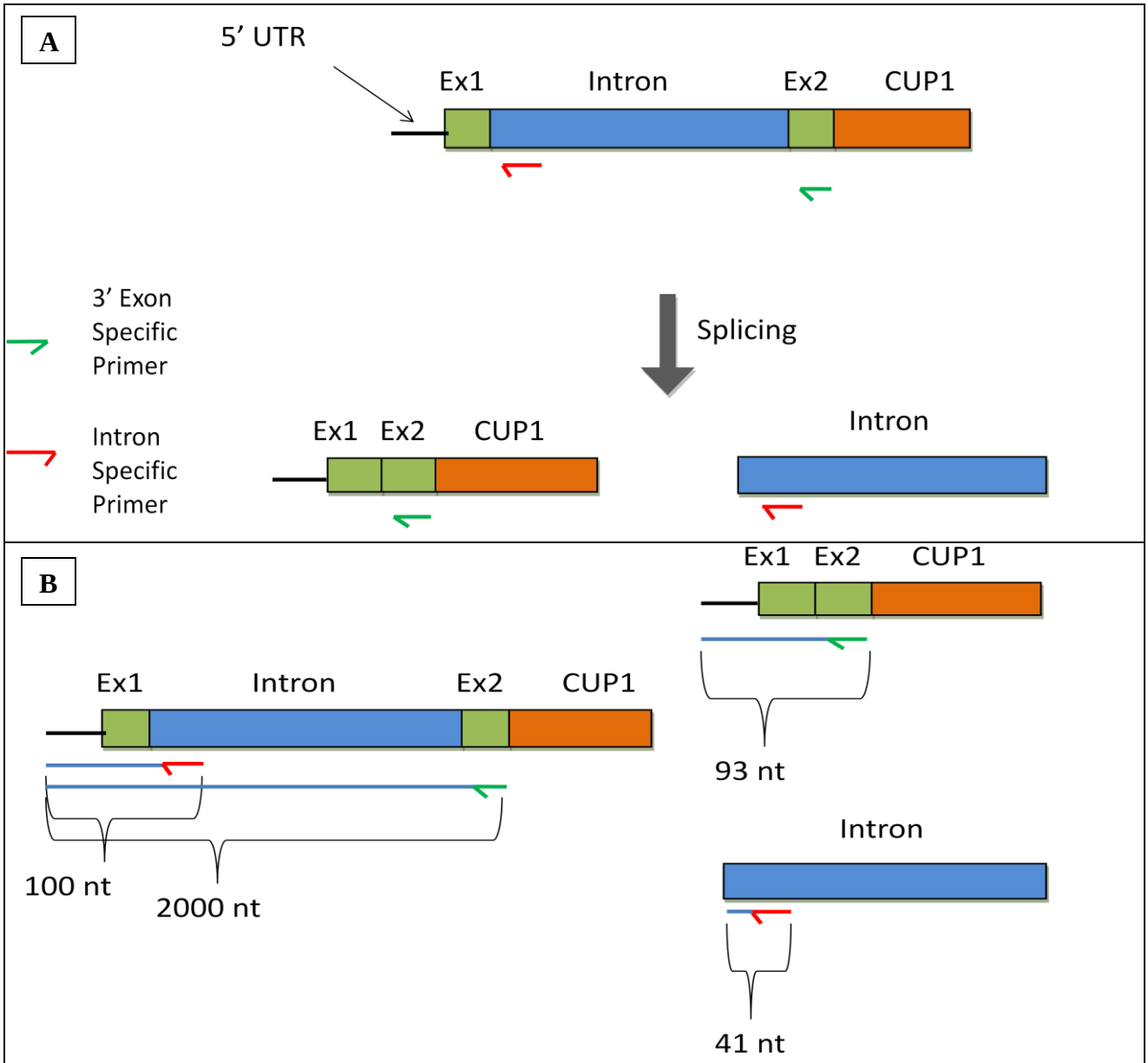

Figure 6—The reverse transcriptase primer extension schematic. (A) Shows the positions of the two primers in the pre-mRNA and the splicing products. (B) Shows the extended primers and their respective cDNA sizes for each primer and each expected splicing product.

the 10x reaction buffer, 2 µL of the polymerase mix, and was brought to a volume of 20 µL with water. The reaction was incubated at 37 ºC for three hours to promote RNA elongation, and treated with DNase (1 µL, then incubated at 37 ºC for 15 minutes) to degrade the template DNA. RNA purification through phenol-chloroform extraction was carried out to isolate the transcription products, and these products were spliced through *in vitro* splicing. Following the protocol previously described for BhI1 *in vitro*

splicing (Toor et al., 2006), RNAs from the BhI1-CUP1 and BhI1ΔORF-CUP1 constructs were first pre-folded in a thermal cycler: 1 min. at 95 ºC, 5 min. at 75 ºC, and 15 min. at 45 ºC. These pre-folded RNAs were then treated with a high-salt buffer: 2 M NH_4Cl, 100 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5). These mixtures were then incubated for 55 ºC for 15 min (Toor et al., 2006). RNA was then precipitated from these reactions with ethanol, sodium acetate (pH 5.2) and glycogen. The RNA pellets were washed and resuspended in water. As a negative control, BhI1-CUP1 and BhI1ΔORF-CUP1 RNAs were pre-folded but were treated only with water instead of salt solutions. A urea gel was run using the splicing products compared to their untreated negative controls.

In Vivo Splicing Assay in Yeast

Next, splicing of the BhI1-CUP1 construct *in vivo* was assayed. For IEP- yeast, glucose containing, trp- media was inoculated with the appropriate yeast strains and grown overnight. For IEP+ yeast, galactose containing, trp-, leu- media was inoculated. Cells were pelleted and phenol-chloroform extractions were done to purify yeast whole cell RNA. This RNA was then used in reverse transcriptase primer extension reactions to assay splicing. Samples used in *in vivo* assays were HS-CUP1, BhI1-CUP1, and BhI1ΔORF-CUP1, each with expression vector combinations as seen in Fig. 4.

Although IEP+ strains were grown in galactose-containing media, attempts to verify its presence via Western blot were inconclusive (data not shown).

Results

Cloning

The cloning strategy of producing BhI1-CUP1 and inserting into *S. cerevisiae* was successful. *E. coli* transformants were screened for pCSR-linker-BhI1-CUP1 by colony PCR and positives were picked up (Fig. 7, and data not shown). DNA sequencing was done for further confirmation. Sequencing came back positive, except for a single base pair mutation—the deletion of an adenine (1829delA)—in the domain V loop, which is not known to be involved in splicing (Toor et al., 2006).

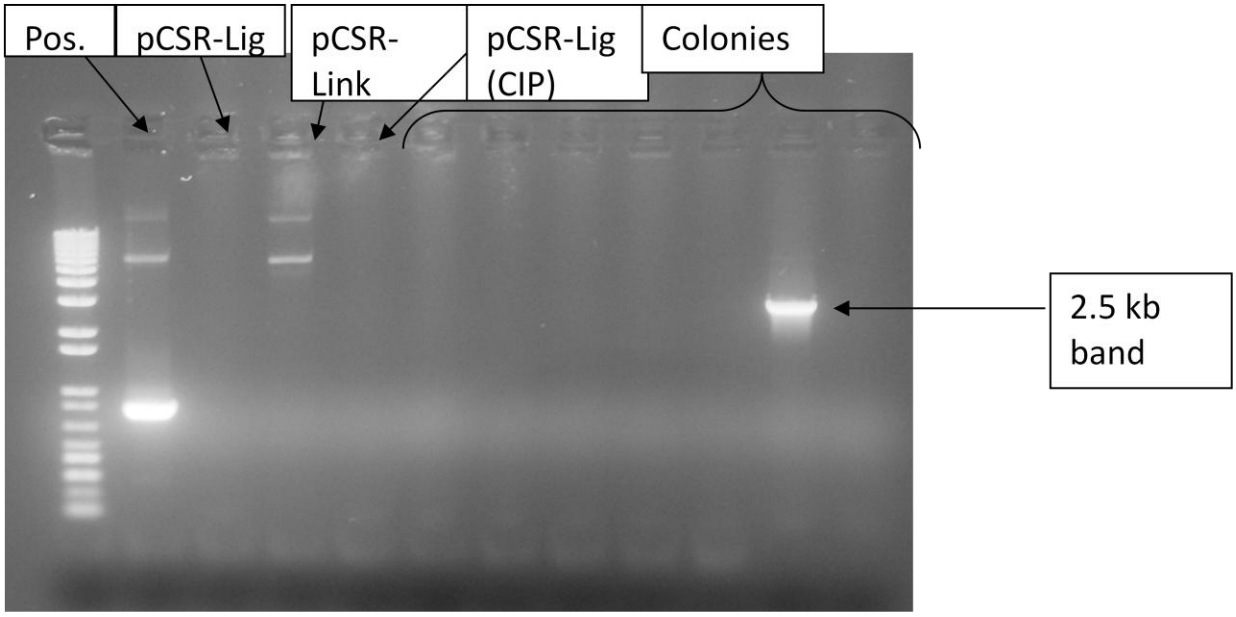

Figure 7—Representative colony PCR of *E. coli* transformed with pCSR-linker-BhI1-CUP1. The positive control used was a homing-site CUP1 (HS-CUP1), expected to be slightly greater than 500 base pairs. pCSR-Lig indicates the negative control colonies which were transformed with the empty pCSR vector after treated with ligase (the BamHI-SalI sites are incompatible). The pCSR-Link is the complete pCSR plasmid ligated to the BamHI-SalI linker. The pCSR-Lig (CIP) is the same as pCSR-Lig, but treated with calf intestinal alkaline phosphatase before ligation. One of the 16 colonies tested returned a positive result at 2.5 kb, the expected band length.

Phenotypic Assay for Copper Resistance

Yeast was successfully transformed with two isolates of the group II intron-reporter vectors, #6 and #25, as well as IEP expression vectors as the combinations seen in Fig. 4. Phenotype assay results are shown in Fig. 8. As expected, when the IEP was absent (IEP-) neither of the two isolates of BhI1-

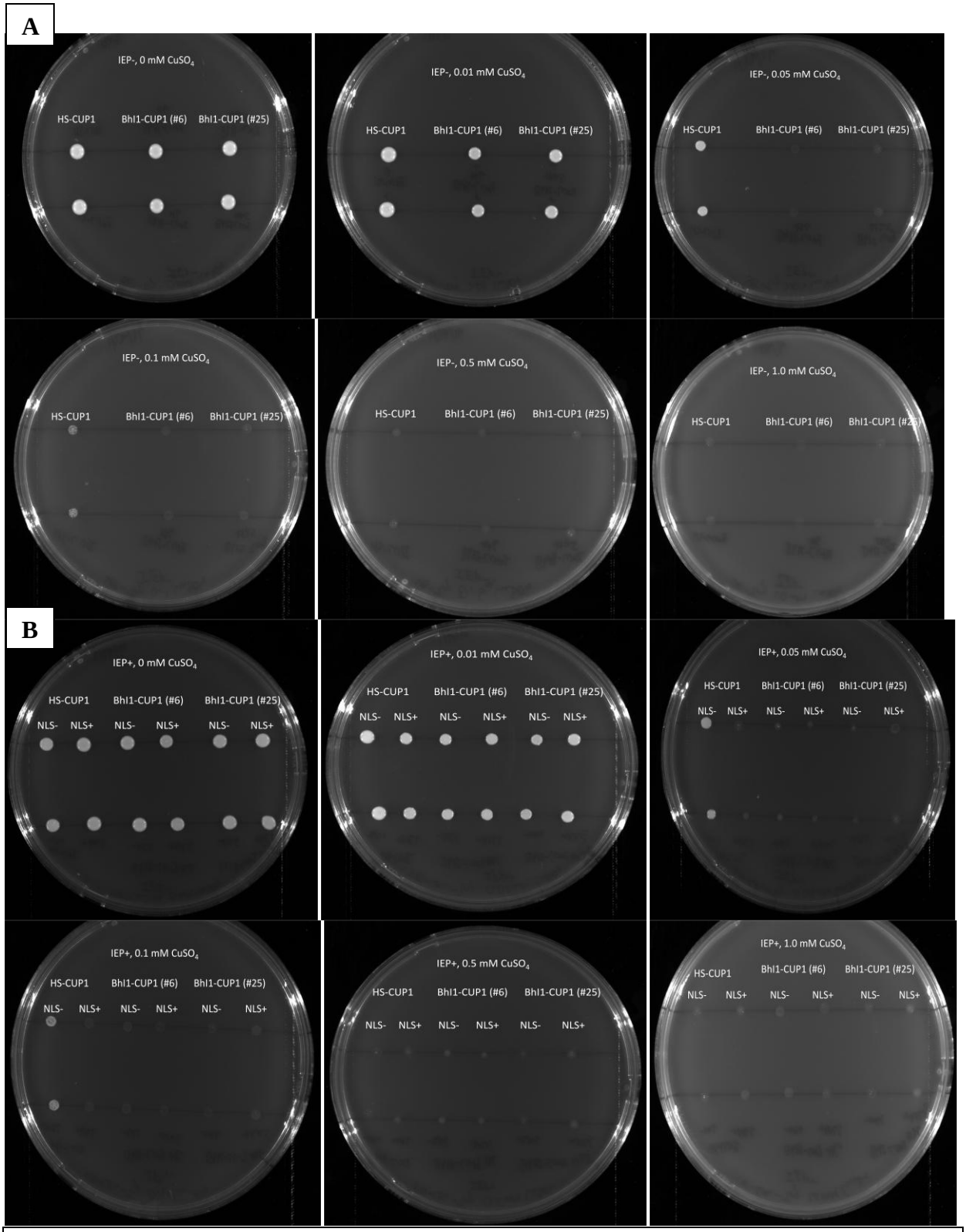

Figure 8—Copper resistance assays. Shows phenotype of BhI1-CUP1-containing yeast by spotting cells with (A) or without (B) an IEP expression vector.

CUP1 yeast grew in $CuSO_4$ concentrations over the threshold value of 0.01 mM, while HS-CUP1 yeast showed viability until $CuSO_4$ reaches 0.1 mM (Fig. 8A). However, this phenotype was not altered even when the IEP was provided. As shown in Fig. 8B, the BhI1-CUP1 IEP+ yeast strains did not show copper resistance, thus suggesting that the group IIC intron-containing host gene might be subject to silencing. As an ancillary result, HS-CUP1 yeast strains possessing the NLS version of the IEP (IEP+NLS+) expression vector experienced a decrease in the copper resistance phenotype.

In Vitro Splicing

Investigation of splicing of group II introns *in vitro* yielded results similar to those described by Toor et al. (2006). To visualize splicing products, a urea-denaturing polyacrylamide gel was run to separate in vitro-transcribed BhI1-CUP1 RNAs that had or had not been subjected to in vitro splicing. As shown in Figure 9, only one large, pronounced band appears, which must correspond to the unspliced pre-mRNA, was seen in the unreacted samples. In contrast, this band became dramatically decreased while multiple smaller bands appeared in the samples that had undergone the splicing reaction. Among them, those with sizes that are relatively smaller than the pre-mRNA but still very large must correspond to the excised intron product, which is linear for BhI1 as opposed to a lariat in IIA and IIB introns (Toor et al., 2006), and of variable intron-containing intermediates. Meanwhile, smaller sized products can be inferred to be ligated exons, and 5'exon (Toor et al., 2006), or perhaps 3' exon product, although this was not reported previously (Toor et al, 2006). Taken together, these observations indicate that splicing must have occurred.

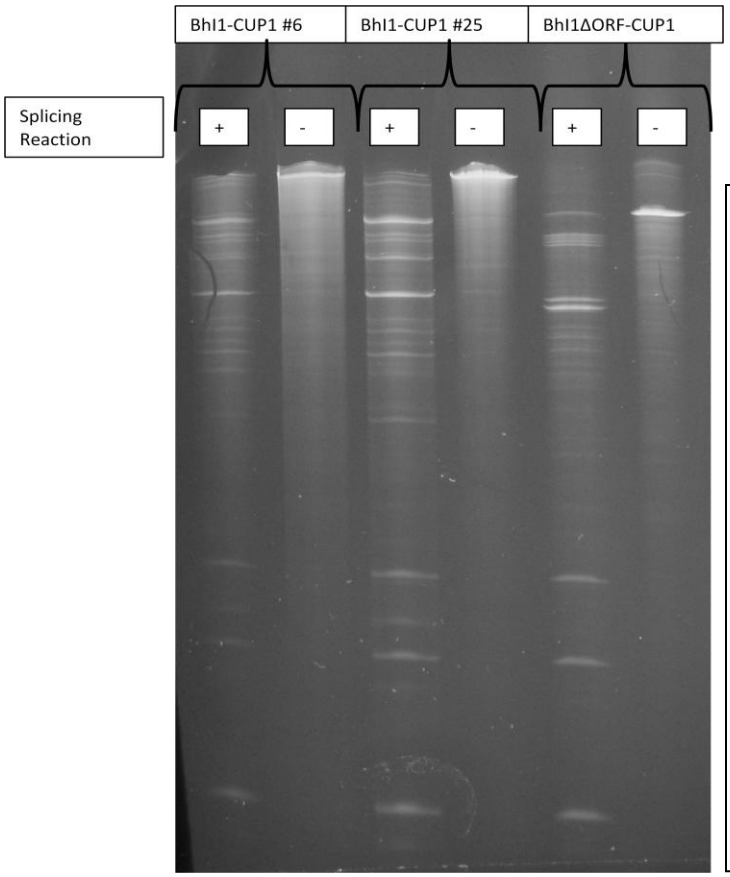

Figure 9—*In vitro* transcription and splicing. Shows differences between *in vitro* transcribed RNA treated with the *in vitro* splicing reaction versus the negative controls. Variable products can be seen, as expected from Toor et al., 2006. Differences between the BhI1-CUP1 and BhI1ΔORF-CUP1 products are seen in the upper part of the gel, where the excised intron is likely to be. This makes sense, as the only difference between the strains is the presence or absence of open reading frame of the IEP.

Primer extension allowed identification of some of the products in the RNA gel (Fig. 10): the 3' exon-specific primer showed a small, about 32 nucleotide cDNA, which would correspond to the 3' exon. Notably, ligated exons (expected at 68 nt) were absent. 3'exon-intron intermediate may be present, although it is difficult to determine with the resolution of this gel. A notable decrease in precursor (~2000 nt) is visible as well in the samples treated with the splicing reaction. Judging by the results of the intron-specific primer, excised intron or 3' exon-intron intermediate are clearly present (41 nt), and a decrease in precursor (75 nt) is visible. A smaller product, at 30 nt with the intron-specific primer, could correspond to cryptic cleavage about 10 nucleotides downstream of the 5' splice site (it is labeled as CSC Product in Figure 10, for Cryptic Site Cleavage Product). Results were similar for both the BhI1-CUP1 and BhI1ΔORF-CUP1 constructs. The presence of 3' exon products could be caused by spliced exon reopening (SER), a reverse splicing-like reaction conducted by the excised intron (Toor et al., 2006; Lambowitz & Zimmerly, 2011). The concentration of these products is highly dependent on salt

concentration and salt makeup. These results likely only occur *in vitro* and not *in vivo*, where the IEP's maturase is present to assure correct splicing of the intron (Toor et al., 2006).

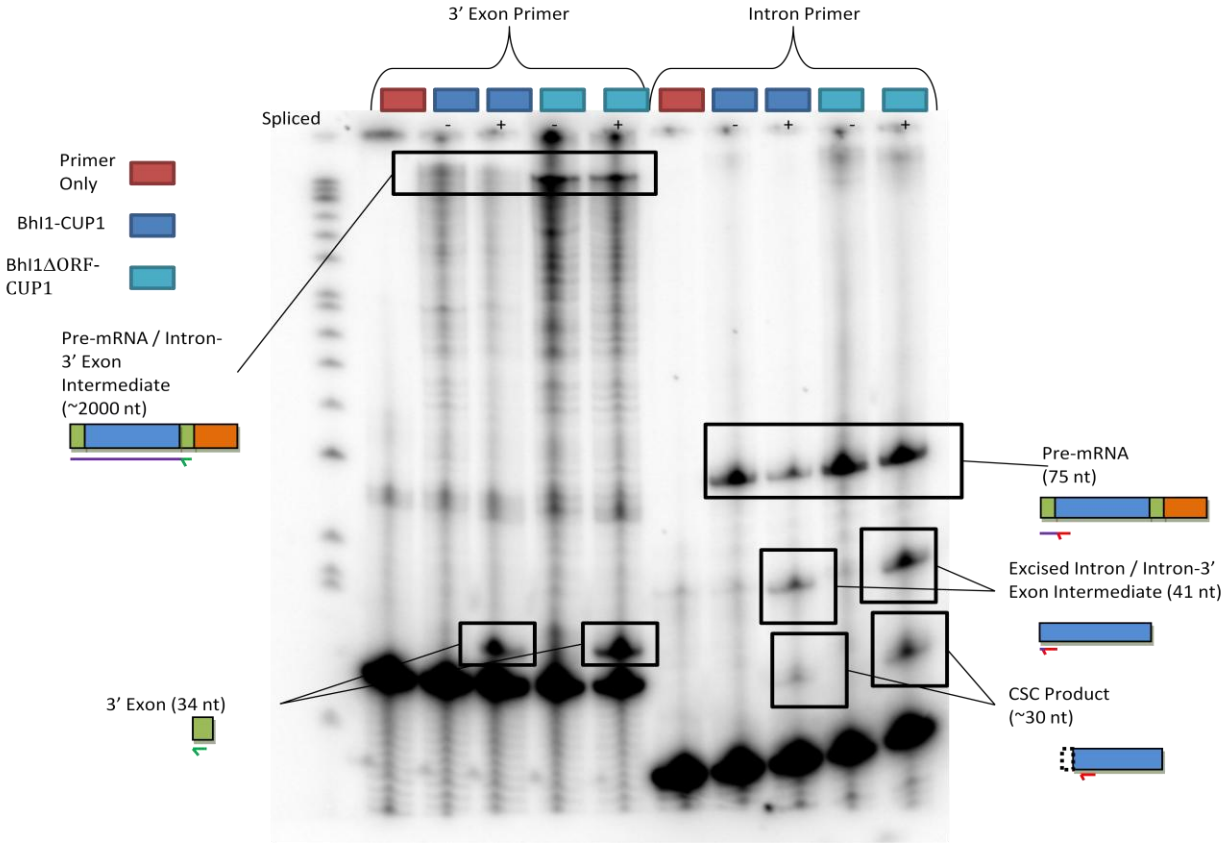

Figure 10—Primer extension assay to identify *in vitro* splicing products. Boxes are placed around products of interest, along with a schematic representation. The dark bands at the bottom of the gel represent excess primer. The 3' exon primer is 30 nt, and the intron primer is 21 nt. These atypical products are not expected *in vivo*, but are *in vitro*.

In Vivo Splicing

Primer extension was conducted to analyze *in vivo* splicing of the intron. Results from the 3' exon-specific primer (Fig. 11) show HS-CUP1 products representing ligated exons (93 nt) as expected. The additional size of the cDNA stems from 5' untranslated region (UTR) produced by the PGD promoter, and this size difference was noted in Qu et al. Supplemental, 2014. The ligated exon product was invisible in any of the BhI1-CUP1 samples regardless of the presence of the IEP. Bands corresponding to pre-mRNA transcripts are visible at the tops of the lanes, although they could alternatively represent intron-3' exon intermediate. These results seem to indicate that the BhI1 intron is

unable to splice in *S. cerevisiae* in the cellular environment but it can be transcribed. Of note is a smaller product within the HS-CUP1 lanes. This is an unexpected product whose exact nature is unknown.

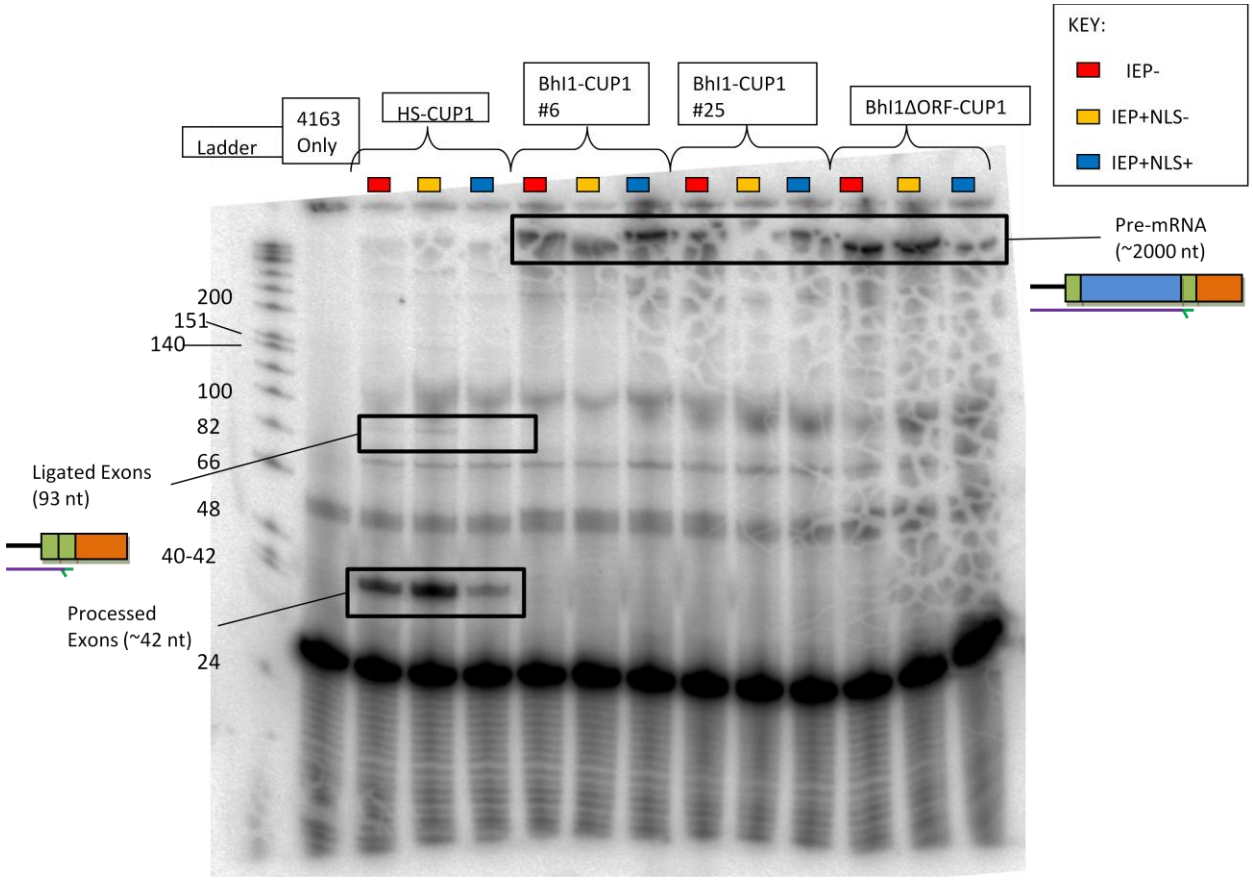

Figure 11—Primer extension analysis of RNA products in yeast using 3' exon specific primer. Shows the *in vivo* splicing products of BhI1-CUP1 constructs in *S. cerevisiae*. HS-CUP1 is a positive control, since it contains no intron. Presence or absence of IEP and/or NLS did not affect splicing. A smaller product is prevalent in the HS sample lanes, although its nature is unknown.

Primer extension using the intron-specific primer *in vivo* further elucidated the nature of BhI1 splicing in *S. cerevisiae* (Fig. 12). HS-CUP1 samples show no products, which is expected since no intron is present for primer to bind to. BhI1-CUP1 samples showed presence of pre-mRNA in each lane (100 nt for BhI1ΔORF-CUP1, 150 nt for BhI1-CUP1), as well as excised intron or intron-3' exon products, regardless of the presence of IEP (41 nt). The difference in expected cDNA sizes for pre-mRNA of the BhI1 constructs is due to the presence of the BamHI-SalI linker (Fig. 3, Fig 5A). Products slightly larger than the pre-mRNA bands were seen in the assay just as in Qu et al. Supplemental, 2014,

and could be due to processing of the pre-mRNA. Presence or absence of the IEP did not affect splicing products.

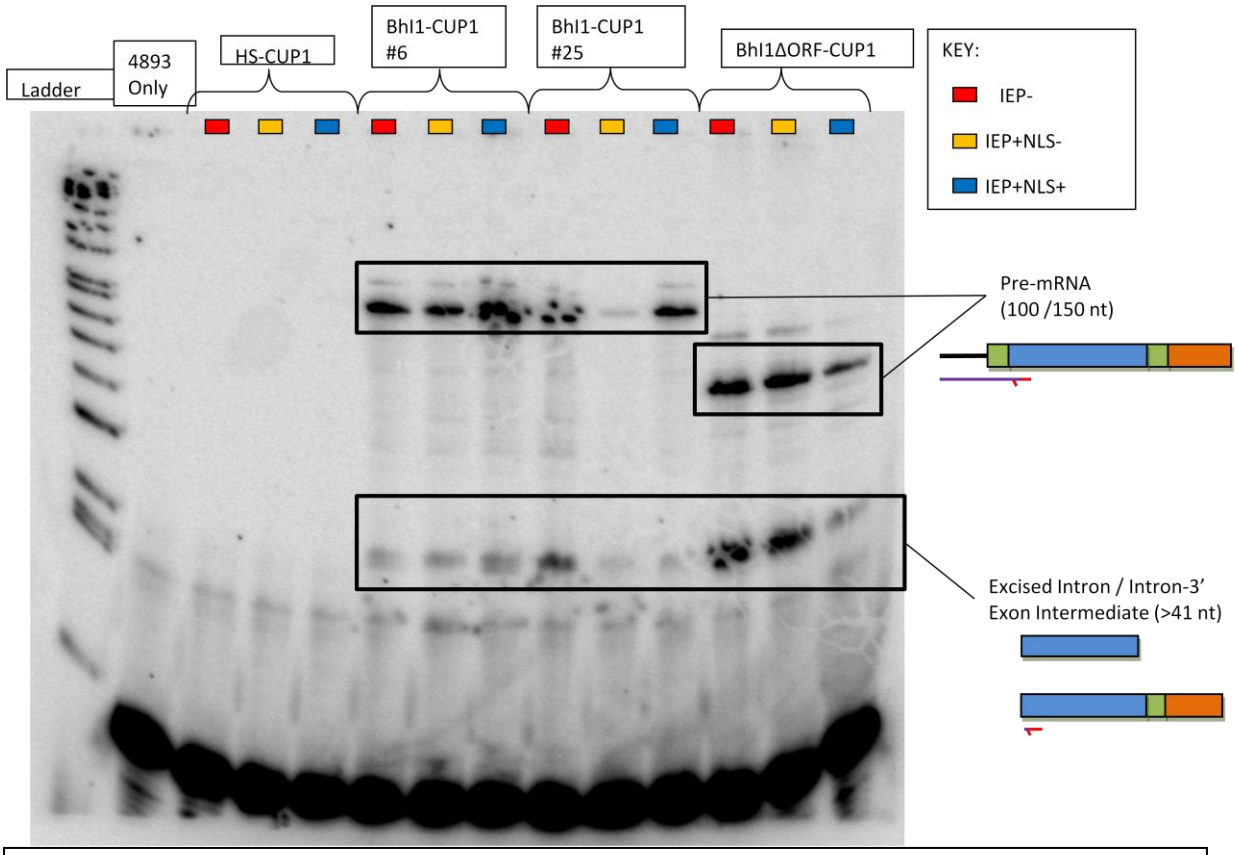

Figure 12—Primer extension analysis of RNA products in yeast using intron specific primer. Shows the *in vivo* splicing products of BhI1-CUP1 constructs in *S. cerevisiae*. HS-CUP1 is a negative control in this gel, since it contains no intron. Pre-mRNA is present in each BhI1-CUP1 lane. In addition, a product corresponding to excised intron or intron-3' exon intermediate is present. Presence or absence of IEP and/or NLS did not affect splicing.

Discussion

Conclusions

Based on the results shown here, it appears that group IIC introns silence their nuclear host genes just as group IIA and IIB introns do, albeit possibly using a different mechanism. It was shown that the BhI1-CUP1 cassette is not expressed in *S. cerevisiae*, unlike its intron-less counterpart, HS-CUP1. Splicing analyses of the intron RNA both transcribed *in vitro* and expressed in yeast shows that exon ligation does not occur, and that alternative products are produced instead, including cryptic site cleavage products. However, these products were anticipated based on the results of Toor et al., 2006 and do not necessarily preclude the ability to form ligated exons *in vivo* when the IEP is present. The complete absence of ligated exons after *in vitro* splicing is somewhat perturbing, but could be due to imbalances in salt concentrations, as the amount of products is highly salt dependent (Toor et al., 2006), or could be due to the difference in exon length in the two constructs, or the previously mentioned mutation in domain V. Primer extension of yeast cellular RNA shows that splicing produces either excised intron or intron-3' exon intermediate regardless of the presence of the BhI1 IEP, and that exon ligation does not occur either. A problem with the construct could be causing the lack of exon ligation both *in vitro* and *in vivo*. This could also be due to a eukaryote-specific silencing mechanism, similar to the host surveillance pathways that silence group IIA and IIB introns.

Rationale

Group II introns are widely accepted as the evolutionary precursors to spliceosomal introns. Martin and Koonin (2006) claimed that the mitochondrial endosymbiosis led to invasion of the host genome by group II introns, necessitating the evolution of the nuclear membrane and the beginning of the eukaryotic lineage. Group IIA and IIB introns introduced into *S. cerevisiae* can transcribe and splice, but are subject to translational repression and cytoplasmic mislocalization. We show that in contrast to these observations, group IIC introns likely cannot splice completely *in vivo* in *S. cerevisiae*. The reason for

this requires further elucidation, but could represent evolutionary differences between the group II intron subclasses.

Group II introns, in the form of degenerate group IIA and IIB subclasses, are found in mitochondrial genomes today. Could it be that group IIC introns represent a more distal clade of mobile retroelement, one that never invaded eukaryotic genomes? Group IIA and IIB introns may have evolved into spliceosomal introns, and eukaryotes may have evolved to allow their splicing but inhibit their expression through host surveillance mechanisms. Group IIC introns, on the other hand, would have never been found in eukaryotic genomes, and as such are unable to be spliced efficiently post-transcription. Alternatively, group IIC introns may have invaded alongside their IIA and IIB relatives, but due to differences in their structure and properties were silenced via a different mechanism.

Group II introns are notable for the part they have likely played as both the ancestors of eukaryotic nuclear spliceosomal introns as well as mobile elements such as retrotransposons. Their proliferation to every domain of life, save for their absence from eukaryotic nuclear genomes, is perplexing. Group IIA and IIB introns, when inserted into *S. cerevisiae*, subject their nuclear host genes to silencing via translational repression and cytoplasmic mislocalization. This study shows that group IIC introns, whose reduced EBS-IBS interactions distinguish them from their IIA and IIB relatives, may be silenced via a distinct post-transcriptional mechanism that inhibits splicing. This information will assist in establishing the evolutionary narrative of the expulsion of group II introns from nuclear genomes and the evolution of group II intron ancestrally-related spliceosomal introns.

Future Experiments

The experiments performed here do not convey the conclusions reached beyond a reasonable doubt. Western blotting to determine the presence of the BhI1 IEP should be done to verify that the lack of efficient and complete splicing in *S. cerevisiae* is due to the eukaryotic cell environment, and not due to lack of maturase. Primer extension experiments both *in vivo* and *in vitro* could be conducted with a greater number of primers to assure a complete picture of the splicing products. A 5' exon-specific

primer would have been beneficial. Running cleaner, higher resolution gels could have contributed to clarity of results and conclusions drawn.

In addition, either short exons or the single base pair mutation shown by sequencing in domain V may contribute to the lack of proper splicing *in vivo* as well as the apparent lack of ligated exons *in vitro*, even though they have not been directly shown to be involved in group IIC intron splicing. This work could attest to the domain V loop's involvement in proper intron splicing. Site-directed mutagenesis of this base pair to its native state or extending exon length followed by repetition of experiments could confirm that this is not the cause of the phenomena observed in this study. Alternatively, testing this aberrant BhI1-CUP1 construct in a bacterial host such as *E. coli* may be beneficial to ascertain that the construct can splice properly *in vivo*.

Provided our experiments do in fact show that group IIC introns are unable to splice properly *in vivo* in yeast, further experiments should probe the factors contributing to this effect. Using directed evolution studies through error-prone PCR, a library of BhI1 mutants could be made and transformed into yeast. Alleviated silencing followed by intron sequencing could identify sequences within the intron contributing to the silencing pathway. A mutagenesis assay to knockout yeast nuclear genes and screen for alleviated silencing could also be conducted. Further study into the exact nature of the silencing mechanism via exploration of likely involved pathways should be done as well.

References

Beauregard, A.; Curcio, M. J.; Belfort, M. The take and give between retrotransposable elements and their hosts. *Annu. Rev. Genet.* [Online] **2008**, 42, 587-617.

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

Koonin, E. V.; Martin, W. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* [Online] **2006**, 440, 41-45

Lambowitz, A. M.; Belfort, M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiol. Spectrum* [Online] **2015**, 3(1):MDNA3-0050-2014

Lambowitz, A. M.; Zimmerly, S. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb. Perspect. Biol.* [Online] **2011**, 3:a003616

Qu, G.; Dong, X.; Piazza, C. L.; Chalamcharla, V. R.; Lutz, S.; Curcio, M. J.; Belfort, M. RNA-RNA interactions and pre-mRNA mislocalization as drivers of group II intron loss from nuclear genomes. *PNAS* [Online] **2014**, 111, 6612-6617.

Toor, N.; Robart, A. R.; Christianson, J.; Zimmerly, S. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* [Online] **2006**, 34, 6461-6471.

Zimmerly, S.; Semper, Cameron. Evolution of group II introns. *Mobile DNA* [Online] **2015**, 6:7