Genetics: Published Articles Ahead of Print, published on November 30, 2011 as 10.1534/genetics.111.136267

Submitted to: **Genetics**

Transcriptional regulatory circuitries in the human pathogen *Candida albicans* involving sense-antisense interactions

Ausaf Ahmad*, Anatoliy Kravets† and Elena Rustchenko†[1]

**Department of Pathology and †Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, New York 14642*

Running title: Long embedded antisense transcripts contribute to utilization of sorbose

Keywords: antisense regulation, sorbose utilization, chromosome 5, *Candida albicans*

[1]*Corresponding author*: Elena Rustchenko, Department of Biochemistry and Biophysics, Box 712, University of Rochester Medical Center, Rochester, NY 14642, USA. Phone: (585) 275-6710. Fax: (585) 275-6007. E-mail: elena_bulgac@urmc.rochester.edu.

Text 33 pages; Figures 1-6. Table 1. Supporting information: Text 4 pages; Figures S1-S5; Tables S1, S2.

Re-submit October 28, 2011

ABSTRACT

Candida albicans, a major human fungal pathogen, usually contains a diploid genome, but controls adaptation to a toxic alternative carbon source L-sorbose, by the reversible loss of one chromosome 5 (Ch5). We have previously identified multiple unique regions on Ch5 that repress the growth on sorbose. In one of the regions, the *CSU51* gene determining the repressive property of the region was identified. We report here the identification of the *CSU53* gene from a different region on Ch5. Most importantly, we find that *CSU51* and *CSU53* are associated with novel regulatory elements, *ASU*s, which are embedded within *CSU*s in an antisense configuration. *ASU*s act opposite to *CSU*s by enhancing the growth on sorbose. In respect to the *CSU* transcripts, the *ASU* long antisense transcripts are in lesser amounts, are completely overlapped, and are inversely related. *ASU*s interact with *CSU*s in natural *CSU/ASU cis* configurations, as well as when extra copies of *ASU*s are placed in *trans* to the *CSU/ASU* configurations. We suggest that *ASU* long embedded antisense transcripts modulate *CSU* sense transcripts.

INTRODUCTION

Genome-wide surveys of eukaryotic transcriptoms have led to the identification of abundant non-coding transcription in plants, mammals (LEE *et al.* 2009), and fungi (DAVID *et al.* 2006; SAMANTA *et al.* 2006; NAGALAKSHMI *et al.* 2008; DUTROW *et al.* 2008; SELLAM *et al.* 2010). The non-coding transcriptom of higher eukaryotes includes a class of various small RNAs that were extensively studied and that were shown to directly modulate gene expression (reviewed by LEE *et al.* 2009; TAFT *et al.* 2010; OLEJNICZAK *et al.* 2010). It was not initially appreciated that another class of long non-coding RNAs appears in 10-12-fold larger amounts than the coding

transcripts (NAGANO and FRASER, 2011). Initially neglected, the long non-coding RNAs are now given much attention, being recognized as important regulators that are implicated in a large range of various functions, including epigenetic control, enhancing or mediating long-range chromatin interactions, as well as serving as scaffolds of chromatin-modifying complexes (HONGAY *et al.*, 2006; CAMBLONG *et al.* 2007; MAHMOUDI *et al.* 2009; GUTTMAN *et al.* 2009; KHALIL *et al.* 2009; HOUSELY *et al.* 2008; reviewed by MORRIS, 2009; NAGANO and FRASER, 2011).

Characterization of transcription in *Candida albicans*, which is considered to be the most common fungal opportunistic pathogen of humans, is in its infancy. RNA interference (RNAi) in this organism has been recently discovered (DRINNENBERG *et al.* 2009). First reports on the transcriptom have been published, revealing an extensive antisense transcriptom (BRUNO *et al.* 2010; SELLAM *et al.* 2010; TUCH *et al.* 2010).

C. albicans is a single cell organism with a diploid genome organized into eight pairs of chromosomes. Survival in various adverse environments is an important property of this pathogen. Integral parts of *C. albicans* adaptation and survival strategies are alterations of large portions of genome including monosomy or trisomy of entire chromosomes. Similar alterations of same chromosomes occur in same environments, allowing survival (RUSTCHENKO, 2007; 2008). A well-known example, which is the topic of this paper, is survival on the toxic sugar L-sorbose, when it is available as a sole source of carbon. The loss and gain of a chromosome 5 (Ch5) up and down regulates, respectively, the *SOU1* (SOrbose Utilization) gene (orf19.2896) on Ch4 and confers growth, Sou^{+}, and no growth, Sou^{-}, on sorbose (RUSTCHENKO and SHERMAN, 2002; RUSTCHENKO, 2007; 2008).

Our early studies already indicated the complexity of regulation by copy number of Ch5. This chromosome carries multiple unique regions for negative control of growth on sorbose; the final number of regions is yet to be established (KABIR *et al.* 2005). A total of five regions A, B, C, 135, and 139 (Figure 1) have been rigorously confirmed by an analysis of Ch5 deletions. The regions were proposed each to encompass at least one unique negative controlling element *CSU* (Control of Sorbose Utilization). The first *CSU*, *CSU51* (orf19.1105.2), has been identified in region A (KABIR *et al.* 2005). Another *CSU*, *CSU53* (orf19.3931) from region 135, is presented in this work.

We report here a new genetic element *ASU* (Activation of Sorbose Utilization), adding an additional layer of control from Ch5. *ASU*s are embedded in *CSU51* and *CSU53* in the opposite orientation and are associated with a distinct, albeit weak, over expression phenotype of the enhanced growth on sorbose, thus, counter-acting the repressive phenotype by *CSU*s. Antisense *ASU* transcripts are long, can be capped and polyadenylated, and seem to act as non-coding transcripts. We present evidence that, as expected, complementary sense *CSU* and antisense *ASU* transcripts interact. The final number of the *CSU*/*ASU* configurations on Ch5 has yet to be determined.

MATERIALS AND METHODS

The co-over expressing system: We used an important tool, a low copy number replicative plasmid pCA88 over expressing the metabolic gene *SOU1*, thus, causing Sou$^-$ recipient strain to utilize sorbose, Sou$^-$ → Sou$^+$ (WANG *et al.* 2004). This plasmid was previously used for preparing Ch5 DNA library and subsequently cloning a negative regulatory gene *CSU51*, as well as unique regions carrying other putative *CSU*s on Ch5, as based on the reversal of sorbose

utilization, Sou$^+$ → Sou$^-$ (KABIR *et al.* 2005). In this work, similarly, we used pCA88 to co-over express *SOU1* and different Ch5 sequences, as presented with cartoons in Figure 2. Care was taken to assure that inserted genes had up to 1.5 kb of the upstream regions, as *C. albicans* is known to have long promoters (GAUR *et al.* 2004; VINCES *et al.* 2006; SRIKANTHA *et al.* 2006). In addition, genes contained approximately 100 bp of the downstream region.

Strains, media, plasmids, and primers: We used the *C. albicans* Sou$^-$ sequencing strain SC5314 and its relatively genetically stable Sou$^-$ Ura$^-$ derivative CAF4-2 (AHMAD *et al.* 2008). Also, we used the well-characterized prototrophic Sou$^-$ strain 3153A (RUSTCHENKO-BULGAC and HOWARD 1993). The strains carrying in their genome either a control empty vector pAK156 or a vector with one *ASU53*, pEA249, or two *ASU53*s, pEA254, were prepared by individually integrating the plasmids into *LEU2* locus on Ch7.

Yeast extract/peptone/dextrose (YPD) and synthetic dextrose (SD) media were previously described (SHERMAN, 2002). 1 M sorbitol was added in SD medium, when growing transformants. Synthetic sorbose or sorbitol media were the same as SD medium, but contained 2% of either L-sorbose or sorbitol, as a sole carbon source (RUSTCHENKO *et al.* 1994). To prepare solid medium, 2% (wt/vol) agar or agarose was added. Uridine (50 μg/ml) was added when needed. The proper growth and handling of cells preventing chromosomal instability was previously reported (RUSTCHENKO-BULGAC 1991; PEREPNIKHATKA *et al.* 1999; WANG *et al.* 2004; AHMAD *et al.* 2008).

All plasmids or primers that were used in this study are described in supporting information and also are presented in Tables S1 and S2, respectively. All replicative plasmids are derivatives

of pCA88 (see "The co-over expressing system" in Results), which is pRC2312 carrying *SOU1* (WANG *et al.* 2004).

RNA preparation and Northern blot analysis: *C. albicans* cells were grown for independent colonies at 37° C on plates with sorbitol medium (see above) and opened with glass beads. Total RNA was isolated according to RUSSO *et al.* (1991) or with RNeasy® Midi kit (QIAGEN Sciences, MD) and additionally treated with RNase free DNase to remove all traces of genomic DNA. mRNA was isolated from total RNA using Oligotex® mRNA Kit (QIAGEN Sciences, MD), as recommended by the manufacturer. Either 15 μg of total RNA or 3-4 μg of mRNA were denatured and size fractioned on 1% formaldehyde gel and then transferred to positively charged nylon membrane from Ambion (Austin, TX) according to RUSSO *et al.* (1991). For loading controls, blots were hybridized with double stranded (ds) DNA probes that were prepared from the *PGK1* gene (orf19.3651) or 18S rRNA and labeled with ^{32}P (RUSSO *et al.* 1991). ^{32}P-labeled strand-specific single-stranded RNAs (riboprobes) were generated by *in vitro* transcription with "MAXIscript® T7" kit (Ambion, Austin, TX) according to the manufacturer's instructions. Riboprobes were hybridized with blots according to CLEMENTS *et al.* (1988). Briefly, blots were incubated in QuikHyb Hybridization solution from Stratagene (La Jolla, CA) overnight at 68° C and washed with low and high stringency solutions at room temperature and at 68° C, respectively. Images were processed using PhosphorImager Storm-820 (Molecular Dynamics, Sunnyvale, CA). Individual bands were quantified using ImageQuant 5 software (Molecular Dynamics, Sunnyvale, CA). Transcript sizes were estimated with RNA Millenium™ Size Markers–Formamide (Ambion, Austin, TX). General approaches for performing Northern analyses were adopted from DING *et al.* (2007). For example, each experiment was repeated

several times with different batches of RNA that were prepared from independently grown cultures.

RT-PCR and semi-quantitative (s.-q.) analysis: In order to synthesize strand specific cDNA, we set the RT reaction as a duplex with gene specific primers for a gene of interest and a control gene. The following genes were used as controls: *CDC6* (orf19.5242); *EMP24* (orf19.6293); *TPK2* (orf19.2277) *PGK1* (orf19.3651). Synthesis was conducted with the MonsterScript Reverse Transcriptase (Epicentre Biotechnologies, Madison, WI), as recommended by the manufacturer. The PCR amplification was also set as a duplex and was conducted for different number of cycles with Phusion Hot Start High-Fidelity DNA Polymerase (New England BioLabs, Ipswich, MA), as recommended by the manufacturer (also see KUAI *et al.* 2004; CERAZIN-LEROY *et al.* 1998). Prior to that, pilot duplex PCR amplifications were undertaken in order to optimize reaction conditions for efficiency of amplification and the lack of non-specific bands.

S.-q. RT-PCR analysis was performed according to KUAI *et al.* (2004) and CERAZIN-LEROY *et al.* (1998). Briefly, images of ethidium bromide stained gels loaded with amplification products from different number of cycles were prepared with the AlphaImager IS2000 Digital Imaging System (Alpha Innotech Corp., San Leandro, CA). ImageQuant 5 software (Molecular Dynamics, Inc., Sunnyvale, CA) was used to measure the brightness of the bands (AHMAD *et al.* 2008). At least three consecutive amplicons in exponential phase, as estimated by the determination coefficient value 0.97 or more were used to normalize the experimental values against the values of the control genes and then to calculate the ratios test/control genes from the averaged values.

Mapping of 5'- or 3'-untranslated region (UTR) with rapid amplification of cDNA ends (RACE): Total RNA was used to analyze 5' and 3' UTRs by RACE with the gene-specific primers and with the "FirstChoice RLM RACE" kit (Ambion, Austin, TX) according to the manufactures' specifications. The RACE products were electrophoretically separated on agarose gel, purified and subsequently ligated into pJET1.2/Blunt vector (Fermentas Life Sciences, Glen Burnie, MD) for further transformation into *E. coli* 5-alpha competent cells supplied by New England BioLabs Inc. (Ipswich, MA). Plasmids from individual transformants were analyzed for the presence of inserts with expected size and sequenced using BigDye Terminator v3.1 from Applied Biosystems (Foster City, CA).

Assay for the Sou phenotype: Spot dilution assay was performed on solid sorbose medium (see above), as described by WELLINGTON and RUSTCHENKO (2005).

Handling transformants carrying replicative plasmids for the growth assay and for RNA isolation: Several transformants grown as colonies on solid SD medium supplied with sorbitol were combined and streaked as patches on solid SD medium. After incubation, some cells were taken for sorbose growth assays, whereas the other cells were suspended in distilled water, plated on solid SD medium for independent colonies, incubated, colonies grown, harvested, and total RNA isolated (see above).

Miscellaneous: Amplicons or plasmid inserts were routinely sequenced in Core Facility at the University of Rochester using BigDye Terminator v3.1 Cycle Sequencing kit on ABI 3730

PRISM Genetic Analyzer. Transformation of *C. albicans* cells was conducted according to (KABIR and RUSTCHENKO, 2005). Site-directed mutagenesis was performed with a QuickChange Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA) according to the manufacturer's recommendations.

RESULTS

Identification of the *CSU53* gene in region 135 of Ch5: In this work, we continue characterizing five well-established regions on Ch5 that are involved in the repression of the growth on sorbose (Introduction, Figure 1). One of the approaches used here is the replicative test-plasmid pCA88 carrying metabolic *SOU1*. As previously reported (see also section "The co-over expressing system" of MATERIALS AND METHODS), this plasmid confers growth on sorbose medium to the Sou^- recipient cells, $Sou^- \rightarrow Sou^+$ (cartoons 1 and 2 in Figure 2). However, the introduction of the 4.3 kb portion of Ch5 that is designated region 135 (Figure 1) into pCA88 results in the plasmid pCA135 (cartoon 3 in Figure 2) that shifts the Sou^+ phenotype of the recipient cells due to *SOU1* back to no-growth, $Sou^+ \rightarrow Sou^-$ (KABIR *et al.* 2005). Compare the repressive Sou^- phenotype due to region 135 with the control Sou^+ phenotype due to *SOU1*, as shown with the spot assay in Figure 3A (see MATERIALS AND METHODS for the assay and Figure S1 for the control growth on glucose medium). Analysis of the sequence of region 135 indicated a single large ORF of 912 bp previously annotated, as the *SFC1* (orf19.3931) gene. We demonstrated that this ORF is implicated in the repression of growth on sorbose, as presented below, and, thus, designated this gene as *CSU53*. In Figure 4A showing the cartoon of region 135, *CSU53* ORF is presented with a gray box.

In order to evaluate whether *CSU53* is relevant to the Sou phenotype, we interfered with *CSU53* translation product by creating three independent frame-shift mutations in region 135 within the *CSU53* ORF at positions +66, +112 or +233, as indicated with "stars" in Figure 4A (see supporting information for details). The mutations were verified by sequencing and each mutated region 135 was individually subcloned into pCA88 resulting in, respectively, plasmids pEA227, pEA158 and pEA201. Note that in Figure 4A the plasmid names are given to the corresponding inserts, a nomenclature that will be kept throughout the text. CAF4-2 cells were individually transformed with the above plasmids and tested for the growth on sorbose with the spot assay. For the handling of *C. albicans* transformants, see MATERIALS AND METHODS. Each mutation consistently abolished the repressive Sou$^-$ phenotype of region 135 and restored the control Sou$^+$ phenotype due to *SOU1*, as presented in Figure 3B (compare pEA227, pEA158 and pEA201 with pCA135 and pCA88; also see Figure S2A for the control growths on glucose medium). These data strongly implicated the *CSU53* (*SFC1*) gene with the repression of growth on sorbose and indicated that the putative Csu53p is important for the Sou phenotype.

In order to produce other evidence of the repressive function of C*SU53*, we interfered with its transcription. Portions of region 135 encompassing *CSU53* ORF with various sizes of the sequence in front of the start codon, were PCR-amplified, individually co-over expressed with *SOU1* and tested for growth on sorbose, as above. We found that diminishing the sequence upstream to the ORF led to diminution of the repressive property. For example, an entire upstream sequence of approximately 2 kb, pEA144, rendered the same repressive phenotype, as the entire region 135 (Figures 3A, bottom, and 4A; also cartoon 4 in Figure 2). Diminishing the upstream sequence to 1.4 kb, pEA105, or to 0.5 kb, pEA143 (Figure 4A) decreased the repression, as indicated by multiple Sou$^+$ colonies, Sou$^-$*, that appeared after prolonged

incubation (Figure 3A, top or bottom, respectively). On the other hand, various portions of region 135 lacking *CSU53* ORF, as, for example, pEA145, pEA146, pEA141, pEA140, pEA130, pEA132, or pEA133 (Figure 4A) consistently displayed the control Sou^{+} phenotype (see Figure 3A, bottom, for representative pEA145 and pEA140), thus, demonstrating no relevance to the repressive property. These results provided clear evidence for a single *CSU53* in region 135.

Identification of new elements, *ASU*s, which are associated with *CSU*s, and which act opposite to *CSU*s by enhancing the growth on sorbose: By co-over expressing *SOU1* and different portions of region 135, as described above, we found a phenotype that was different from the repressive Sou^{-} or the control Sou^{+}. This phenotype was an increased growth on sorbose medium, Sou^{++}, that occurred, when the region upstream to *CSU53* was removed, thus, resulting in pEA104, pEA156, and pEA155 (Figure 4A), as exemplified with the growth assay of pEA104 in Figure 3A (top).

We addressed the question whether the Sou^{++} phenotype depends on the insert orientation toward *SOU1* by preparing plasmid pEA234, which carried a 2,158 bp insert from plasmid pEA104 in opposite orientation (see supporting information and Table S1 for the pEA234 construction). As exemplified in Figure 3C, plasmids with different orientation of the insert, pEA104 and pEA234, rendered the same increase of growth, which was independent of insert orientation, in multiple experiments. See Figure S3A for the control growth on glucose medium.

We next addressed the question whether the Sou^{++} phenotype depends on the large insert acting, for example, as a stabilizing factor. We prepared and analyzed plasmid pEA261 carrying an insert of the same size, 2,158 bp, as the above pEA104, encompassing the sequence upstream

to *CSU53* with an adjacent 169 bps extending to outside the region 135, Figure 4A. Unlike pEA104, Sou^{++}, new pEA261 conferred the control Sou^{+} phenotype.

It did not seem that the Sou^{++} phenotype was an artifact due to a certain sequence, as various shorter portions encompassing the region downstream to *CSU53* alone or with a portion of the *CSU53* ORF, such as pEA145, pEA146, pEA141, pEA140, and pEA130 (Figure 4A) showed the control Sou^{+} phenotype (see Figure 3A , bottom, for the representative pEA145 and pEA140).

Importantly, all three portions conferring the Sou^{++} phenotype, pEA155, pEA156, and pEA104 contained a relatively small ORF of 120 bp embedded in *CSU53* ORF of 912 bp in opposite orientation, 461 bp downstream from the start codon (Figure 4A, a black box). Apparently, there is a critical 20 bp region, which causes the difference between pEA155 (Sou^{++}) and pEA142 (Sou^{+}). This critical region lies in front of a putative transcription start site of the embedded ORF. We, thus, designated the ORF embedded in *CSU53* of region 135, as *ASU53* (Activation of Sorbose Utilization).

Similar to *CSU53* from region 135, we previously reported that *CSU51* from region A, Figure 1, also has a smaller ORF of 105 bp embedded in opposite orientation, 32 bp downstream from the start codon (KABIR *et al.* 2005) (Figure 4B). Also similarly, the authors introduced frame-shift mutations in region A to target either *CSU51* or the embedded ORF and demonstrated that the embedded ORF lacked the repressive property. Importantly, when the repressive property of *CSU51* was abolished, the phenotype shifted to Sou^{+}, *i.e.*, no Sou^{++} growth occurred. In this work, we changed the approach and analyzed portions of region A, instead of introducing mutations in the intact region. We found that the portions pEA219 of 792 bp, pEA221of 802 bp, and pEA209 of 404 bp encompassing *CSU51* with the embedded ORF,

but lacking the region upstream to the *CSU51* ORF, rendered, as expected, the Sou^{++} phenotype (Figures 4B and S3B). Increased growth was independent of the insert orientation, as shown with plasmid pEA236 carrying the same insert as in pEA209, but in opposite orientation (Figure S3B). We, thus, designated the ORF embedded in *CSU51* of region A, as *ASU51*, by analogy with *ASU53* of region 135.

We confirmed the Sou^{++} phenotype of *ASU51* by trimming the region upstream to its ORF to 20 bp, a portion pEA240 (Figure 4B) that shifted the phenotype Sou^{++} → Sou^{+}. We also extended the region downstream to the *ASU51* ORF from a portion pEA209 (Sou^{++}) into the region upstream to the *CSU51* ORF, a portion pEA183, which shifted the phenotype Sou^{++} → Sou^{+} (the phenotypes are indicated on the schematics in Figure 4B). The 44 bp sequence downstream to the *ASU51* ORF on pEA209 or 75 bp on pEA183 included, respectively, 11 bp or 44 bp from the region upstream to *CSU51*. Sequence analysis of the critical 33 bp difference between pEA209 (Sou^{++}) and pEA183 (Sou^{+}) revealed a putative TATA box between position -12 to -15 in the 5'-UTR of *CSU51*, suggesting that this TATA box was sufficient for some transcriptional activity of *CSU51* that prevailed over the weaker Sou^{++} phenotype.

Importantly, the Sou^{++} phenotype consistently occurred in multiple independent experiments, as well as in the series of experiments that were conducted at 37° C, 30° C, and 22° C, as established with pEA104 or pEA234 carrying *ASU53* or with pEA104 or pEA234 carrying *ASU51*. This phenotype could be clearly observed within approximately the first three to five days of incubation, until the control cells caught up with the growth, thus, obscuring a convenient comparison. At 37º C, the phenotype due to *ASU51* was not as pronounced, as those at two other temperatures. Also, overall, the phenotype due to *ASU51* was not as pronounced, as the phenotype due to *ASU53* (Figure S3).

In conclusion, we have found that each studied locus, *CSU51* or *CSU53*, contains a *CSU* element and an *ASU* element, which is embedded in *CSU* in the opposite orientation. In the natural *CSU/ASU* configurations, the Sou^{-} repressive phenotype of *CSU*s dominates the weaker Sou^{++} phenotype of the *ASU*s (cartoons 3 or 4 in Figure 2 and Figures 3A, S3, and 4). The *ASU* element can be revealed by interfering with the transcription, but not the translation of the *CSU* element. A simple explanation would be that mutations destroying the Csu proteins leave the corresponding sense RNAs nearly intact. However, the elimination of the upstream regions, hence promoters, interferes with the production of the sense *CSU* transcripts which subsequently effects translation products. Such phenotypic differences are expected if sense *CSU* and antisense *ASU* transcripts interact, for example, by forming dsRNA molecules that subsequently either inhibit translation or lead to degradation. Then, the lack of the *CSU* transcript from the *CSU/ASU* configuration from the plasmid would lead to the increased abundance of the *ASU* transcript from the plasmid and ultimately would lead to the increase of combined amount of plasmid and chromosomal *ASU* transcripts. This would increase the interaction with the chromosomal *CSU* transcripts and cause more depletion of *CSU* inhibitory transcripts, thus, up regulating *SOU1*, and, finally, resulting in the better growth.

Visualization of sense *CSU* and antisense *ASU* transcripts with Northern blots: We determined *CSU* and *ASU* transcripts from chromosomal *CSU*/*ASU* configurations in the strains SC5314, 3153A, and CAF4-2 with Northern blots from three or four independent cultures of each strain (MATERIALS AND METHODS), as exemplified with SC5314 and 3153A in Figure 5A. Highly abundant *CSU51* sense transcript was revealed with total RNA and dsDNA probe prepared from *CSU51*. However, the low abundance complementary *ASU51* antisense transcript

could be only revealed with mRNA and a riboprobe prepared from *ASU51* (Figure 5B). The low abundance sense *CSU53* and antisense *ASU53* were also revealed with mRNA and riboprobes prepared from the corresponding genes (Figure 5B). Each probe repeatedly produced hybridization signal(s) of the pattern, which was identical in all three examined strains. Specifically, the *CSU51* probe revealed more abundant transcript of approximately 700 nt and less abundant transcript of approximately 1,200 nt. The *ASU51* probe revealed a pool of transcripts that were distributed around a 1,000 nt size. A *CSU53*/*ASU53* pair produced one sense and one antisense transcript of approximately 1,300 nt and 1,200 nt, respectively.

Relative amounts of antisense *ASU* and sense *CSU* transcripts: We estimated comparative amounts of sense *CSU* and antisense *ASU* transcripts from chromosomal *CSU/ASU* configurations in the strains SC5314, 3153A, and CAF4-2 using s.-q. RT-PCR analysis with gene-specific primers (MATERIALS AND METHODS). The transcript levels of *ASU51*, as compared to *CSU51*, were 1.5% in SC5314, 4% in CAF4-2, and 3.8% and 4% in 3153A in two independent experiments, thus, ranging from 1.5% to 4%. Also, the transcript levels of *ASU5*3, as compared to *CSU53*, were 16% in SC5314, 8% in CAF4-2, and 37% and 44% in 3153A in two independent experiments, thus, ranging from 8% to 44%.

The RT-PCR data are in agreement with the Northern blot analyses that indicated the lower abundance of the antisense transcript for at least *CSU51*/*ASU51* configuration (see above and Figure 5A). Furthermore, the low amounts of the *ASU* transcripts are consistent with their phenotype being recessive in respect to the phenotype of *CSU*s (see above and Figure 3A). It should be pointed out that sense and antisense transcripts were analyzed using the same RNA preparations, implying that sense and antisense transcripts may exist in the same cells.

Mapping UTRs of sense and antisense transcripts: In order to clarify how sense and antisense transcripts overlap, we mapped UTRs of each *CSU* and *ASU* transcript using the RACE technique with a total of 3 to 6 clones from each of two different strains, as described in MATERIALS AND METHODS. This approach, although designed to recognize 5' cap or 3' polyadenylated [poly(A)] structures, fails, however, to determine the sizes of poly(A) tails.

We found that *CSU/ASU* loci in both of the examined strains 3153A and CAF4-2 produced multiple sense or antisense 5'- and 3'-UTRs. The sizes of the largest *CSU51*, *CSU53*, and *ASU53* transcripts, 945 nt, 1318 nt, and 1003 nt, respectively, that were calculated from the largest 5'- and 3'-UTRs, as well as the corresponding ORF, but not the poly(A) (see above), never exceeded, but approximately corresponded, to the transcript sizes on the Northern blots in Figure 5A. A large difference occurred with the *ASU51* transcript: 395 nt by RACE versus approximately 1,000 nt by Northern blot. This difference could be, for example, due to a large poly(A) tail. In fact, large poly(A) tails of up to 365 nt, 400 nt or 650 nt have been reported in eukaryotes (CARRAZANA *et al.* 1988; SALLES and STRICKLAND 1995).

Mutational analysis demonstrates that *ASU* ORFs are not critical for the Sou^{++} phenotype: We addressed the question whether the *ASU* ORF is necessary for the Sou phenotype by introducing mutations in the ORFs of *ASUs* (cartoon 7, Figure 2). We used a 1.2 kb portion pEA104 from region 135 that carries *ASU53* and that confers the Sou^{++} phenotype (Figures 3A and C and 4A) in order to create two mutations: a frame-shift mutation at positions +4 and +7 of the *ASU53* ORF, pEA232, and a stop codon at position +12, pEA205, as indicated by "stars" in Figure 4A. Also, we used a 400 bp portion pEA209 from region A that carries *ASU51* (Figure

4B) that confers the Sou^{++} phenotype (Figure S3B) in order to create a frame-shift mutation at position +39 of the *ASU51* ORF, pEA243, as indicated by a "star" in Figure 4B. (See Table S1 and supporting information for the description of mutations and plasmids). We found no difference in the Sou^{++} growth of the cells carrying plasmids with intact or mutated *ASU*s, as exemplified in Figure 3D with the *ASU53* mutations carried on pEA232 and pEA205. See Figure S2B for the control growth on glucose medium.

***ASU*s interact with *CSU*s at the phenotypic and the transcriptional levels:** We directly addressed the question of interaction between the corresponding *CSU* and *ASU* elements at the phenotypic level. *SOU1*, a natural *CSU/ASU* configuration (Sou^{-}), and an extra *ASU* (Sou^{++}); were co-over expressed on a replicative plasmid (cartoon 8 and 9, Figure 2), the *CSU/ASU* ratio, thus, resulting in 1:2. The plasmid pEA238 included *CSU51*/*ASU51* and *ASU51* carried, respectively, on region A and the portion pEA209 from region A (Figure 4B). The plasmid pEA162 included *CSU53*/*ASU53* and *ASU53* carried, respectively, on the portions pEA105 and pEA104 of region 135 (Figure 4A). (See Table S1 and supporting information for plasmid preparations. See Figures 3, S1, and S3 for the spot assays on plates of the individual elements).

The plasmids with extra *ASU*s, pEA238 and pEA162, were tested on sorbose medium in multiple spot assays, as described above. We found that these plasmids substantially diminished the dominating Sou^{-} repressive phenotype of corresponding *CSU* elements. As exemplified in Figure 3E with the plasmid pEA162, an extra *ASU53* resulted in the growth, which was almost indistinguishable from the control growth, Sou^{+}. Consistently, an extra *ASU51* resulted in either intermediate growth between no growth and control growth, *i. e*., neither Sou^{-} nor Sou^{+} occurred, or high frequency large Sou^{+} colonies (Figure S4A). We would like to emphasize that

the Sou$^+$ colonies never occurred, when region A, *i.e.*, *CSU51/ASU51*, was co-over expressed alone with *SOU1*. This variability, presumably, occurred due to the highly expressed and, thus, strong *CSU51* (see above). Also, see Figures S4A and B for the control growths on glucose medium.

In order to substantiate the finding of the interaction between *ASU*s and *CSU*s, we integrated one or two copies of *ASU53*, vectors pEA249 or pEA254, respectively, in the *LEU2* locus on Ch7 in the strain CAF4-2. The Control integration was with no *ASU53*, pAK156 (Table S1). We have previously demonstrated that integration in *LEU2* did not interfere with the Sou phenotype (Wang *et al*., 2004). The proper integration of pAK156, pEA249, and pEA254 was verified by PCR amplifications with a pair of primers, one for the *LEU2* gene, AF237U, and another for the lacZ N-terminal sequence of integration vector, M-13FU (Table S2). All three amplicons had the expected sizes of approximately 2.0 kb, 3.8 kb, and 6.2 kb, respectively, while no amplicon was produced with genomic DNA from the control strain SC5314. The presence of *ASU*53 in corresponding amplicons was confirmed with primers AF110 and AF111.

We prepared three batches of total RNA from three independent cultures of each construct. We then synthesized cDNA and used it for PCR amplifications with gene specific primers (Table S2). Reactions were set in duplex with control genes and were conducted with different number of cycles followed by quantitation (MATERIAL AND METHODS). See Figure S5 for the examples of amplicons that were quantitated. As expected, the amount of the *ASU53* transcript increased consistent with the increase of the copy number of *ASU53*, as compared to the control strain with two regular chromosome copies carrying the integration vector with no extra copies of *ASU53* (Table 1). Introduction of one extra copy of *ASU53* increased the *ASU53* transcript to 1.4; 1.5; and 1.6 resulting in an average of 1.5 +/- 0.1; introduction of two extra copies of

ASU53 increased the *ASU53* transcript to 1.7; 1.9, and 1.9 resulting in an average of 1.9 +/- 0.1. We then asked if the amount of the *CSU53* transcript changed. We found a copy-dependent inhibition of *CSU53* by *ASU53*. In the presence of one extra copy, the *CSU53* transcript was 0.6; 0.7; 0.8 resulting in an average of 0.7 +/- 0.1; whereas in the presence of two extra copies, the *CSU53* transcript was 0.5; 0.5; and 0.6 resulting in an average of 0.5 +/- 0.1 (Table 1).

The effect of chromosomal extra copies of *ASU51* on the expression of the corresponding *CSU51* was not analyzed, because of a large difference in the expression of the two genes.

***SOU1* is up regulated by *ASU53* and down regulated by *CSU51* or *CSU53*:** We next determined the effect of genomic extra *ASU*s on the metabolic *SOU1* expression. RT-PCR amplifications were conducted and analyzed as above. In the presence of one or two extra copies of *ASU53*, the amount of the *SOU1* transcript increased in a copy-dependent fashion (Table 1). In the presence of one extra copy, the *SOU1* transcript was 1.6; 1.8; and 1.9 resulting in an average of 1.8 +/- 0.2; whereas in the presence of two extra copies, the *SOU1* transcript was 2.6; 2.6; and 2.7 resulting in an average of 2.6 +/- 0.2.

We next used a co-over expression system to address whether *CSU51* or *CSU53* controls *SOU1* expression. Northern blot analyses were carried out with independently prepared batches of total RNA that was extracted from CAF4-2 cells transformed with the replicative plasmids pEA105 or pCA135 (Figure 4A) co-over expressing *SOU1* with *CSU53*, as well as with the plasmid pAK65 (Figure 4B), co-over expressing *SOU1* with *CSU51*. Also see cartoons 3 and 4 of Figure 2. Note that *CSU*s are represented by the natural configurations *CSU*/*ASU*, because *CSU*s cannot be separated from *ASU*s. However, in these natural configurations, *CSU*s dominate *ASU*s (see above). As exemplified in Figure 6, the co-over expression with each *CSU* gene

clearly diminished the amount of the *SOU1* transcript, as compared to the control pCA88, which over expressed *SOU1*, but lacked *CSU*. Specifically, *SOU1* was down regulated 0.55; 0.63; 0.33; and 0.36 fold, resulting in an average of 0.47 +/- 0.15 by *CSU53* in four independent experiments; as well as 0.18 and 0.29 fold by *CSU51* in two independent experiments. As expected from the phenotypes and transcript abundances (see above), *CSU51* repressed *SOU1* stronger than *CSU53*.

DISCUSSION

We found that the previously identified *CSU51* and currently identified *CSU53* each contain a genetic element, *ASU*, which is embedded in *CSU* in the opposite orientation. *ASU*s are manifested in several ways. They produce antisense transcripts that are, however, significantly less abundant than the corresponding *CSU* sense transcripts. Extra copies of *ASU* in the genome, lead to an decrease or increase of, respectively, the transcript of the regulatory *CSU* or the metabolic *SOU1*. *ASU*s are also manifested at the phenotypic level. A copy of *ASU* over expressed from the low copy number plasmid, possesses a distinct, albeit weak, phenotype, enhancing the growth on sorbose. This is, presumably, due to a combined action of the plasmid and chromosomal copies of *ASU* on the chromosomal copies of *CSU*. Consistently, a copy of *ASU* co-over expressed from plasmid with the corresponding natural *CSU*/*ASU* configuration greatly diminishes the *CSU* repressive phenotype.

The interaction between *CSU*s and *ASU*s is, thus, clearly, revealed at the transcription and at the phenotypic levels. It is also clear that the levels of the *CSU* and *ASU* transcripts are in a reverse relationship. Conversely, the expression of metabolic *SOU1* on Ch4 directly depends on the transcription of *ASU*s and reversely depends on the transcription of negative regulatory

*CSU*s. It is reasonable to suggest that *ASU* elements act as modulators repressing *CSU*s, because the phenotype of *CSU*s dominates and because the *ASU* transcripts are less abundant and are completely imbedded in the sense transcripts. Despite not specifically addressed, both *CSU* sense and *ASU* antisense transcripts are expected to be produced in the same cell, as this is required for their interaction. The complementarity of sense/antisense elements is also a strong indication of their interaction. As discussed in the Results, we obtained preliminary evidence that, as expected for the complementary sequences, sense and antisense RNA interact. This is because *ASU* phenotype is manifested, when the sequence upstream to *CSU* ORF is abrogated, thus, greatly interfering with the *CSU* transcription, but not when the Csu protein is mutated leaving the *CSU* transcription normal.

All sense *CSU* and antisense *ASU* transcripts contain ORFs and are also capped and polyadenylated, which is indicative of translation. However, while both *CSU* ORFs are essential for the Sou^{-} repressive function, *ASU* ORFs are not essential for the Sou^{++} enhancing function, as clearly demonstrated by mutational analysis. This could be interpreted, as Asu proteins are not produced or, alternatively, are not relevant for the Sou phenotype. In this respect, non-coding RNA can be capped and polyadenylated; furthermore some RNAs can function both as mRNA and as non-coding RNA (RAPICAVOLI and BLACKSHAW 2009; DINGER *et al.* 2008; CALLAHAN and BUTLER 2008). Future studies should establish whether proteins are translated from *ASU*s and, if so, how they function. Currently, we consider that *ASU* transcripts are implicated with the Sou phenotype as non-coding RNAs.

Our current model proposes that interactions between three kinds of elements *CSU*, *ASU*, and *SOU1* are based on the transcript ratios. *ASU*s repress *CSU*s that, in turn, repress *SOU1*. Because the amount of *CSU* transcript diminishes, when corresponding *ASU* is over expressed, a

plausible scenario is that the complementary *CSU* and *ASU* RNAs form ds RNA molecules that are subsequently degraded.

In summary, identification of an additional *CSU53* confirms the previous evidence of multiple *CSU*s on Ch5. The inverse relationship between *CSU51* or *CSU53* transcripts on one hand and *SOU1* transcript on the other hand is consistent with the previous proposal that *CSU*s are negative regulators of *SOU1* (KABIR *et al.* 2005) that determine the regulatory role of Ch5 copy number. Discovery of *ASU* positive elements that counter-act *CSU* negative elements reveals an unanticipated layer of complexity in the negative regulation of growth on sorbose.

ACKNOWLEDGMENTS

We thank Fred SHERMAN, Scott Butler, and Yi-Tao Yu for the inspiring discussions, as well as for the critical reading of the manuscript. We thank M. Anaul Kabir for the plasmid pAK156. This work was supported in part by National Institutes of Health Grant GM12702. We are also grateful to The University of Rochester Funds that enabled this study.

LITERATURE CITED

Ahmad, A., M. A. Kabir, A. Kravets, E. Andaluz, G. Larriba *et al*., 2008 Chromosome instability and unusual features of some widely used strains of *Candida albicans*. Yeast **25**: 433-448.

Bruno, V. M., Zh. Wang, S. L. Marjani, Gh. M. Euskirchen, J. Martin *et al*., 2010 Comprehensive annotation of the transcriptome of the human gungal pathogen *Candida albicans* using RNA-seq. Genome Research **20**: 1451-1458.

Cabezón, V., A. Llama-Palacios, C. Nombela, L. Monteoliva and C. Gil, 2009 Analysis of *Candida albicans* plasma membrane proteome. Proteomics **9**: 4770-4786.

Callahan, K. P., and J. S. Butler, 2008 Lifting the veil on the transcriptome. Genome Biol**. 9**: 218.1-218.4.

Camblong, J., N. Iglesias, C. Fickentscher, G. Dieppois and F. Stutz F, 2007 Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *Saccharomyces cerevisiae*. Cell **131**: 706-717.

Carranzana, E. J., K. B. Pasieka and J. A. Majzoub, 1988 The vasopressin mRNA poly (A) tract is unusually long and increases during stimulation of vasopressin gene expression *in vivo*. Mol. Cell. Biol. **8**: 2267-2274.

Cerazin-Leroy, V., D. Denis-Henriot, M. Morot, P. deMazancourt and Y. Giudicelli, 1998 Semi quantitative RT-PCR for comparison of mRNAs in cells with different amounts of housekeeping gene transcripts. Mol. Cell. Probes **12**: 283-291.

Clements, J. M., T. M. Laz and F. Sherman, 1988 Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8**: 4533-4536

David, L., W. Huber, M. Granovskaia, J. Toedling, C. J. Palm *et al*., 2006 A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. USA **103**: 5320-5325.

Dinger, M. E., K. C. Pang, T. R. Mercer and J. S. Mattick, 2008 Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput. Biol. **4**: e1000176.

Drinnenberg, I. A., D. E. Weinberg, K. T. Xie, J. P. Mower, K. H. Wolfe *et al*., 2009 RNAi in budding yeast. Science **326**: 533-534.

Ding, Y., L. Xu, B. D. Jovanovic, I. B. Helenovski, D. L. Kelly *et al*., 2007 The methodology used to measure different gene expression affects the outcome. J. Biomol. Tech. **18**: 321-330.

Dutrow, N., D. A. Nix, D. Holt, B. Milash, B. Dalley *et al*., 2008 Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping. Nat. Genet. **40**: 977-986.

Gaur, N. A., N. Puri, N. Karnani, G. Mukhopadhyay, S. K. Goswami *et al*., 2004 Identification of a negative regulatory element which regulates basal transcription of a multidrug resistance gene *CDR1* of *Candida albicans*. FEMS Yeast Res. **4**: 389-399.

Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin *et al*., 2009 Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458**: 223-227.

Hongay, C. F., P. L. Grisafi, T. Galitski and G. R. Fink, 2006 Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. Cell **127**: 735-745.

Houseley, J., L. Rubbi, M. Grunstein, D. Tollervey and M. Vogelauer, 2008 A ncRNA modulates histone modification and mRNA induction in the yeast *GAL* gene cluster. Mol. Cell **32**: 685-695.

Janbon, G., F. Sherman and E. Rustchenko, 1998 Monosomy of a specific chromosome determines L-sorbose utilization: a novel regulatory mechanism in *Candida albicans*. Proc. Natl. Acad. Sci. USA **95**: 5150-5155.

Kabir, M. A., A. Ahmad, J. R. Greenberg, Y. K. Wang and E. Rustchenko, 2005 Loss and gain of chromosome 5 controls growth of *Candida albicans* on sorbose due to dispersed redundant negative regulators. Proc. Natl. Acad. Sci. USA **102**: 12147-12152.

Kabir, M. A., and E. Rustchenko, 2005 Determination of gaps by contig alignment with telomere-mediated chromosomal fragmentation in *Candida albicans*. Gene **345**: 279-287.

Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj *et al*., 2009 Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA **106**: 11667-11672.

Kuai, L., F. Fang, J. S. Butler and F. Sherman, 2004 Polyadenylation of rRNA in *Saccharomyces cerevisiae***.** Proc Natl Acad Sci USA **101**: 8581-8586.

Lee, T. L., A. L. Pang, O. M. Rennert and W. Y. Chan, 2009 Genomic landscape of developing male germ cells. Birth Defects Res. (Part C) **87**: 43–63.

Mahmoudi, S., S. Henriksson, M. Corcoran, C. Méndez-Vidal, K. G. Wiman *et al*., 2009 Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol. Cell **33**: 462-471.

Morris, K. V., 2009 Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. Epigenetics **4**: 296-301.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou C, D. Raha *et al*., 2008 The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320**: 1344-1349.

Nagano, T., and P. Fraser, 2011 No-nonsense function for long noncoding RNAs. Cell **145**: 178-181.

Olejniczak, M., P. Galka and W. J. Krzyzosiak, 2010 Sequence-non-specific effects of RNA interference triggers and microRNA regulators. Nucleic Acids Res. **38**: 1-16.

Perepnikhatka, V., F. J. Fischer, M. Niimi, R. A. Baker, R. D. Cannon *et al*., 1999 Specific chromosome alterations in fluconazole-resistant mutants of *Candida albicans*. J. Bacteriol. **181**: 4041-4049.

Rapicavoli, N. A., and S. Blackshaw, 2009 New meaning in the message: noncoding RNAs and their role in retinal development. Dev. Dyn. **238**: 2103-2114.

Russo, P., W. Z. Li, D. M. Hampsey, K. S. Zaret and F. Sherman, 1991 Distinct *cis*-acting signals enhance 3′-endpoint formation of *CYC1* mRNA in the yeast *Saccharomyces cerevisiae*. EMBO J. **10**: 563-571.

Rustchenko, E., 2007 Chromosome instability in *Candida albicans*. FEMS Yeast Res. **7**: 2-11.

Rustchenko, E., 2008 Specific chromosome alterations of *Candida albicans*: mechanism for adaptation to pathogenicity, pp 197-212 in *Evolutionary biology of bacterial and fungal pathogens*, edited by C. Nombela, G. H. Cassel, F. Baquero and J. A. Gutiérrez-Fuentes. ASM Press, Washington.

Rustchenko-Bulgac, E.P., 1991 Variations of *Candida albicans* electrophoretic karyotypes. J. Bacteriol. **173**: 6586-6596.

Rustchenko-Bulgac, E. P., and D. H. Howard, 1993 Multiple chromosomal and phenotypic changes in spontaneous mutants of *Candida albicans*. J. Gen. Microbiol. **139**: 1195–1207.

Rustchenko, E., D. H. Howard and F. Sherman, 1994 Chromosomal alterations of *Candida albicans* are associated with the gain and loss of assimilating functions. J. Bacteriol. **176**: 3231-3241.

Rustchenko, E., and F. Sherman, 2002 Genetic instability of *Candida albicans*, pp 723-776 in *Fungi Pathogenic for Humans and Animals*, edited by D. H. Howard. Marcel Dekker, Inc., New York.

Sallés, F.J., and S. Strickland, 1995 Rapid and sensitive analysis of mRNA polyadenylation states by PCR. Genome Res. **4**: 317-321.

Samanta, M. P., W. Tongprasit, H. Sethi, C. S. Chin and V. Stolc, 2006 Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway. Proc Natl Acad Sci USA **103**: 4192-4197.

Sellam, A., H. Hogues, Ch. Askew, F. Tebbji, M. van het Hoog *et al*., 2010 Experimental annotation of the human pathogen *Candida albicans* coding and noncoding transcribed regions using high resolution tiling arrays. Genome Biol. **11**: R71 doi:10.1186/gb-2010-11-7-r71.

Sherman, F., 2002 Getting started with yeast. Methods Enzymol. **350**: 3-41.

Srikantha, T. H., A. R. Borneman, K. J. Daniels, C. Pujol, W. Wu *et al*., 2006 *TOS9* regulates white-opaque switching in *Candida albicans*. Eukaryot. Cell **5**: 1674-1687.

Taft, R., K. C. Pang, T. R. Mercer, M. Dinger and J. S. Mattick, 2010 Non-coding RNAs: regulators of disease. J. Pathol. **220**: 126-139.

Tuch, B. B., Q. M. Mitrovich, O. R. Homann, A. D. Hernday, C. K. Monighetti *et al*., 2010 The transcriptomes of two heritable cell types illuminate the circuit governing their differentiation. PLoS Genetics **6**: e1001070.

Vinces, M. D., C. H. Haas and C. A. Kumamoto, 2006 Expression of the *Candida albicans* morphogenesis regulator gene *CZF1* and its regulation by Efg1p and Czf1p. Eukaryot. Cell **5**: 825-835.

Wang, Y-K., B. Das, D. H. Huber, M. Wellington, M. A. Kabir *et al*., 2004 Role of the 14–3–3 protein in carbon metabolism of the pathogenic yeast *Candida albicans*. Yeast **21**: 685–702.

Wellington, M., and E. Rustchenko, 2005 5-Fluoro-orotic acid induces chromosome alterations in *Candida albicans*. Yeast **22**: 57–70.

TABLE 1

Expression changes of *CSU53* and *SOU1* in the presence of genomic extra copies of *ASU53*

	Expression change*			
Gene	1 extra *ASU53*	Mean +/- SD	2 extra *ASU53*s	Mean +/- SD
ASU53	1.4; 1.5; 1.6	1.5 +/- 0.1	1.7; 1.9; 1.9	1.9 +/- 0.1
CSU53	0.6; 0.7; 0.8	0.7 +/- 0.1	0.5; 0.5; 0.6	0.5 +/- 0.2
SOU1	1.6; 1.8; 1.9	1.8 +/- 0.2	2.6; 2.6; 2.7	2.6 +/- 0.2

* Expression change for each gene was determined by the ratio of the amount of transcript in the integration construct having one or two extra copies of *ASU53*s versus the control integration construct having no extra *ASU53*.

FIGURE LEGENDS

FIGURE 1.- Schematic presentation of two *CSU*/*ASU* sense/antisense configurations in the context of Ch5. *CSU51*/*ASU51* configuration from region A or *CSU53*/*ASU53* configuration from region 135 is located within a 209 kb portion of Ch5, which is critical for growth on sorbose. This portion also carries other regions B, C, and 139, as indicated (Kabir *et al*. 2005). The size of Ch5 is indicated and the centromere (C) and telomeres (T) of Ch5 are shown. Also indicated are the ORF sizes of *CSU*s and *ASU*s.

FIGURE 2.- Cartoons representing various co-over expression plasmids that were derived from the backbone plasmid pCA88. Shown are the following: 1, an original low copy number replicative plasmid pRC2312; 2, its derivative pCA88 over expressing the metabolic *SOU1* gene; 3 to 9, various plasmids co-over expressing *SOU1* with a sequence of interest. The sequences are indicated by an open, gray, and black blocks. These correspond to the Ch5 region, to the portion of region encompassing the *CSU* gene, which always contains an embedded *ASU* gene (*CSU* + *ASU*); and to the *ASU* gene, as indicated. Mutations are marked with "X". The phenotype of the recipient Sou^{-} cells of the strain CAF4-2 conferred by each type of plasmid on L-sorbose medium, is indicated by Sou^{-}, Sou^{+}, Sou^{++} , and Sou $^{+/-}$. The Sou^{-} or Sou^{+} phenotypes due to, respectively, the original pRC2312 or pCA88 (*SOU1*) are considered, as the general negative or positive controls for growth. Note that in the *CSU/ASU* configuration, the stronger repression phenotype of *CSU*, Sou^{-}, dominates the weaker phenotype of *ASU*, Sou^{++}, in the *CSU*/*ASU* configuration. Also note that the intact *CSU* always contains the embedded *ASU*. However *ASU* can be removed from the larger region without *CSU*.

FIGURE 3.- Phenotypes conferred to a recipient Sou^{-} strain CAF4-2 by various replicative plasmids. Two major elements *CSU53/ ASU53* or *ASU53* that derived from region 135 of Ch5 (Figures 1 and 4A), are co-over expressed with the *SOU1* gene in different combinations, as indicated in parentheses. For more information see the legend of Figure 2. Note, that the names of the plasmids and tested portions that are presented schematically in Figure 4A are the same. Shown are examples of spot assay for

growth on medium containing sorbose as a sole carbon source. Approximately 5 x 10^5 cells per spot were plated in duplicates. A) Analysis of different portions of region 135 (Figure 4A) resulting in the identification of *CSU53* and the *CSU53*/*ASU53* configuration. The Sou phenotypes conferred by plasmids are indicated, as follows: the control Sou$^-$ lack of growth in the presence of an empty vector pRC2312; the control Sou$^+$ growth due to pCA88 (see also the legend of Figure 2); the Sou$^-$ repression of growth due to pCA135 or pEA144; and the Sou^{-*} repression of growth with multiple Sou$^+$ colonies due to pEA105 or pEA143. When the *CSU53*/*ASU53* configuration is carried on a shorter portion than an entire region 135, it is designated *CSU53***. Note that repression of growth occurs due to the natural *CSU53*/*ASU53* configuration in which *CSU53* always dominates *ASU53*. Also shown is the Sou^{++} enhanced growth due to pEA104 co-over expressing *SOU1* with *ASU53*. Also shown is the control Sou$^+$ growth due to pEA145 and pEA140 that co-over express *SOU1* with the representative portions of region 135 lacking *CSU* and/or *ASU* elements. B) Mutational analysis of the *CSU53* ORF implicates the putative Csu53 protein with the Sou phenotype. Plasmids pEA227, pEA201, and pEA158, each containing a region 135 with a different frame-shift mutation of *CSU53* ORF (cartoon 5 in Figure 2, Figure 4A) abolish the Sou$^-$ repression phenotype of region 135 in favor of the control growth due to *SOU1*, Sou$^- \rightarrow$ Sou$^+$. C) Plasmids pEA104 or pEA234 co-over expressing *ASU53* and *SOU1*, enhance growth on sorbose medium, Sou^{++}, independent of the insert orientation on a plasmid, as indicated by an arrow. See above for more explanations. D) Mutational analysis of the ASU53 ORF does not implicate the putative Asu53 protein with the Sou phenotype. Plasmids pEA232 or pEA205 carrying, respectively, frame-shift or stop codon mutation in the *ASU53* ORF (cartoon 7 in Figure 2, Figure 4A), do not abolish the original Sou^{++} phenotype. See above for more explanations. E) *ASU*53 and *CSU*53 interact at the phenotypic level, as shown with the plasmid pEA162 (cartoon 8 in Figure 2) co-over expressing three elements: *SOU1*; *CSU53* in a natural configuration *CSU53*/*ASU53*, which is designated *CSU53***; and *ASU53*. An extra *ASU53* changes the repressive phenotype of *CSU53* to the growth, which is almost equal to the control Sou$^+$ growth. See above for more explanations.

FIGURE 4.- Schematic presentation of the analyses of *CSU*/*ASU* configuration in region 135 or A of Ch5 (Figure 1). The entire region or a portion of the region was individually co-over expressed with *SOU1* on a plasmid (Figure 2) and assayed on sorbose medium, as exemplified in Figure 3. The co-over expression phenotypes of portions fell into three categories: repression of growth Sou^{-}, enhanced growth Sou^{++}, and control growth Sou^{+}, as indicated on the right (see the legend of Figure 3A for more explanations). A) *CSU53*/*ASU53* configuration in the context of region 135, which is also designated pCA135 by the name of the corresponding plasmid. Also shown are the sense *CSU53* and antisense *ASU53* transcripts, as determined by RACE in the strains 3153A and CAF4-2 (see the section "Mapping UTRs of sense and antisense transcripts"). Also shown are representative portions of region 135 that were co-over expressed with *SOU1* in order to identify *CSU53* and *ASU53*. Mutations within an entire region 135 in *CSU53* ORF or within a portion of the region in *ASU53* ORF are indicated with "stars". B) *CSU51*/*ASU51* configuration in the context of region A, which is also designated pAK65 by the name of the corresponding plasmid. Also shown are the sense *CSU51* and antisense *ASU51* transcripts that were determined as indicated in A. Also shown are representative portions of region A that were co-over expressed with *SOU1* in order to identify *ASU51*.

FIGURE 5.- Northern blot analyses reveal the *CSU53/ ASU53* and *CSU51/ ASU51* sense/antisense transcripts, as exemplified with the strains 3153A and SC5314. A) Hybridization signals obtained with total RNA and dsDNA probe for *CSU51* transcript, as well as mRNA and riboprobes for *CSU53*, *ASU51* and *ASU53* transcripts, as indicated. For quantitative estimate of amounts of the corresponding *CSU* and *ASU* transcripts see the section "Relative amounts of antisense *ASU* and sense *CSU* transcripts". B) Nucleotide sequence of *CSU53/ASU53* or *CSU51/ASU51* ORF configuration. Start codons of *CSU* sense elements, as well as sequences comprising *ASU* antisense elements are in bold. The sequences that were used to prepare riboprobes hybridizing with mRNA transcribed from either Watson or Crick strand on the Northern blot in A are underlined.

FIGURE 6.- Amount of the *SOU1* transcript is inversely related to the amount of *CSU* transcript. Co-over expression of *SOU1* with *CSU51* from a low copy number plasmid (cartoon 3 in Figure 2) decreases the

SOU1 transcript, while increasing, as expected, the *CSU51* transcript, as determined with Northern blot analysis. The recipient CAF4-2 cells carry the following plasmids: 1, negative control pRC2312, an empty vector; 2, positive control pCA88 over expressing *SOU1*; 3, pAK65 co-over expressing *SOU1* with region A carrying a natural *CSU51/ASU51* configuration in which *CSU51* always dominates *ASU51*. Hybridization signals were obtained with total RNAs and dsDNA probes prepared from *SOU1* or *CSU51*, as indicated. Normalization for the amounts of loaded RNA was done with the same blot using dsDNA probe prepared from 18S rDNA. Note the decrease of the *SOU1* signal from pAK65, as compared to pCA88.

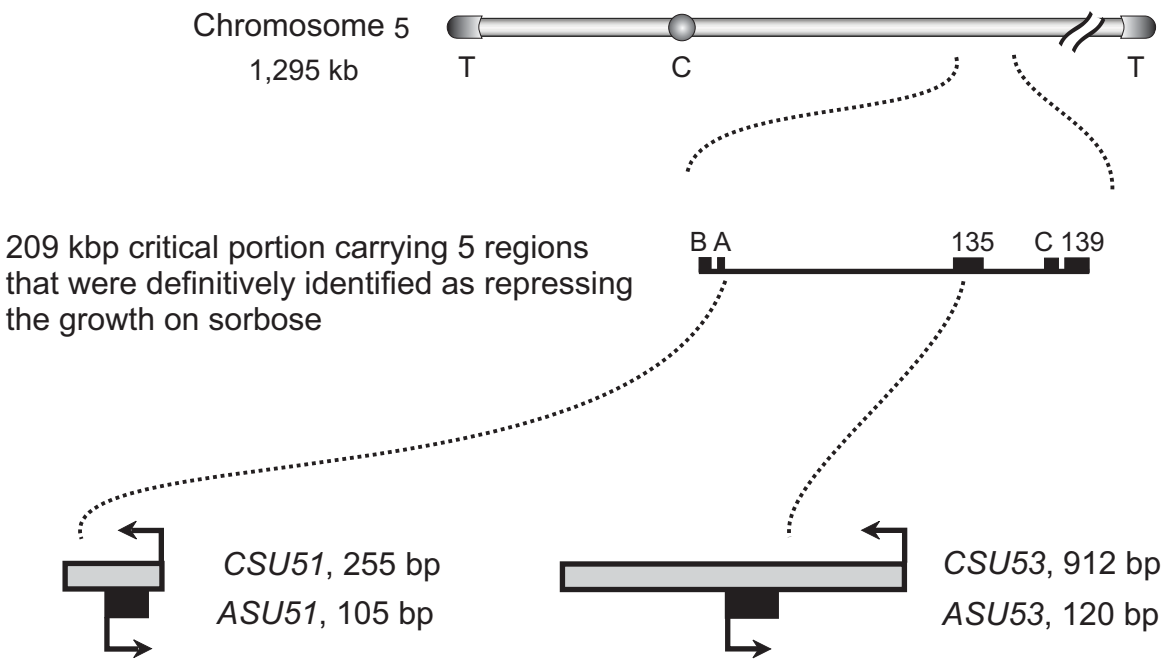
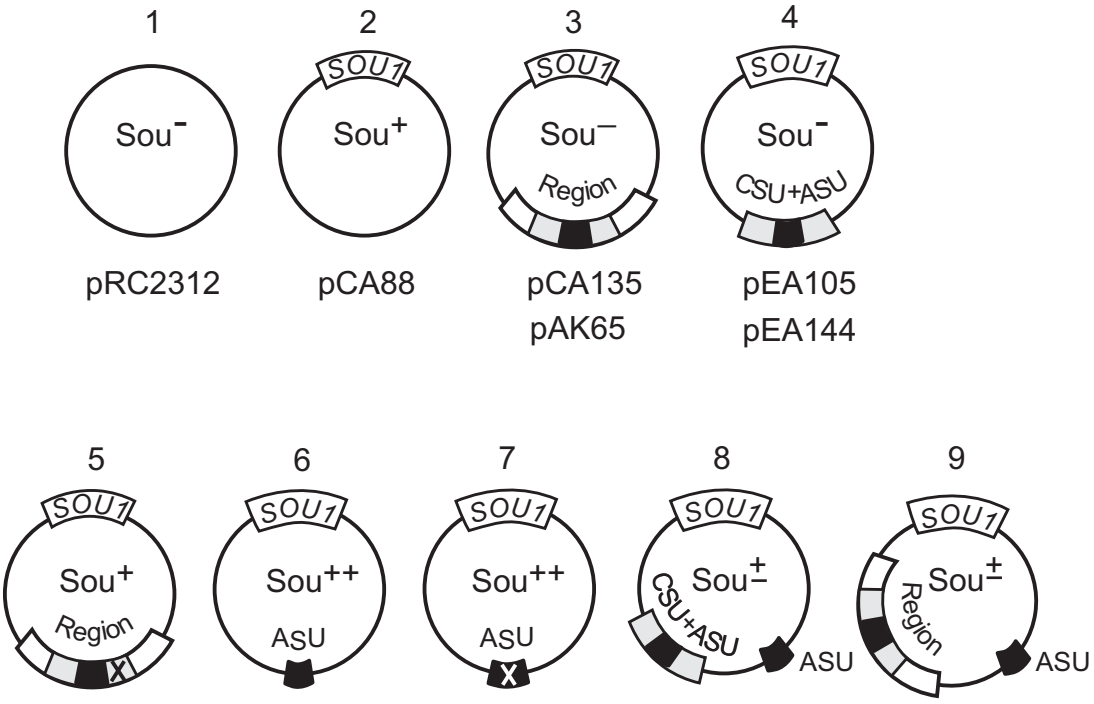

Chromosome 5
1,295 kb
T
C
T
209 kbp critical portion carrying 5 regions that were definitively identified as repressing the growth on sorbose
B A
135
C 139
CSU51, 255 bp
ASU51, 105 bp
CSU53, 912 bp
ASU53, 120 bp

Fig. 1



1	2	3	4	5	6	7	8	9
pRC2312	pCA88	pCA135 pAK65	pEA105 pEA144	pEA158 pEA201 pEA227	pEA104 pEA155 pEA156 pEA219 pEA221 pEA209 pEA234 pEA236	pEA205 pEA232 pEA243	pEA162	pEA238

Fig. 2

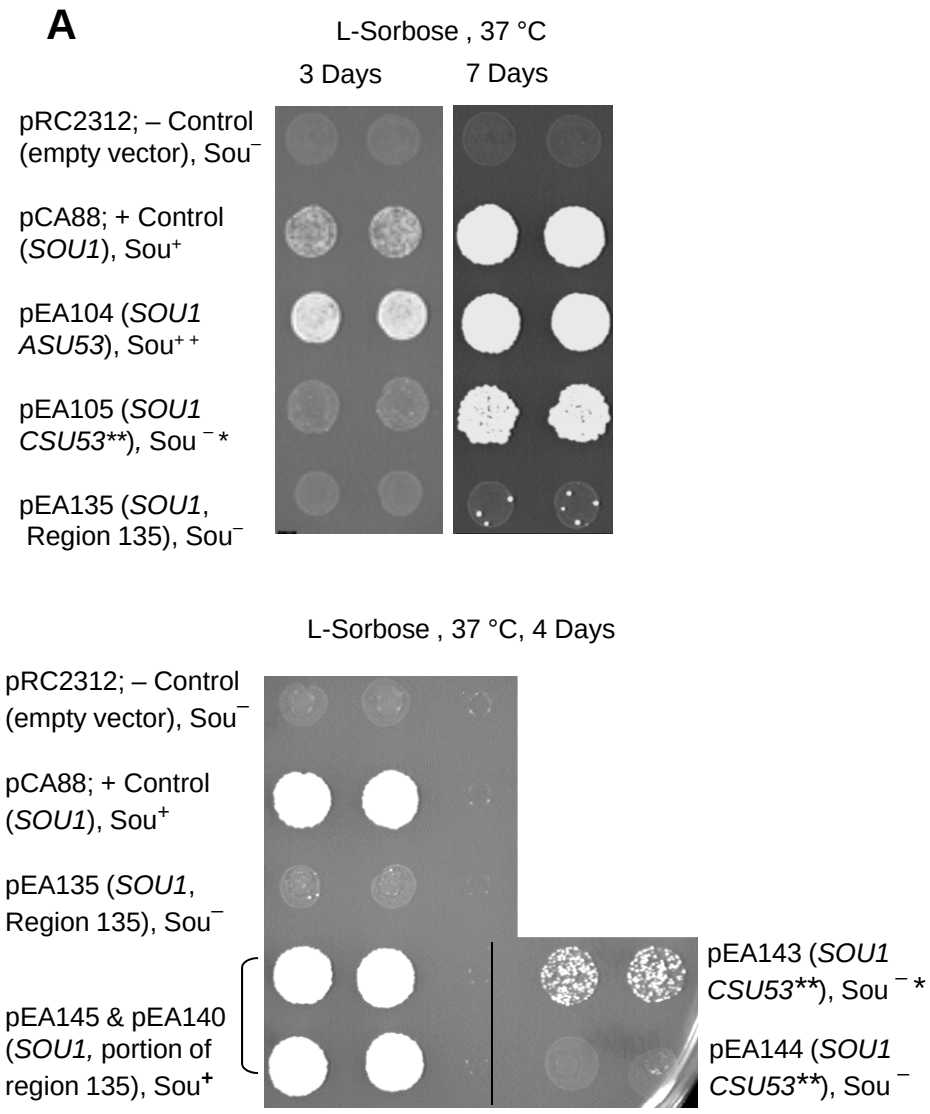

A
L-Sorbose , 37 °C
3 Days
7 Days
pRC2312; – Control (empty vector), Sou⁻
pCA88; + Control (SOU1), Sou⁺
pEA104 (SOU1 ASU53), Sou⁺⁺
pEA105 (SOU1 CSU53**), Sou⁻*
pEA135 (SOU1, Region 135), Sou⁻
L-Sorbose , 37 °C, 4 Days
pRC2312; – Control (empty vector), Sou⁻
pCA88; + Control (SOU1), Sou⁺
pEA135 (SOU1, Region 135), Sou⁻
pEA145 & pEA140 (SOU1, portion of region 135), Sou⁺
pEA143 (SOU1 CSU53**), Sou⁻*
pEA144 (SOU1 CSU53**), Sou⁻

Fig. 3

B

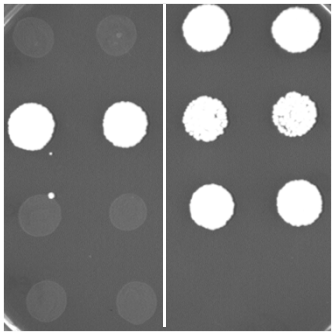
L-Sorbose, 3 days, 37°C

pRC2312; - Control (empty vector), Sou$^-$

pCA88; + Control (*SOU1*), Sou$^+$

pCA135 (*SOU1*, Region 135), *S*ou$^-$

pEA105 (*SOU1 CSU53***), Sou$^-$



pEA227 (*SOU1 CSU53*** with f.-sh.), Sou$^+$

pEA201 same as above

pEA158 same as above

Fig. 3

C

L-Sorbose,

2 days, 37 °C

pRC2312; – Control (empty vector), Sou^{-}

pCA88; + Control (*SOU1*), Sou^{+}

pEA104 (*SOU1 ASU53*◄), Sou^{++}

pEA234 (*SOU1 ASU53*►), Sou^{++}

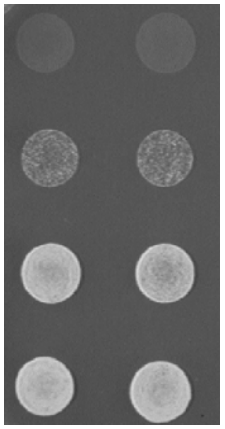


Fig. 3

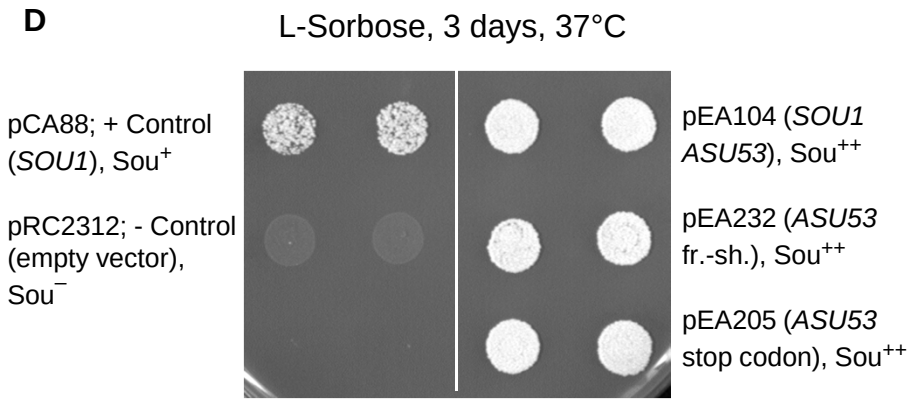

D
L-Sorbose, 3 days, 37°C
pCA88; + Control (SOU1), Sou+
pRC2312; - Control (empty vector), Sou–
pEA104 (SOU1 ASU53), Sou++
pEA232 (ASU53 fr.-sh.), Sou++
pEA205 (ASU53 stop codon), Sou++

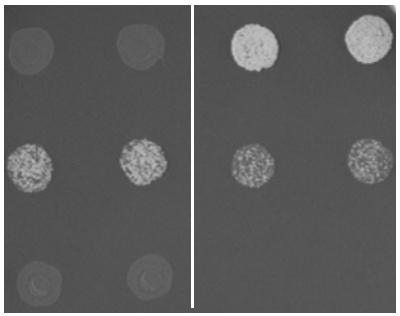
Fig. 3

E L-Sorbose, 2 days, 37°C

pRC2312; - Control (empty vector), Sou^{-}

pCA88; + Control (*SOU1*), Sou^{+}

pEA105 (*SOU1 CSU53***), Sou^{-}



pEA104 (*SOU1, ASU53*), Sou^{++}

pEA162 (*SOU1, ASU53 CSU53***), Sou±

Fig. 3

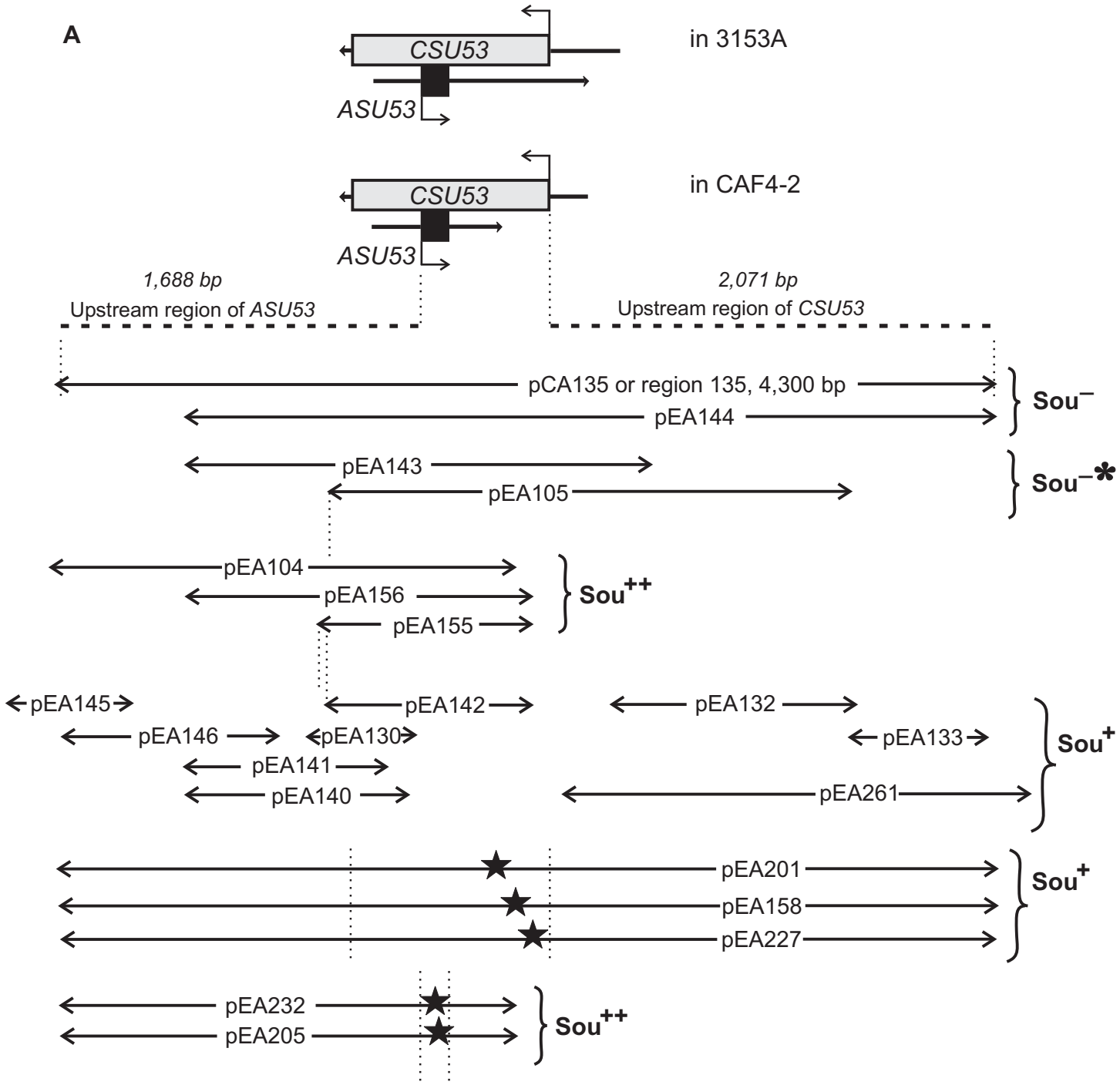

A
CSU53
ASU53
in 3153A
CSU53
ASU53
in CAF4-2
1,688 bp
Upstream region of ASU53
2,071 bp
Upstream region of CSU53
pCA135 or region 135, 4,300 bp
pEA144
Sou−
pEA143
pEA105
Sou−*
pEA104
pEA156
pEA155
Sou++
pEA145
pEA142
pEA132
pEA146
pEA130
pEA133
pEA141
pEA140
pEA261
Sou+
pEA201
pEA158
pEA227
Sou+
pEA232
pEA205
Sou++

Fig. 4

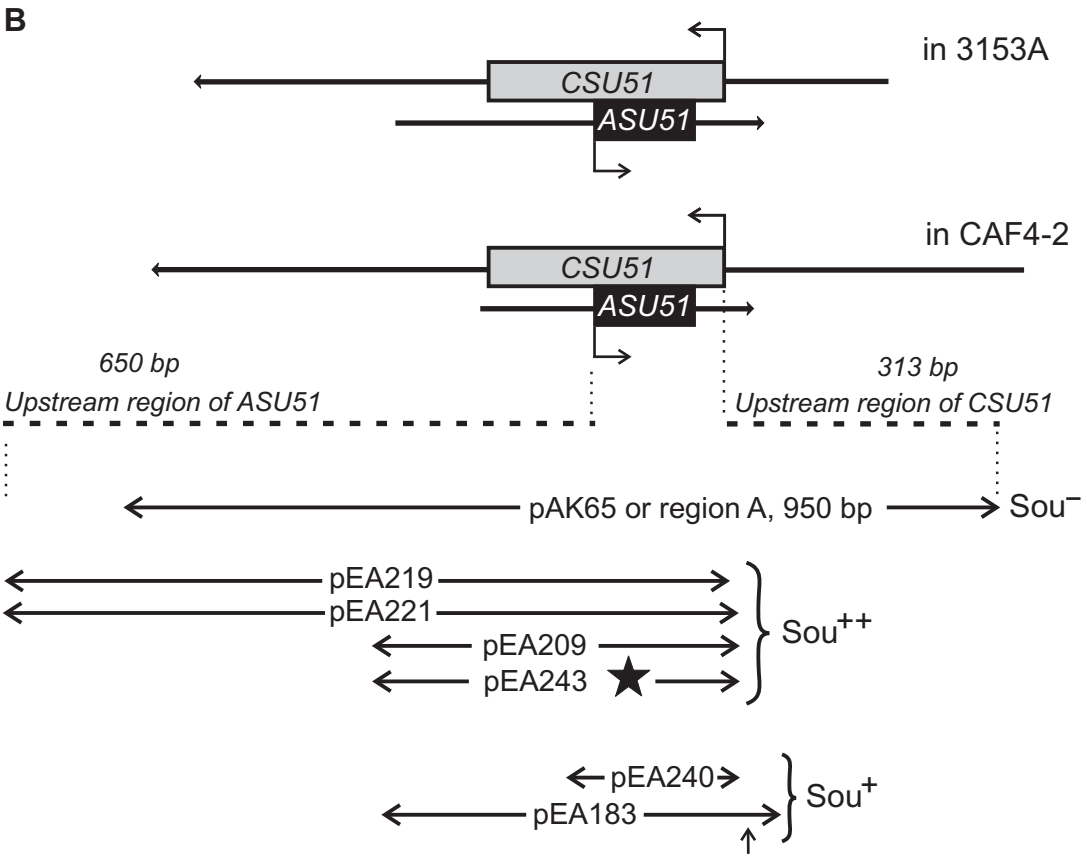

B
in 3153A
CSU51
ASU51
in CAF4-2
CSU51
ASU51
650 bp
Upstream region of ASU51
313 bp
Upstream region of CSU51
pAK65 or region A, 950 bp
Sou⁻
pEA219
pEA221
pEA209
pEA243
Sou++
pEA240
pEA183
Sou+

Fig. 4B

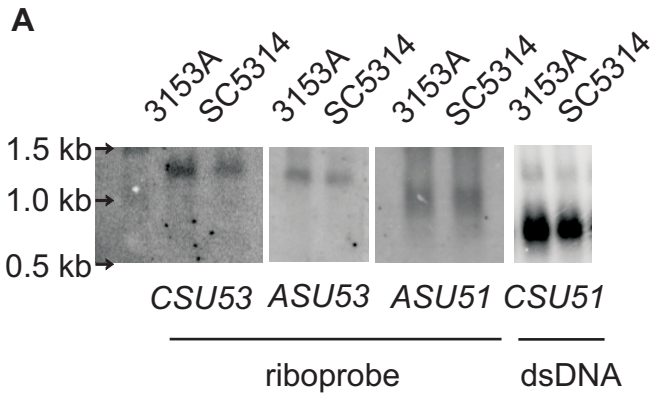

A
3153A
SC5314
3153A
SC5314
3153A
SC5314
3153A
SC5314
1.5 kb
1.0 kb
0.5 kb
CSU53
ASU53
ASU51
CSU51
riboprobe
dsDNA

Fig. 5

B

CSU53/ASU53

```
  1  ATGTCTTCTACACAAAAACAAAAGAAGGGTGCTGTTGATTTCGTTGCTGG
 51  GGGTGTTGCTGGTTTATTCGAAGCTTTGTGTTGTCATCCATTAGATACCA
101  TCAAAGTGAGAATGCAATTATACAAAAAATCCGGTCAAAAACCACCAGGT
151  TTTATTAAAACAGGTGTCAATATTGTCCAAAAAGAAGGATTTTTGTCATT
201  ATATAAAGGTTTGGGGGCCGTTGTTATTGGTATTGTGCCAAAAATGGCTA
251  TCAGATTCAGTTCATATGAATTTTACCGTTCATTCTTTTTAGACGAAAAT
301  GGCAAAATTTCCACTGGTAAGACTTTCCTTGCTGGTGTTGGTGCTGGTAT
351  TACCGAATCTGTCATGGTTGTTAATCCTATGGAAGTTGTGAAAATTAGAT
401  TACAAGCACAACATCATTCTATGAAGGACCCATTGGACATTCCAAAATAC
451  AGAAACGCTCCTCATGCTGCATATCTTATTGTCAAGGAAGAAGGTTTCAG
501  TACTTTATACCGTGGTGTTTCTTTAACTTGTGCCAGACAAGCTACCAACC
551  AAGGTGCTAACTTTGCTACATATTCTACCATCAAAGCATATCTTCAAAAA
601  CAACAAAACACTGAATTATTACCAGCATGGCAAACCAGTATTGTCGGTTT
651  GATTTCTGGTGCAGTCGGTCCATTAACCAATGCTCCATTGGATACCATTA
701  AAACAAGATTACAAAAGAGTAAGTTTACCAACAAGGAAAACGGATTGGTT
751  CGTATTGTCAAAATCGGTAAACAATTAGTCAAAGAAGAAGGTATTAACGC
801  TTTGTACAAGGGTATCACTCCAAGAATCATGAGAGTTGCTCCAGGTCAAG
851  CTGTGGTATTCACAGTGTATGAAGCTGTCAAACATTATTTGACAAATGAA
901  CCTACTGCTTAA
```

CSU51/ASU51

```
  1  ATGCAATTCACCAAAGTTATCGCTTCATTAGCCTTAGTTGCTTCCATCAA
 51  CGCTAAATTCCACAACACCACTGGTGCTGCTGCTGGTAACGGTACCGTTG
101  CTGGTGGTTCCAACTCTACTGGTGGTGCTGGTTCCCATAACGGTACTGGT
151  GCTTCTAACGGTTCTTCCAGTAAATCTTCTGGTTCCGGTGCTGCTGTCAA
201  CTCCGTCACTGGTTTGGCTGCTTTAGCTGCTGTTGGTGCTGCTTTATTGT
251  ACTAG
```

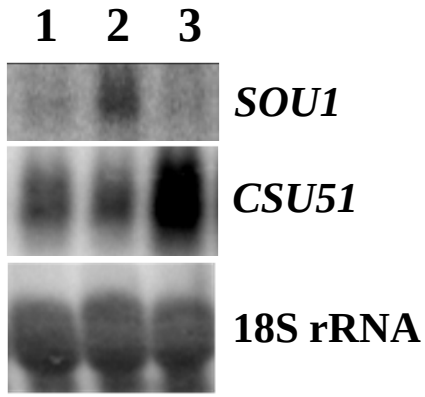


Fig. 6