Published in final edited form as:

Nat Struct Mol Biol. 2016 May ; 23(5): 450–455. doi:10.1038/nsmb.3192.

Expansion of antisense lncRNA transcriptomes since the loss of RNAi

Eric A. Alcid[1,2] and **Toshio Tsukiyama**[1]

[1]Divison of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, U.S.A

[2]Molecular and Cellular Biology Program, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, Washington, U.S.A

Abstract

Antisense long noncoding RNAs (ASlncRNAs) have been implicated in regulating gene expression in response to physiological cues. However, little is known about ASlncRNA evolutionary dynamics, and what underlies the evolution of their expression. Here, using budding yeast species *Saccharomyces* and *Naumovozyma* as models, we show that ASlncRNA repertoires have expanded since the loss of RNAi, in terms of their expression levels, their lengths, and their degree of overlap with coding genes. Furthermore, we show RNAi is inhibitory to ASlncRNA transcriptomes, and that elevation of ASlncRNAs in the presence of RNAi is deleterious to *Naumovozyma castellii*, a natural host of RNAi. Together, our work suggests that the loss of RNAi had a substantial impact on the genome-wide increase in expression of ASlncRNAs across budding yeast evolution.

Keywords

lncRNAs; antisense RNAs; evolution; RNAi

Recent advancement in genome-wide analyses of RNA has revealed that long noncoding RNAs (lncRNAs) are transcribed throughout eukaryotic genomes. One class of lncRNAs includes those that overlap with open reading frame boundaries in an anti-sense orientation (ASlncRNAs). Many lncRNAs have been shown to play key regulatory roles in metazoans, such as *HOTAIR* and *Xist*[1–4]. In the budding yeast *Saccharomyces cerevisiae*, ASncRNAs that overlap with the *GAL10*[5] and *PHO84* genes[6–8] repress the expression of overlapping mRNAs in response to environmental cues. However, given the large number of ASlncRNAs expressed in eukaryotes, biological functions of the vast majority of them remain unknown. Similarly, while evolutionary principles of mRNA and intergenic lncRNAs expression have

Correspondence should be addressed to T.T., (; Email: ttsukiya@fredhutch.org

Accession Codes

Sequencing reads have been deposited to the Sequence Read Archive under Bioproject SRP056928

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

been extensively studied[9,10], the evolutionary dynamics of ASlncRNAs have not been determined across any species phylogeny. As a result, how ASlncRNA transcriptomes evolve, and what affects their evolution, remain largely unknown. To determine how ASlncRNA transcriptomes have evolved, we used budding yeast as a model and found evidence that the loss of RNA interference (RNAi) has permitted expansion of the ASlncRNA transcriptomes among *Saccharomyces* species.

RESULTS

Expression of ASlncRNAs at *PHO84* and *GAL10*

To survey ASlncRNA expression across the budding yeast phylogeny, we first measured antisense expression at the *PHO84*[6–8] and *GAL10*[5] genes which attenuate transcription of overlapping mRNAs in *Saccharomyces cerevisiae,* relative to two control genes *IPP1* or *ACT1*. Strand-specific reverse transcription followed by quantitative PCR across six species of budding yeast revealed that, at *PHO84,* antisense expression is at a very low level in *Naumovozyma castellii,* and is more highly expressed in the *Saccharomyces* species (Fig. 1a). In contrast, antisense expression at the *GAL10* locus is very low in *N. castellii* and *Saccharomyces uvarum,* but robustly expressed in all other species of budding yeast that were tested (Fig. 1b). Antisense *GAL10* expression was not measured in *Saccharomyces kudriavzevii* due to degeneration of the 3' end of the gene and deterioration of *GAL* genes (including *GAL10*) in *S. kudriavzevii*[11]. The differences in antisense *PHO84* and antisense *GAL10* is not due to the differences in the expression of *IPP1* and *ACT1*, which were used as control mRNAs, because they are expressed at similar levels in our RNAseq data across all species tested (w Fig S1a, b). Together, these results are consistent with the possibility that levels of ASlncRNA expression might have globally increased since divergence from *N. castellii.*

ASlncRNA transcriptomes in budding yeast species

To globally identify ASlncRNAs across budding yeast evolution, we next performed strand-specific, high-throughput RNA sequencing in *N. castellii*, *S. uvarum, S. kudriavzevii, Saccharomyces mikatae,* and *S. cerevisiae* to measure steady-state ASlncRNA levels genome-wide, using total RNA depleted of ribosomal RNA. We adopted the dUTP method, as this has been shown to be the leading protocol for strand-specific, high throughput RNA sequencing[12] (Supplementary Table S1). We utilized the Yeast Gene Order Browser[13] and homology searches to identify a total of 5,031 orthologous genes for each species (Supplementary Table S2). It has been demonstrated that absolute levels of mRNA transcripts per cell across the budding yeast species we tested do not significantly vary[14,15]. We therefore counted all RNA reads from each species, then quantified and normalized antisense reads mapping to every orthologous gene using a negative binomial distribution for every species, to serve as a proxy for the ASlncRNA transcriptome[16].

To initially determine the similarity of global ASlncRNA profiles among the species, we performed principal component analysis using open reading frame (ORF)-antisense expression values. Along the first principal component, the ASlncRNA transcriptomes of the genus *Saccharomyces* yeast species clearly separated away from *N. castellii,* and clustered

together, suggesting that the difference in ASlncRNA transcriptomes between *N. castellii* and the rest of the species explains a substantial portion of the variance (Fig. 1c). We observed a similar clustering trend among the *Saccharomyces* species for sense RNA transcriptomes along the first principal component (Supplementary Fig. S1c). Furthermore, ASlncRNA transcriptome similarity correlated with the budding yeast phylogeny, as measured by Spearman's rho correlation coefficient (Supplementary Fig. S2b). It should be noted that the *S. cerevisiae* ASlncRNA transcriptome clearly separated from the other species along the second principal component, which might be due to selection in labs, and is consistent with previous reports suggesting that *S. cerevisiae* often acts as an outlier in growth assays, though this needs to be further investigated[17]. Together, this data suggests extensive rewiring of ASlncRNA transcriptomes since divergence from *N. castellii.*

To investigate the evolution of ASlncRNA transcriptomes further, we constructed distance matrices for each species using the Jensen-Shannon distance metric[18], and constructed ASlncRNA and mRNA expression trees (Fig 1d, Supplementary Fig. S1d). Both the ASlncRNA and mRNA expression trees resolve the relationship between *N. castellii* and the *Saccharomyces* species. However, when total tree branch length is measured, the ASlncRNA expression tree is much greater. This is likely due to mRNA transcriptomes being evolutionarily much more stable and highly conserved, making the tree highly sensitive to even more subtle differences (Supplementary Fig. S2a). These results suggest that substantial changes in ASlncRNA transcriptomes occur across evolutionary transitions, and that they are much more divergent than mRNA transcriptomes.

We next investigated how the global levels of ASlncRNA transcripts have changed along the budding yeast phylogeny. When we measured the distribution of the transcript levels of all ASlncRNAs overlapping 5031 orthologous ORFs, we found a clear increase in ASlncRNA levels across budding yeast evolution since divergence from *N. castellii* ($p \ll 2.2e^{-16}$ for *S. cerevisiae* and *N. castellii,* Wilcoxon rank-sum test, Fig. 2a, c). This increasing pattern was not found when mRNA distributions for each species were assessed ($p = 0.9583$ for *S. cerevisiae* and *N. castellii,* Wilcoxon rank-sum test, Fig. 2b, d). This striking result suggested that ASlncRNA transcriptomes in budding yeast started rapidly expanding immediately after divergence from *N. castellii.* The increase in ASlncRNA expression since divergence from *N. castellii* could have come from at least two possible sources: transcription termination defects at convergent genes, or divergent promoters at nucleosome-depleted regions (NDRs) at genes arranged in tandem that overlap the upstream gene (Fig. 2e and 2f, bottom). To assess the possible contributions of termination defects and divergent promoters to the ASlncRNA transcriptome, we separated all ORFs into whether they are arranged convergently with their downstream gene, or in tandem. We then measured antisense tag density for each gene, and performed metagene analysis for each orientation category for each species. For every species, expression of ASlncRNA from convergent genes was ~4-fold higher than ASlncRNA arising from tandem genes (Fig. 2e and f, Supplementary Fig. S3a). At convergently oriented genes, the ASlncRNA levels were consistently higher in the *Saccharomyces* species than in *N. castellii,* with the difference ranging from more modest *(S. uvarum*) to large *(S. cerevisiae)* amounts (Fig 2e). It should be noted that, for the convergent genes analyzed, the transcripts analyzed for the downstream genes were in the sense orientation which shows striking similarities in abundance between

yeast species (Fig. 2e and f). Similarly, antisense levels were low across the gene bodies for tandemly oriented genes in *N. castellii*, and were consistently higher in *Saccharomyces* species, with the highest levels in *S. cerevisiae*. Furthermore, the difference between these two species at tandem genes was even more pronounced than at convergent genes, (~8 vs ~4 fold, respectively) (Fig. 2e and f, bottom panels, Supplementary Fig. S3a). Similar analysis examining sense (mRNA) expression revealed minimal differences between all species (Supplementary Fig. S3b,c). Taken together, this analysis revealed that, after the divergence from *N. castellii*, ASlncRNA levels have increased at both convergent and tandem genes, though more so at tandem genes, suggesting that increased divergent transcription is one of the driving forces underlying robust ASlncRNA transcription programs in *the* genus *Saccharomyces*.

We next determined the lengths of the ASlncRNAs across the budding yeast species. To this end, we identified all putative ASlncRNA units in all species, which afforded genomic "start" and "end" coordinates (Supplementary Tables S3–7, see Methods,[19,20]). As shown in Fig. 3a, our analysis revealed that the length of the ASlncRNA transcripts in *N. castellili* (mean 571 bases, median 324 bases) was significantly shorter than that in *Saccaromyces* species (mean 626 bases, median 436 bases, $p \ll 2.2e^{-16}$, two-sample Kolmogorov–Smirnov test). This result suggested a possibility that the extent to which ASlncRNAs overlap with mRNAs might be different between *N. castellili* and *Saccaromyces* species. To test this model, we identified all putative ASlncRNA units in all species and calculated the number of base pairs each ASlncRNA overlaps with its cognate ORF (Supplementary Tables 3–7, see Methods[19,20]). Supporting our model, this analysis revealed that the ASlncRNA transcripts in *N. castellili* overlap with ORF boundaries much less extensively as compared to *Saccharomyces* species (Fig. 3b). Together, these results showed that budding yeast species expanded ASlncRNA transcriptomes in terms of the steady-state levels, the lengths, as well as the degree of overlap with mRNAs after divergence from *N. castellili*.

We predict that ASlncRNAs playing important biological roles more likely represent discrete transcription units, rather than transcription noise. If an ASlncRNA and a mRNA share a preinitiation complex (PIC) at their initiation sites, it is possible that the ASlncRNA is transcribed by a RNA polymerase that is recruited for mRNA transcription. In this case, the ASlncRNA may represent transcriptional noise, or erratic mRNA initiation. On the other hand, if a PIC is formed at an ASlncRNA initiation site and not shared by a neighboring mRNA, the PIC is likely dedicated for the ASlncRNA. This implies that the ASlncRNA is a discrete transcription unit, and is meant to be transcribed. Notably, 33% of ASlncRNAs in *S. cerevisiae* transcribed from divergent promoters have a PIC dedicated to them ($p = 0.01$, hypergeometric test) based on high-resolution PIC (TFIIB) mapping data[21], suggesting that they are discrete transcription units.

The effects of the exosome on ASlncRNA evolution

The majority of lncRNAs, including ASlncRNAs, are rapidly degraded by the exosome, a highly conserved exonuclease[22,23]. Mutation of the exosome would then lead to the identification of so-called cryptic unstable transcripts (CUTs)[22,23]. Because all the analyses so far were performed in the presence of fully functional exosome, ASlncRNAs identified

thus far are considered stable unannotated transcripts (SUTs). We therefore investigated how the levels of CUT-ASlncRNAs have changed since *S. cerevisiae* and *N. castellii* diverged. To this end, we mutated *RRP6,* an exosome component, in *N. castellii* and globally compared its cryptic ASlncRNA transcriptome to that of *S. cerevisiae*[19]. As reported[22,23], the abundance of ASlncRNAs strongly increases in *S. cerevisiae* when *RRP6* is mutated (Figure 4a: note that only ASlncRNAs that increase in abundance in *rrp6* mutant were analyzed[23]), due to stabilization of CUTs ($p \ll 2.2e^{-16}$, Wilcoxon rank-sum test). In *N. castillii*, *DCR1* can also degrade ASlncRNA-mRNA duplexes, which can confound our analyses of CUT-ASlncRNAs. We therefore introduced null *RRP6* mutations in *N. castellii* in a *dcr1* background. As was the case in *S. cerevisiae*, our analyses revealed that deletion of *RRP6* in *N. castellii* caused a significant increase in ASlncRNA levels when compared to the control strain (*dcr1* alone) ($p \ll 2.2e^{-16}$, Wilcoxon rank-sum test, Fig 4a). However, the ASlncRNA levels in *S. cerevisiae rrp6* strain were still much higher than that of *N. castellii dcr1 rrp6* strain. As a result, the difference in the ASlncRNA levels between *S. cerevisiae rrp6* and *S. castillii dcr1 rrp6* mutants was comparable, if not larger, than that between wild type *S. cerevisiae* and *S. castillii dcr1* mutant (Supplementary Fig. S4a–c). Together, this data suggests that, similar to SUT-ASlncRNA expression (Fig. 2), CUT-ASlncRNA expression has also increased since divergence from *N. castellii.*

The effects of RNAi on ASlncRNA evolution

We next sought to identify the basis for the relative increase in ASlncRNA expression in *Saccharomyces* budding yeast since divergence from *N. castellii*. One pathway present in *N. castellii* and absent in *Saccharomyces* lineage that can affect the stability of ASlncRNAs is RNA interference (RNAi)[24,25]. If both mRNA and ASlncRNAs are transcribed from the same locus, they can form double strand RNA, which can be processed by RNAi machinery, destabilizing both mRNA and ASlncRNA transcripts genome-wide[26]. Indeed, we have recently demonstrated that global elevation of ASlncRNA levels in the presence of reconstituted RNAi in *S. cerevisiae* is deleterious[19]. Therefore, it is conceivable that the loss of RNAi in the *Saccharomyces* lineage has alleviated the selective pressure to attenuate ASlncRNA levels genome-wide. In support of this, *S. uvarum,* which still retains *DCR1,* globally expresses ASlncRNAs at a level intermediate to *N. castellii* and other *Saccharomyces* species (Fig. 2). To test whether RNAi can have a negative effect on ASlncRNA expression, we compared genome-wide levels of ASlncRNAs in our wild type *S. cerevisiae* strain, and an *S. cerevisiae* strain where RNAi was reconstituted[19,24]. This analysis showed that reconstitution of RNAi led to a significant decrease of ASlncRNA expression at both convergently and tandemly oriented genes (Fig. 4b and c, $p \ll 2.2e^{-16}$ for both orientations, Wilcoxon rank-sum test: note that only ASlncRNAs that increase abundance in *rrp6* mutant were analyzed). Interestingly, we found that disabling RNAi in *N. castellii* by *dcr1* mutation had no statistically significant effect on endogenous ASlncRNA levels at both convergent and tandem genes ($p = 0.52$ and $p = 0.08$, respectively, Wilcoxon rank-sum test) (Fig. 4b and 4c), suggesting that *N. castellii* may have mechanism(s) to alleviate the effects of RNAi on ASlncRNA stability. The apparently higher antisense read counts of wild type *N. castellii* over *dcr1* mutant (Fig. 4b,c) were not statistically significant ($p = 0.07185$, Wilcoxon rank-sum test).

To further test our model, we next investigated the phenotypic consequences of expressing ASlncRNAs while maintaining RNAi machinery in the genome in *N. castellii,* a natural host of RNAi[25]. Mutation of *RRP6* in *N. castellii* led to a slow growth phenotype (Fig 4d) at elevated temperature. This suggested that, although the effect of this mutation on the abundance of ASlncRNAs was not as strong as in *S. cerevisiae* (Fig 4a), it did cause a fitness defect in *N. castillii*. If this temperature sensitivity was at least partly due to RNAi globally destabilizing transcripts, deletion of *DCR1* was expected to rescue the growth defect. As shown in Figure 4d, this turned out to be the case, supporting our model that processing of mRNA-ASlncRNA hybrids by RNAi is at least one of the underlying mechanisms by which RNAi has helped maintain low levels of ASlncRNA expression in the *N. castellii* genome. This could be due to compromised heat shock response by elevated ASlncRNAs in the presence of RNAi, which has been observed in *S. cerevisiae*[19]. Furthermore, the partial rescue of the growth defects of the *rrp6* mutant by *DCR1* mutation is associated with a modest, though statistically significant increase in the levels of all ASlncRNAs identified in the *rrp6 dcr1* mutant ($p = 3 \times 10^{-9}$, Wilcoxon signed-rank test, Supplementary Fig. S4d). Among all ASlncRNAs with statistically significant differences in levels (increase or decrease) ($p <= 0.05$, negative-binomial distribution) between the *rrp6* and *rrp6 dcr1* mutants, we found that the levels of these ASlncRNAs mostly increased upon *DCR1* mutation in a *rrp6* background ($p = 0.00264$, Wilcoxon signed-rank test, Supplementary Figure S4e), suggesting that abrogated siRNA production might underlie the enhanced growth of *rrp6 dcr1* double mutant. Together, these data support our model in which the loss of RNAi enabled the global elevation of ASlncRNAs across the budding yeast phylogeny.

Discussion

We have shown that global ASlncRNA transcriptomes have significantly expanded in *Saccharomyces* species of budding yeast after divergence from *N. castellii*, in terms of steady-state levels, lengths and the degrees of overlaps with mRNAs. We further provided supporting evidence that the loss of RNAi has alleviated the selective pressure to maintain the expression levels of ASlncRNA low, allowing steady expansion of ASlncRNA transcriptome in *Saccharomyces* species. To our knowledge, this is the first report to provide evidence that RNAi profoundly affects the evolution of lncRNA transcriptomes, though it has been speculated before[20]. In this regard it is interesting to note that, despite possessing an active RNAi pathway, *N. castellii* still has detectable antisense expression at a large number of genes. This is analogous to many higher eukaryotes that keep RNAi while having abundant lncRNAs transcribed. As such, organisms that maintain both RNAi and ASlncRNAs likely possess currently unknown mechanisms that mitigate the deleterious effects of having both systems coexist. Given that RNAi attenuates ASlncRNAs, and elevation of ASlncRNAs in the presence of RNAi leads to a substantial fitness cost to both *S. cerevisiae* and *N. castellii*, it is likely that the incompatibility between the presence of RNAi and high levels of ASlncRNA transcription extends to metazoans.

ONLINE Methods

Yeast strains

A list of all strains used in this study can be found in Supplementary Table S8. The identities of the strains were confirmed by RNA-seq. We carried out single-step gene deletions by standard lithium acetate transformation using NatMX drug-resistance markers as described for *S. cerevisiae* [28]. For *N. castellii,* we performed gene deletions as previously described[29]. Strains were also created using standard genetic crosses. For *S. cerevisiae,* genome sequences and annotations were downloaded from Ensembl[30] or the Saccharomyces Genome Database[31]. For all other yeast strains, genome sequences and annotations were downloaded from the Yeast Gene Order Browser[13].

Yeast growth conditions

Strains were cultured at 30°C or 25°C in YPD until OD600 = 0.4–0.7 before being harvested for RNA using standard hot acid phenol extraction.

Strand-specific library preparation and high-throughput RNA sequencing

For every strain, 3μg of Total RNA was depleted of ribosomal RNA species using Ribo-Zero magnetic rRNA removal kit (Human/Mouse/Rat) (Epicentre). Strand-specific libraries were then prepared using the dUTP method combined with TruSeq (Illumina) as previously described [32,33]. Our protocol includes actinomycin D during reverse transcription construction to prevent artifacts[32]. 50 cycles of paired-end sequencing was performed on an Illumina HiSeq 2500 on either high-output mode or rapid run mode (FHCRC Shared Resources). All sequencing experiments were performed in biological duplicate.

Identification of orthologous genes among *S. cerevisiae, S. mikatae, S. kudriavzevii, S. uvarum, N. castellii*

An initial set of orthologous genes was identified using the “Pillars.tab” file from YGOB, corresponding to 4894 orthologous genes. To identify additional orthologous genes, we aligned all open reading frame amino acid sequences for each species to all open reading frame amino acid sequences for *S. cerevisiae* using LAST[34]. We then identified the 20th percentile alignment score and set this as the minimum threshold. All remaining amino acid sequences not previously identified in “Pillars.tab” but had an alignment score at or above the minimum threshold were then identified as additional orthologs, resulting in a total of 5031 gene orthologs shared among the 5 yeast species.

RNA-seq analysis

Alignment—Reads were aligned to the species-specific genome using TopHat2[35] with the following settings: tophat2 -p 4 -G <gene_annotation_file> -I 2000 --library-type=fr-firststrand -o <output_directory> <bowtie_index> <Read1.fastq> <Read2.fastq>. Reads were then trimmed of adapter sequences with a custom Python script using the Python module HTSeq[36].

Heuristic of RNA-seq data to identify putative untranslated regions (UTRs)— Because it is possible that ASlncRNAs might overlap mRNA transcripts at untranslated regions (UTRs), we identified putative UTRs by finding local minima of sequencing read density within 300 basepairs of open reading frame (ORF) boundaries. After reads were aligned, reads were filtered such that only properly aligned, uniquely mapped reads were kept using a custom Python script and pysam[37]. After confirming high reproducibility of replicates, reads for each replicate were combined to make per-base, strand-specific pileup files using pysam. Using this pileup file, putative 5' and 3' UTRs were identified by starting at either the start codon or stop codon coordinate, respectively, for each orthologous gene and extending away from the open reading frame boundary until a local minimum in the per/bp read density was encountered within 300 bp from the gene boundary. The coordinate where this is achieved served as the outer UTR coordinate. A custom python script was written for this implementation (available upon request).

Identification of ORFs with differentially expressed antisense reads—Using the putative orthologous transcript list (with adjusted UTRs) for each species, differentially expressed ASlncRNA units were defined by first enumerating the number of reads in each replicate that overlap antisense to each transcript, then using a negative binomial distribution (R-package DESeq2)[16] to determine differential expression. ASlncRNAs that had a p-adjusted value <= 0.2 were determined to be differentially expressed. Fold-change, as well as absolute expression (in normalized count values) were determined using DESeq.

Construction of CUT-ASlncRNA distributions—To identify CUT-ASlncRNAs, only ASlncRNAs whose log2-fold change >= 0 were kept, leading to 2420 and 2481 CUT-ASlncRNAs for *S. cerevisiae* and *N. castellii*, respectively. This was done as previously described[23]. These populations were then used as distributions for boxplots and histograms.

Meta-analyses of RNA-seq data (Fig. 2, Supplemental Fig. S5 and S6)—To perform meta-analysis, we first normalized reads/per-base coverage files by the genome-wide average, excluding tRNA and rRNA loci. Full-length transcripts (starts and ends adjusted by putative UTRs) were then binned into 10 equally-sized bins, while upstream regions, downstream regions, and intergenic regions, were divided into 3 equally sized bines. Every binned region was then aligned by the putative transcription start-site, and the average of each aligned bin was found. This data was used to construct the ribbon plots (see below).

Segmentation heuristic of RNA-seq data to identify putative transcript units

After reads were aligned, reads were filtered such that only properly aligned, uniquely mapped reads were kept using a custom Python script and pysam [37]. Because replicates were highly reproducible (data not shown), reads for each replicate were combined to make per-base, strand-specific pileup files using pysam. Using this pileup file, putative transcript units were segmented by defining a minimum expression threshold, defined below. tRNAs, and rRNAs were excluded for every step in analysis.

Defining a threshold level using empirically determined tag density—For a known open reading frame (ORF), expression was calculated by the following equation:

$$\text{Tag Density} = [\sum \text{count}_i (i = \text{start}..\text{end}) / (\text{end} - \text{start})] / \text{count}_{\text{genome average}}$$

where *i* is the genomic position, *count* is the number of reads overlapping *i, end* is the last genomic position of the ORF, *start* is the beginning position of the ORF. This was repeated for every ORF in the genome. The threshold was defined by the bottom 5th percentile expression value for transcripts longer than 250 bp (inclusive). For transcripts between 100bp and 249 bps (inclusive), the threshold was the bottom 25th percentile expression value.

Segmentation heuristic of pileup files—Using the threshold defined above, putative transcripts were identified by computing the tag density within a 100bp sliding-window using a 1bp step size. “Starts” and “Ends” of transcript units were defined by whether the tag density exceeded the defined threshold and were at least 100 bp in length. Segments closer than 50 bp, and were less than 2-fold different in tag density, were joined, which is commonly performed. See above for threshold differences based on length.

Construction of heatmaps, plots, statistical and phylogenetic analysis

Heatmaps, plots, and meta-gene plots were constructed in R[27] using the packages “ggplot.” Jensen-shannon distance metrics were calculated as previously described[18]. Neighbor-joining trees were then created using the R-package “ape”[38]. 2-sided Wilcoxon rank-sum and Wilcoxon signed-rank tests were performed using the R function wilcoxon.test with the open “paired = FALSE” and “paired = TRUE”, respectively.

Strand-specific RT-PCR

Strand-specific RT-PCR was performed for *PHO84* and *GAL10* as previously described[39]. Calculation of relative expression was performed using the __Ct method, normalized to either *ACT1* or *IPP1.* The nucleotide sequences of the primers used are listed in Supplementary Table S9.

Gene Ontology Analysis

All gene ontology analysis was performed using GOSeq[40]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank H. Malik, I.A. Drinnenberg, and the members of the Tsukiyama lab for helpful discussions; H. Malik, I.A. Drinnenberg for critical reading of the manuscript; M. Dunham (University of Washington), D. Bartel (Massachusetts Institute of Technology) and D. Gottschling (Fred Hutchinson Cancer Research Center) for yeast strains; A. Marty and FHCRC shared resources for deep sequencing. This work was supported by a grant from US National Institutes of Health (R01 GM058465 to T.T.) and a predoctoral fellowship from US National Institutes of Health (F31 GM101944 to E.A.A.). E.A.A. contributed in planning and performing experiments, analyzing and interpreting data, and writing this manuscript. T.T. contributed in planning experiments, interpreting data and writing this manuscript.

References

1. Gupta RA, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–6. [PubMed: 20393566]
2. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–23. [PubMed: 17604720]
3. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–93. [PubMed: 20616235]
4. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell. 2013; 152:1308–23. [PubMed: 23498939]
5. Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. Molecular cell. 2008; 32:685–95. [PubMed: 19061643]
6. Camblong J, et al. Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae. Genes & development. 2009; 23:1534–45. [PubMed: 19571181]
7. Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell. 2007; 131:706–17. [PubMed: 18022365]
8. Castelnuovo M, et al. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. Nature structural & molecular biology. 2013; 20:851–8.
9. Necsulea A, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014; 505:635–40. [PubMed: 24463510]
10. Brawand D, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011; 478:343–348. [PubMed: 22012392]
11. Hittinger CT, Rokas A, Carroll SB. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:14144–14149. [PubMed: 15381776]
12. Levin JZ, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Meth. 2010; 7:709–715.
13. KPB, Wolfe KH. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. Genome research. 2005; 15:1456–1461. [PubMed: 16169922]
14. Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. PLoS Biol. 2010; 8:e1000414. [PubMed: 20625544]
15. Thompson DA, et al. Evolutionary principles of modular gene regulation in yeasts. eLife. 2013; 2
16. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11:R106. [PubMed: 20979621]
17. Warringer J, et al. Trait Variation in Yeast Is Defined by Population History. PLoS Genet. 2011; 7:e1002111. [PubMed: 21698134]
18. Merkin J, Russell C, Chen P, Burge C. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science. 2012; 338:1593–1599. [PubMed: 23258891]
19. Alcid EA, Tsukiyama T. ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. Genes Dev. 2014; 28:2348–60. [PubMed: 25367034]
20. Yassour M, et al. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. Genome Biol. 2010; 11:R87. [PubMed: 20796282]
21. Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature. 2012; 483:295–301. [PubMed: 22258509]
22. Neil H, et al. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature. 2009; 457:1038–42. [PubMed: 19169244]
23. Xu Z, et al. Bidirectional promoters generate pervasive transcription in yeast. Nature. 2009; 457:1033–7. [PubMed: 19169243]

24. Drinnenberg IA, Fink GR, Bartel DP. Compatibility with Killer Explains the Rise of RNAi-Deficient Fungi. Science. 2011; 333:1592. [PubMed: 21921191]
25. Drinnenberg IA, et al. RNAi in budding yeast. Science. 2009; 326:544–50. [PubMed: 19745116]
26. Lasa I, et al. Genome-wide antisense transcription drives mRNA processing in bacteria. Proceedings of the National Academy of Sciences. 2011; 108:20172–20177.
27. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.
28. Goldstein G, McCusker J. Three New Dominant Drug Resistance Cassettes for Gene Disruption in Saccharomyces cerevisiae. Yeast. 1999; 15:1541–1553. [PubMed: 10514571]
29. Krawchuck M, Wahls W. High-efficiency Gene Targeting in Schizosaccharomyces pombe Using a Modular, PCR-based Approach with Long Tracts of Flanking Homology. Yeast. 1999:1419–1427. [PubMed: 10509024]
30. Cunningham F, et al. Ensembl 2015. Nucleic Acids Res. 2015; 43:D662–9. [PubMed: 25352552]
31. Cherry JM, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012; 40:D700–5. [PubMed: 22110037]
32. Parkhomchuk D, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic acids research. 2009; 37:e123. [PubMed: 19620212]
33. Sultan M, et al. A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. Biochemical and biophysical research communications. 2012; 422:643–6. [PubMed: 22609201]
34. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011; 21:487–93. [PubMed: 21209072]
35. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013; 14:R36. [PubMed: 23618408]
36. Anders S, Pyl PT, Huber W. HTSeq – A Python framework to work with high-throughput sequencing data. 2014 bioRxiv.
37. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–9. [PubMed: 19505943]
38. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20:289–290. [PubMed: 14734327]
39. Chatterjee SN, Devhare PB, Lole KS. Detection of negative-sense RNA in packaged hepatitis E virions by use of an improved strand-specific reverse transcription-PCR method. Journal of clinical microbiology. 2012; 50:1467–70. [PubMed: 22205803]
40. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010; 11:R14. [PubMed: 20132535]

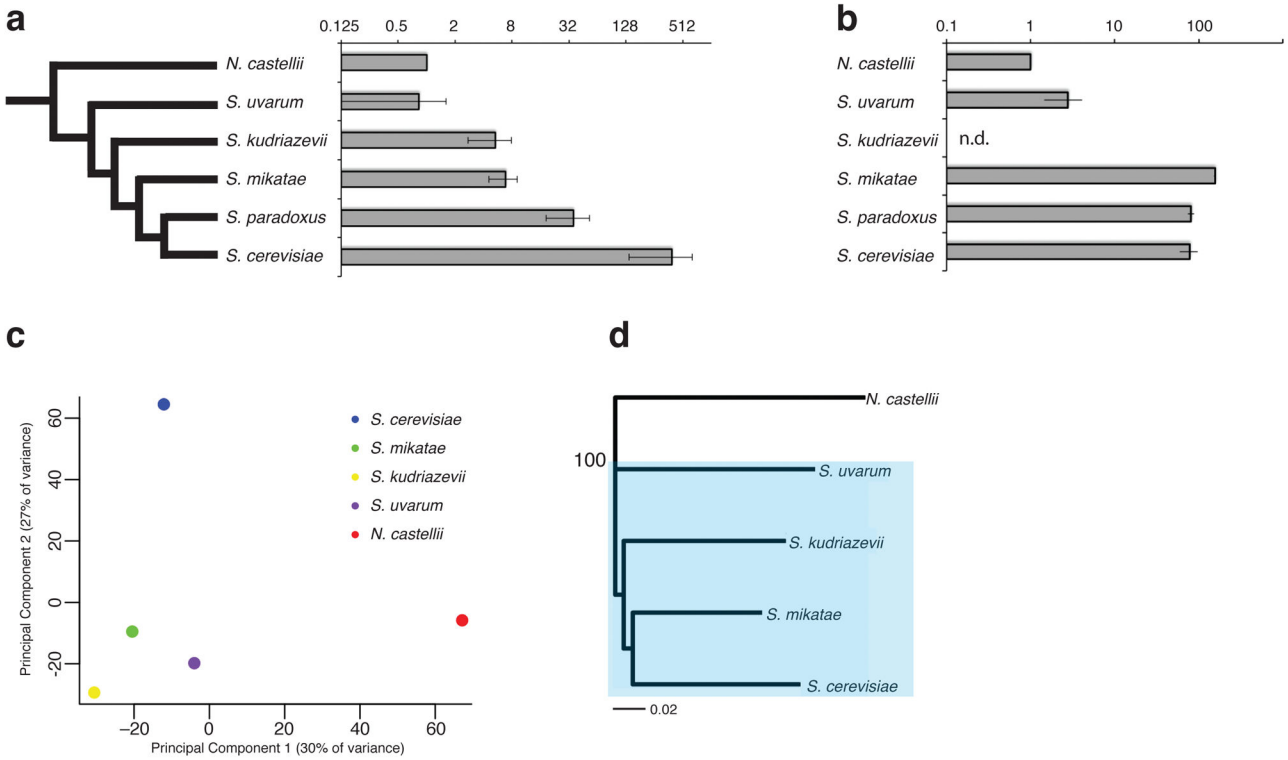

Figure 1. ASlncRNA expression patterns among budding yeast

(a) *PHO84* ASlncRNA expression levels. From left to right: Cladogram of *N. castellii* and *the* genus *Saccharomyces* budding yeasts, and expression levels of *PHO84* lncRNA for each species as determined by RT-qPCR. Expression level is relative to *N. castellii,* whose value was set to 1. The data is presented in logarithmic scale. (b) Expression levels of *GAL10* ASlncRNA for each species as determined by RT-qPCR. As in (a), the expression level is relative to *N. castellii,* which was set to 1. For (a) and (b), mean and standard error of the mean were determined using RNA isolated from 2 different cultures, 3 technical replicates per culture. The data is presented in logarithmic scale. (c) Principal Component Analysis (PCA) of ASlncRNA transcriptomes in budding yeast. (d) Neighbor-joining tree based on pairwise distance matrices (Jensen-Shannon distance metric) for the genus *Saccharomyces* budding yeasts and *N. castellii.* Bootstrap value showing *N. castellii* as an outgroup (out of 100). Highlighted in blue are all yeast species of the genus *Saccharomyces.*

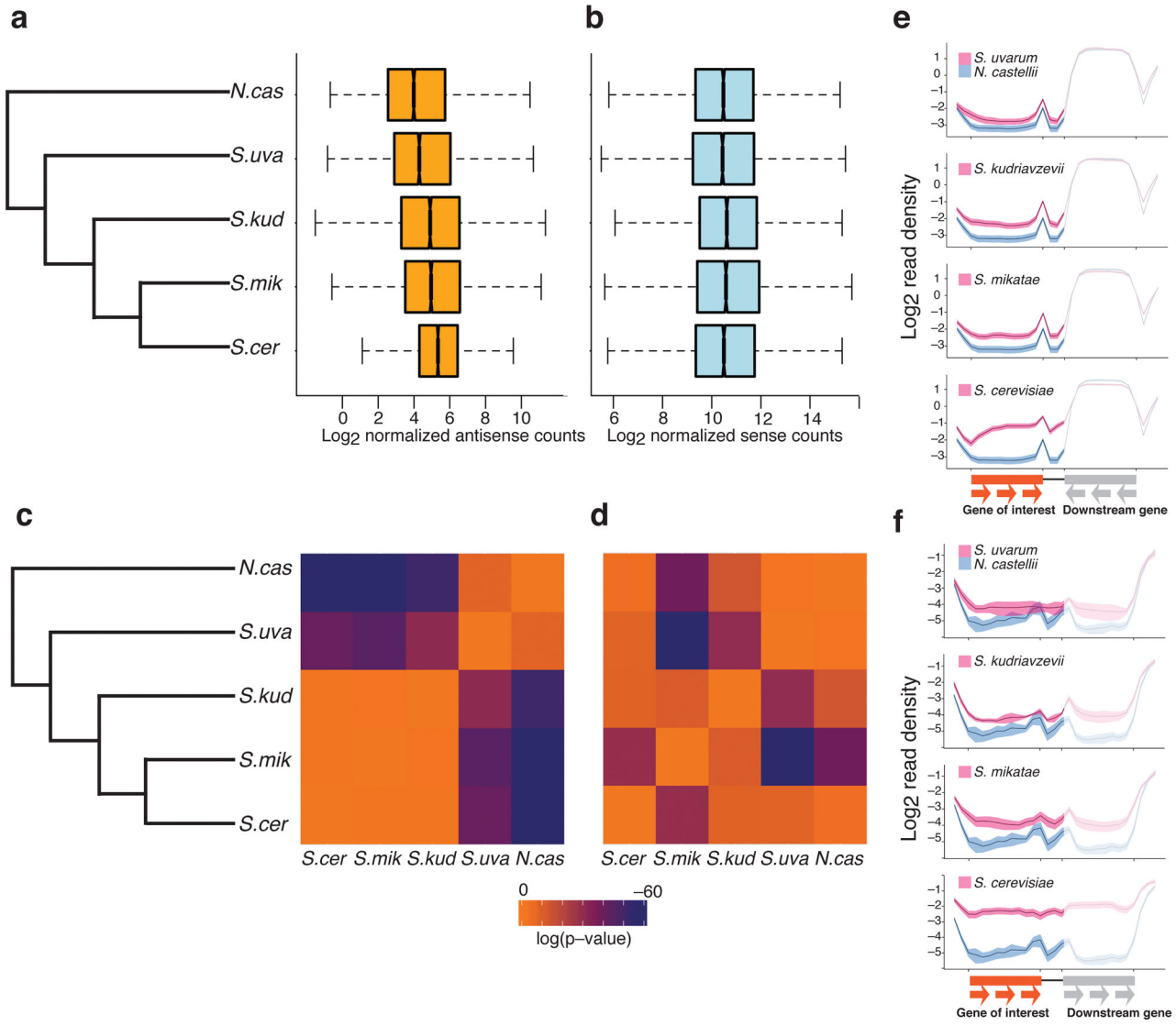

Figure 2. Elevation of ASlncRNA levels across budding yeast phylogeny

(a) Global ASlncRNA levels among budding yeast species. From left to right: Cladogram of *N. castellii* and *the* genus *Saccharomyces* budding yeast. *N. cas, S. uva, S. kud, S. mik* and *S. cer* denote *N. castellii,* S. uvarum, *S. kudriazevii, S. mikatae*, and *S. cerevisiae*, respectively. Boxplots of distributions of normalized read counts (log2 scale) mapping antisense 5031 orthologous open reading frames for budding yeasts. For all box plots, the midline represents the median value, the borders of the box represent the values at the 25th (first quartile) and 75th percentiles (3rd quartile), and the whiskers represent the following:upper whisker = min(max(x), Q_3 + 1.5 * IQR), lower whisker = max(min(x), Q_1 − 1.5 * IQR), where IQR = 3rd quartile value – 1st quartile value[27] . The notches surrounding the median value represent the 95% confidence interval estimation for the medians. Data for all RNA-sequencing experiments was collected from RNA extracted from two different isogenic cultures. (b) Global sense RNA levels among budding yeast species. As in (a), except reads mapping in the sense orientation. (c) Heatmap representation of pair-

wise Wilcoxon-rank-sum tests for ASlncRNA transcriptomes. (d) Heatmap representation of pair-wise Wilcoxon-rank-sum tests for mRNA transcriptomes. (e) Antisense read density at convergent genes in *Saccharomyces* species as compared to *N. castellii*. Ribbon Plots of antisense read density in log2-scale at genes arranged in convergent orientation for (top to bottom) *S. uvarum* (n = 3172 genes), *S. kudriavzevii* (n= 3208 genes), *S. mikatae* (n= 3438 genes), *S. cerevisiae* (n = 3656 genes). *N. castellii* (n = 3064 genes) is represented in all the plots by the blue ribbon. The lines represent the antisense RNA-seq signal, while the outer borders of the ribbon represent 1 standard-error of the mean away from the mean. (f) Antisense read density at tandem genes in *Saccharomyces* species as compared to *N. castellii*. Ribbon Plots of antisense read density in log2-scale at genes arranged in convergent orientation for (top to bottom) *S. cerevisiae* (n= 3046 genes), *S. mikatae* (n= 3366 genes), *S. kudriavzevii* (n= 3141 genes), *S.uvarum* (n=3146 genes), *N. castellii* (n= 2846 genes).

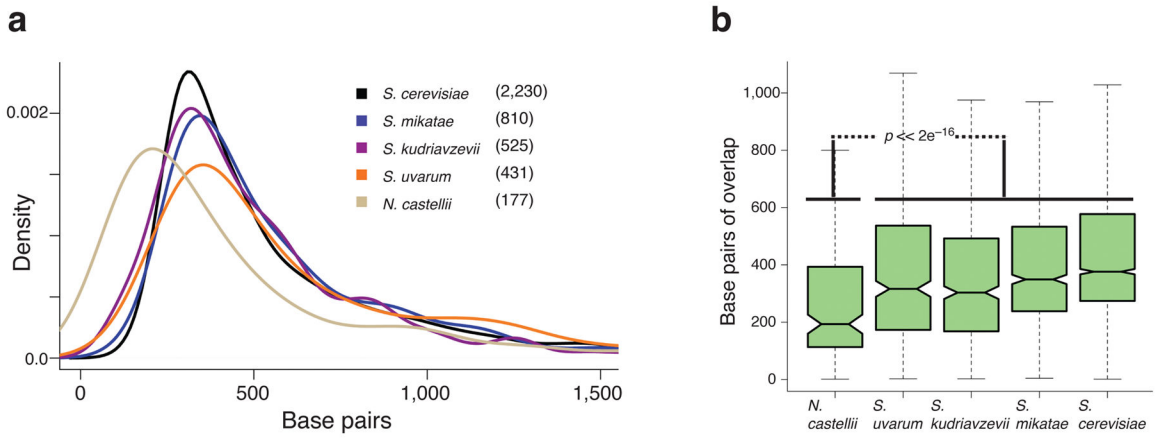

Figure 3. ASlncRNAs have increased in length, and overlapped mRNAs to a greater degree, since divergence from *N. castellii*

(a) Kernal density estimates of the length distributions of ASlncRNAs in the indicated species of budding yeast. The number of identified ASlncRNAs is shown in parentheses (see Methods). **(b)** Boxplot representation of the distribution of the amount of overlap in base-pairs between ASlncRNA-mRNA pairs in *S. cerevisiae* (n = 2543), *S. mikatae* (n = 810), *S. kudriavzevii* (n = 525), *S. uvarum* (n = 431), *N.castellii* (n = 177). P-value ($p \ll 2.2e^{-16}$) was determined using a two-sided Wilcoxon rank-sum test. See Figure 2 legend for description of boxplot features.

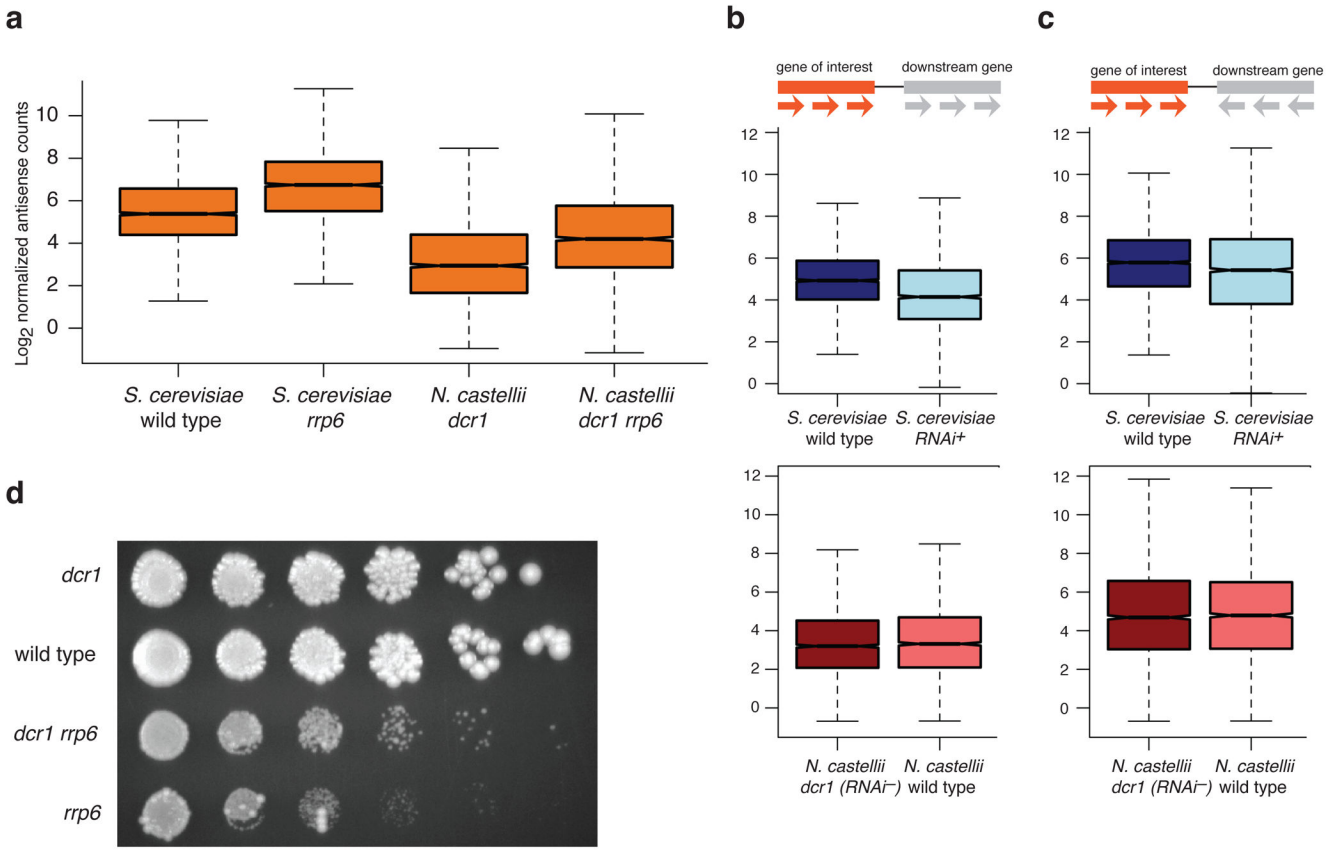

Figure 4. RNAi constrains ASlncRNA expression

(a) The effects of the exosome on global ASlncRNA levels in *S. cerevisiae* and *N. castellii*. Boxplots of distribution of normalized read counts at CUT-ASlncRNAs (ASlncRNAs that increase levels in *rrp6* mutant) in control and *rrp6* strains of *S. cerevisiae* (n = 2420), $p \ll 2.2e^{-16}$ determined using a two-sided Wilcoxon rank-sum test (*S. cerevisiae* WT vs *rrp6*). and *N. castellii* (n = 2481), $p \ll 2.2e^{-16}$ determined using a two-sided Wilcoxon rank-sum test (*N. castellii dcr1* vs *dcr1 rrp6*). (b) The effects of RNAi on global ASlncRNA levels at tandem genes in *S. cerevisiae* and *N. castellii*. Boxplots of the distribution of normalized read counts of ASlncRNAs at tandem oriented genes for wild type and RNAi+ *S. cerevisiae* (Top, n = 3656 genes) or wild type and *dcr1 N. castellii* (Bottom, n = 3064 genes). (c) The effects of RNAi on global ASlncRNA levels at convergent genes in *S. cerevisiae* and *N. castellii*. As in (b), except at convergent oriented genes, *S. cerevisiae* (n = 3046 genes), *N. castellii* (n = 2846 genes). See Figure 2 legend for description of boxplot features. (d) Growth defects of *N. castellii rrp6* mutant are partially recued by *dcr1* mutation. A spot test of 5-fold serial dilutions of *N. castellii dcr1,* wild type, *rrp6, dcr1 rrp6* strains on YEPD at an elevated temperature (*N. castellii* grows optimally at 25°C.)