frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 February 2024
DOI 10.3389/fmicb.2024.1362067

OPEN ACCESS

EDITED BY
Alexander DeLuna,
National Polytechnic Institute of Mexico (CINVESTAV), Mexico

REVIEWED BY
Sandra Coral Garrett,
UCONN Health, United States
Pengfei Ding,
University of Maryland, United States
Selene L. Fernandez-Valverde,
University of New South Wales, Australia

*CORRESPONDENCE
Uciel Chorostecki
✉ ucielp@gmail.com
Toni Gabaldón
✉ toni.gabaldon@bsc.es

RECEIVED 27 December 2023
ACCEPTED 07 February 2024
PUBLISHED 26 February 2024

CITATION
Chorostecki U, Saus E and Gabaldón T (2024) Probing RNA structural landscapes across *Candida* yeast genomes.
Front. Microbiol. 15:1362067.
doi: 10.3389/fmicb.2024.1362067

Probing RNA structural landscapes across *Candida* yeast genomes

Uciel Chorostecki[1,2,3]*, Ester Saus[2,3] and Toni Gabaldón[2,3,4,5]*

[1]Faculty of Medicine and Health Sciences, Universitat Internacional de Catalunya, Barcelona, Spain, [2]Barcelona Supercomputing Centre (BSC-CNS), Barcelona, Spain, [3]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain, [4]ICREA, Barcelona, Spain, [5]Centro de Investigación Biomédica En Red de Enfermedades Infecciosas (CIBERINFEC), Barcelona, Spain

Understanding the intricate roles of RNA molecules in virulence and host-pathogen interactions can provide valuable insights into combatting infections and improving human health. Although much progress has been achieved in understanding transcriptional regulation during host-pathogen interactions in diverse species, more is needed to know about the structure of pathogen RNAs. This is particularly true for fungal pathogens, including pathogenic yeasts of the *Candida* genus, which are the leading cause of hospital-acquired fungal infections. Our work addresses the gap between RNA structure and their biology by employing genome-wide structure probing to comprehensively explore the structural landscape of mRNAs and long non-coding RNAs (lncRNAs) in the four major *Candida* pathogens. Specifically focusing on mRNA, we observe a robust correlation between sequence conservation and structural characteristics in orthologous transcripts, significantly when sequence identity exceeds 50%, highlighting structural feature conservation among closely related species. We investigate the impact of single nucleotide polymorphisms (SNPs) on mRNA secondary structure. SNPs within 5′ untranslated regions (UTRs) tend to occur in less structured positions, suggesting structural constraints influencing transcript regulation. Furthermore, we compare the structural properties of coding regions and UTRs, noting that coding regions are generally more structured than UTRs, consistent with similar trends in other species. Additionally, we provide the first experimental characterization of lncRNA structures in *Candida species*. Most lncRNAs form independent subdomains, similar to human lncRNAs. Notably, we identify hairpin-like structures in lncRNAs, a feature known to be functionally significant. Comparing hairpin prevalence between lncRNAs and protein-coding genes, we find enrichment in lncRNAs across *Candida* species, humans, and *Arabidopsis thaliana*, suggesting a conserved role for these structures. In summary, our study offers valuable insights into the interplay between RNA sequence, structure, and function in *Candida* pathogens, with implications for gene expression regulation and potential therapeutic strategies against *Candida* infections.

KEYWORDS

secondary structure, lncRNA, *Candida*, fungi, evolution

Introduction

Yeasts belonging to the *Candida* genus are among the most common human fungal opportunistic pathogens (Brown et al., 2012). Up to 30 distinct, phylogenetically diverse *Candida* species have been reported to infect humans, and the risk of infection is especially high in immunocompromised hospitalized patients (Papon et al., 2013; Gabaldón et al., 2016). Although their incidence may differ from region to region, the four most common species, *Candida albicans*, *Candida glabrata*, *Candida parapsilosis* and *Candida tropicalis*, generally in this order, collectively account for over 95% of the cases (Diekema et al., 2012), all of them included in the fungal priority pathogens list, by the world health organization (Alastruey-Izquierdo, 2022).

The transcriptome comprises all the RNA transcribed by the genome in a specific cell type or tissue, at a particular developmental stage, and under a specific condition (Jacquier, 2009). Thus, transcriptome analysis allows for understanding the genome's activity and provides comprehension of gene expression regulation, structure and function. Accumulating data overwhelmingly supports the idea that any RNA's function is always determined by its structure (Vicens and Kieft, 2022). For instance, recent studies have shown that the regulation of mRNA stability through RNA modification is a crucial step for achieving a tight regulation of gene expression (Delaunay and Frye, 2019), and mRNA stability depends on the mRNA nucleotide sequence, which affects the secondary and tertiary structures of the mRNAs. Several methods are available for the computational prediction of RNA secondary structure, and the performance of the methods varies across the datasets (Bugnon et al., 2022).

Moreover, the accuracy of existing algorithms still needs to be improved to model long RNA molecules since the number of possible structures scales dramatically with the length of the sequence (Wan et al., 2012; Bugnon et al., 2022). The structure prediction accuracy is improved either by searching for conservation in a set of homologous sequences or by using experimental data. In this regard, several experimental methods have been developed to study RNA structure at single-nucleotide resolution (Kertesz et al., 2010; Underwood et al., 2010; Wan et al., 2013; Saus et al., 2018). In particular, nextPARS is an enzymatic-based technique that allows probing of the secondary structure of RNAs *in vitro* at a genome-wide scale (Saus et al., 2018). The nextPARS method has advantages compared to similar probing methods like Structure-seq (Ding et al., 2014) and SHAPE-seq (Lucks et al., 2011). Its experimental procedure is straightforward, the scripts necessary for obtaining secondary structure profiles and complete details of the methodology are freely available (Chorostecki et al., 2021), and the results demonstrate accuracy comparable to or exceeding previous methodologies (Saus et al., 2018).

The structural content of the UTRs compared to the coding regions has been found to vary from organism to organism (Mortimer et al., 2014). Secondary structures of 5' UTR, the coding region, and 3' UTR are mainly independent since base pairs across domain boundaries are infrequent (Shabalina et al., 2006). 5′ UTR structures in eukaryotic mRNAs can modulate the translation initiation. Moreover, regulatory elements in the mRNA, especially in the 3′ UTR, can also modulate translation (Leppek et al., 2018). It is important to note that even within 5' UTRs, unstructured (linear) regulatory elements are likely to have a crucial impact on translation (Hinnebusch et al., 2016). The median length of mRNA 5′ UTRs in humans is around 200 nt, exceeding those of other mammals and tripling that of budding yeast (about 50 nt) (Leppek et al., 2018).

Sequence-structure relationships in mRNA coding regions remain elusive, and their secondary structure is largely unknown. The correlation between sequence and structure in *Saccharomyces cerevisiae* mRNAs coding regions is weaker than in small noncoding RNAs (sncRNAs), and profiles of paralogous mRNAs show a strong correlation with sequence for identity levels upwards of 85–90% (Chursov et al., 2011). However, pairs of more distantly related yeast transcripts' secondary structures appear to be unrelated (Chursov et al., 2011). Furthermore, in a previous study, the structures of orthologous mRNAs from *Saccharomyces cerevisiae* and *C. glabrata* were studied only *in-silico*. That study found no correlation for pairs with low sequence identity levels, and there was no conclusion for similar sequences due to a lack of data (Chursov et al., 2011).

The regulation of human body temperature is a robust first line of defense against fungal infections, particularly in cases of fever (Bergman and Casadevall, 2010). High temperatures significantly limit fungal growth. During Candida infections, the host's fever response exposes fungal cells to temperatures between 37°C and 42°C. These temperature changes profoundly impact various aspects of *Candida*, such as its appearance, reproduction, characteristics, and drug resistance (Shapiro et al., 2011). This insight can guide strategies for combating *Candida* infections. Moreover, RNA secondary structures exhibit high sensitivity to temperature variations and modifying temperature conditions can induce alterations in RNA folding and stability. Exploring RNA structures across various temperatures unveils dynamic structural changes that may play a pivotal role in governing gene expression, especially in response to temperature shifts encountered during infection processes.

Long non-coding RNAs (lncRNAs) represent a heterogeneous group of ncRNAs, longer than 200 nucleotides (Ponting et al., 2009), and the function of most of them remains uncertain. These molecules exhibit distinct characteristics consistently observed across diverse taxa, including mammals, insects, and plants. Compared to protein-coding genes, they tend to have lower expression levels and greater cell type specificity (Cabili et al., 2011). Their sequence conservation is generally poor, and they undergo rapid evolutionary changes, frequently displaying species-specific traits (Kutter et al., 2012). Recent research sheds light on the significance of lncRNAs in *Candida* species. In a comprehensive study encompassing five major *Candida* pathogens, hundreds of lncRNAs were identified from publicly available sequencing data (Hovhannisyan and Gabaldón, 2021). These lncRNAs exhibit unique evolutionary characteristics, and despite limited sequence conservation among these species, some lncRNAs share common sequence motifs and show co-expression with specific protein-coding transcripts, suggesting potential functional connections.

Understanding RNA structure is vital as it can help uncover the regulation of mRNAs and the roles of lncRNAs and even reveal potential functional consequences of synonymous or non-coding genetic variations. Thus, to gain further insights into the role of RNA structure in *Candida* species, we performed RNA probing experiments and comparative genomics analysis of five yeast species, including the four major *Candida* species and the model yeast *S. cerevisiae*. To our knowledge, our study constitutes the most comprehensive examination of structures of mRNAs and lncRNAs in yeasts. We determined and compared pairs of orthologs across the considered

species to shed new light on the relationships between sequence and structure conservation. In addition, for one of the species, we compared the RNA structures at varying temperatures and measured their relative stabilities. Finally, we reached the nextPARS score across the coding sequence (CDS) and untranslated regions (UTRs) in *C. glabrata*. Our results show that the coding regions in *C. glabrata* transcripts are more structured than untranslated regions.

Our study provides a comprehensive analysis of mRNA and lncRNA structures in the major *Candida* pathogens, shedding light on the conservation of structural features and their correlation with sequence conservation. Candida species have developed resistance mechanisms to antifungal agents, making treatment increasingly challenging. Understanding the intricate relationship between sequence, structure, and function will contribute to understanding the evolution of *Candida* species, potentially identifying new strategies to combat *Candida* infections.

Methods

Sample preparation, secondary structure probing with nextPARS and sequencing

We performed four different experiments to probe the secondary structure of transcriptomes of four *Candida* species: *C. albicans* SC5314, *C. parapsilosis* GA1, *C. glabrata* CBS138 and *C. tropicalis* CSPO. Cultures were set up for each species and were grown in YPD medium in an orbital shaker at 200 rpm, overnight, at 37°C for *C.glabrata* and 30°C for the rest of the species. Total RNA was extracted from these cultures using the RiboPure™-Yeast Kit according to the manufacturer's instructions (Thermo Fisher Scientific), starting with a total amount of 3×10^8 cells per sample as recommended for a maximum yield. To obtain PolyA+ RNA samples, total RNA from yeast were purified by two rounds of selection using Dynabeads mRNA purification kit following the manufacturer's instructions (Thermo Fisher Scientific). The quality (RNA integrity) and quantity of both total and PolyA+ RNA samples were assessed using the Agilent 2,100 Bioanalyzer with the RNA 6000 Nano LabChip Kit (Agilent), the NanoDrop 1,000 Spectrophotometer (Thermo Fisher Scientific), and the Qubit Fluorometer with the Qubit RNA BR (Broad-Range) Assay Kit (Thermo Fisher Scientific).

We used the nextPARs approach (Saus et al., 2018), to probe the secondary structure of the transcriptomes at 23°C, 37°C and 55°C for *C. parapsilosis* and at 23°C for the remaining species. Two μg of the corresponding polyA+ RNA were used per each reaction tube. 0.03 U of RNase V1 (Ambion) and 200 U of S1 nuclease (Fermentas) were used to digest the corresponding samples when probing the structure at 23°C. When samples were probed at 37°C and 55°C, the final enzyme concentration was adapted to prevent overdigestion of RNAs due to higher temperatures, according to previous studies (Wan et al., 2012). Thus, in the digestion reactions, 0.015 U or 0.0075 U of Rnase V1 (Ambion), and 100 U or 50 U of S1 nuclease (Fermentas) were used to probe the transcriptomes at 37°C or 55°C, respectively. We performed quality controls of the final digested samples to confirm that they were not over-digested after enzymatic treatment. Before proceeding with the library preparation of the digested samples, we always inspected them visually to confirm they were not over-digested, as we can see from the images of the Agilent 2,100 Bioanalyzer profiles of RNA samples used in the experiments before and after the enzymatic digestions of our experiments (Supplementary Figure S10).

Once the good quality of the final digested samples was confirmed, libraries were prepared using the TruSeq Small RNA Sample Preparation Kit (Illumina) following a modified protocol previously described (Saus et al., 2018; Chorostecki et al., 2021). After performing quality control of each library using Agilent 2,100 Bioanalyzer with the DNA 1000 kit (Agilent), libraries were sequenced in single-reads with read lengths of 50 nucleotides in Illumina HiSeq2500 sequencers at the Genomics Unit of the CRG (CRG-CNAG).

For *S. cerevisiae*, we used available data previously generated by the nextPARS technique (Bioproject id PRJNA380612) (Saus et al., 2018).

RNA secondary structure prediction and visualization

First, we converted the nextPARS score to SHAPE-like reactivities with the nextPARS2SHAPE v1.0 script,[1] and we used it for the secondary structure prediction. Then, to obtain the secondary structure of *Candida* mRNAs and lncRNAs, we use the Fold software (Version 6.2) from the RNAstructure package (Reuter and Mathews, 2010), using pseudo energy restraints. Residues for which no nextPARS data were assigned a reactivity of 999, as suggested by the Fold manual. Additionally, we calculated the similarity score for lncRNAs folded with and without nextPARS data using the R package RNAsmc (Wang et al., 2023). This package provides a similarity score for two RNA secondary structures, with larger values indicating greater similarity between the structures. The score ranges from 0, representing no similarity, to a maximum value of 10, signifying identical RNA secondary structures. To determine significance, we applied a threshold of 6, where higher values indicate greater structural similarity, and scores below this threshold are considered not significantly similar, based on the RNAsmc publication (Wang et al., 2023).

To obtain the secondary structure of the lncRNA and PCG, we use the RNAstructure software (Reuter and Mathews, 2010). Using the nextPARS2SHAPE.py script,[2] we converted the nextPARS score to SHAPE-like normalized reactivities that were used to provide pseudoenergy restraints to the Fold software. For residues for which there was no nextPARS data, we assigned a reactivity of 999, as suggested by the RNAfold manual. The RNA arc diagram was built using R-chie (version 2.0.) (Tsybulskyi et al., 2020). RNA structures were constructed using VARNA (Version 3–93) (Darty et al., 2009).

Stem-loop prediction

We retrieve the sequence information of the four *Candida* species, human and *Arabidopsis thaliana*. Then, to obtain the secondary structure of protein-coding genes (PCG) and lncRNAs, we use the

1 https://github.com/Gabaldonlab/MutiFolds/blob/master/scripts/nextPARS2SHAPE.py

2 https://github.com/Gabaldonlab/MutiFolds/

Fold software (Version 6.2) from the RNAstructure package (Reuter and Mathews, 2010). Finally, the stem-loop prediction was made by developing an in-house script (stem-loop_predictor_public.py) using forgi (version 2.0.0), a Python library for manipulating RNA secondary structure. The folding and the search for stem-loop in different species have been performed on MareNostrum 4 supercomputer at the Barcelona Supercomputing Center, Spain. The code is available on GitHub.[3]

Sequence-structure correlation

We retrieved pairs of orthologous genes between the five species using the MetaPhOrs (v2.5) web server (Chorostecki et al., 2020), which provides orthology and paralogy relationships derived from gene family trees from different source repositories across ~4,000 different fully-sequenced species. We developed a custom Python (version 3.10.9) script (get_positions_alignments.py) to perform the analysis, and we added it to the GitHub repository (see Footnote 3). Briefly, for each pair of orthologous mRNAs that we have good nextPARS information (Supplementary Figure S1), we performed a multiple sequence alignment (MSA) using T-Coffee (version 13.46.0) (Di Tommaso et al., 2011) with the method flag equal to slow_pair. Then, we measure the Spearman correlation for each pair of orthologous using the nextPARS score for the aligned position.

We performed a shuffled analysis as a control to elucidate further the relationship between sequence conservation and structural features in orthologous transcripts across yeast species. For that, we randomly shuffled the positions of nextPARS scores and their alignments for orthologous pairs of genes retrieved from different species. We then calculated the correlation of the nextPARS scores for these shuffled positions in the same way for the real data. This approach allowed us to assess whether the observed sequence-structure correlations in orthologous transcripts were significantly higher than those expected by random chance.

Statistical methods

To compare nextPARS data among *C. parapsilosis* mRNAs at various temperatures, the non-parametric Kruskal-Wallis test was employed, considering the non-normal distribution of the data. Similarly, the assessment of nextPARS scores between UTRs and coding regions in *C. glabrata* involved the application of the Kruskal-Wallis test, accommodating the non-normality of the dataset. Furthermore, in the analysis focusing on SNP positions within the 5' UTR in *C. glabrata*, a paired T-test was utilized for the comparison, considering the paired nature of the data.

SNP analysis

To investigate the impact of secondary structure induced by SNPs in *Candida* transcripts, we used variant calling data obtained from CandidaMine,[4] an integrative data warehouse for *Candida*. We focused on identifying SNPs within the coding regions of genes for all four *Candida* species studied. We compared the nextPARS scores in loci with identified SNPs to those in loci with no reported SNPs. Additionally, for *C. glabrata*, we extended our analysis to the untranslated regions (UTRs), as these regions remain unannotated in other *Candida* species. Specifically, we assessed how SNPs within the 5' UTRs of *C. glabrata* transcripts correlate with RNA secondary structure, focusing on identifying any potential structural changes associated with these SNPs. We developed custom Python (version 3.10.9) scripts (compare_score_SNP.py) using the pandas library (version 2.1.1) to process and compare the SNPs and nextPARS scores. The code is available on GitHub (see Footnote 3). Specifically, we assessed how SNPs within the 5' UTRs of *C. glabrata* transcripts correlate with RNA secondary structure, focusing on identifying any potential structural changes associated with these SNPs. Here, a threshold of 0.2 was applied to enhance precision and minimize noise. This threshold was explicitly implemented to reduce potential interference from positions with a score of 0, which could arise due to the absence of nextPARS data at those particular positions.

Calculation of long-range interactions

The percentage of long-range interactions in the secondary structure of each lncRNA sequence was determined using a custom Python (version 3.10.9) script (get_long_range_interactions.py) to perform the analysis, and we added it to the GitHub repository (see Footnote 3). The script parses the dot-bracket notation, representing RNA secondary structure, and identifies base pairs. We use the Fold software (Version 6.2) from the RNAstructure package (Reuter and Mathews, 2010) to obtain the secondary structures with and without using the nextPARS experiments as constraints. A dynamically calculated threshold based on a specified percentage of the sequence length is employed to define long-range interactions. We used a default threshold of 25%, indicating that interactions with a base pair distance exceeding 25% of the sequence length are considered long-range. Hence, we compute the percentage of long-range interactions and generate a plot illustrating the variation across different lncRNAs.

Data availability statement

The raw sequencing data of nextPARS experiments used in this project have been deposited in the Short Read Archive of the European Nucleotide Archive under the Bioproject IDs PRJNA714002 and PRJNA838569. To enhance accessibility to our identified RNA structures, we have made the nextPARS score information available in CSV format for each mRNA from the four *Candida* species and *Saccharomyces cerevisiae*. This dataset can be accessed on our GitHub repository (see Footnote 3) inside the nextPARS_score directory. For details on the protocol for the computation of the nextPARS scores, see the book chapter publication (Chorostecki et al., 2021).

3 https://github.com/Gabaldonlab/Candida_nextPARS

4 https://candidamine.org/

Results

Genome-wide view of the structural landscape of *Candida* mRNAs

To compare the structural landscape of *Candida* RNAs, we performed nextPARS experiments (Saus et al., 2018; Chorostecki et al., 2021) in the four most commonly infecting *Candida* species: *C. albicans*, *C. glabrata*, *C. parapsilosis,* and *C. tropicalis*, which collectively account for over 95% of all Candida infections (Consortium OPATHY and Gabaldón, 2019). The nextPARS protocol, detailed in the original publication, uses Illumina sequencing technology for high-throughput and multiplexed probing of RNA secondary structures. This methodology provides an effective means to evaluate the secondary structure of RNAs experimentally. The resulting score file encapsulates a structural profile of the RNA transcript, assigning a score to each nucleotide represented by the structural profile that ranges from −1 (highest preference for single-stranded) to 1 (highest preference for double-stranded). After filtering out transcripts with low counts we obtained structural information on mRNAs and lncRNAs (see Methods, Supplementary Table S1). This dataset, comprising experimental probing information from hundreds of transcripts, provides the first genome-wide view of the structural landscape of *Candida* mRNAs and lncRNAs. For comparative purposes, we also used, in downstream analyses, publicly available nextPARS data for the model yeast *Saccharomyces cerevisiae* (Saus et al., 2018).

Assessing sequence-structure relationships in *Candida* mRNAs

There needs to be a better understanding of how sequence conservation relates to the conservation of structural features in orthologous transcripts across different species. We measured the sequence and structure similarity correlation in mRNAs coding regions to assess this for yeast species. We retrieved pairs of orthologous genes between the five species using the MetaPhOrs (v2.5) web server (Chorostecki et al., 2020). For all pairs of orthologous mRNAs with nextPARS information (Supplementary Figure S1), we calculated the correlation of the nextPARS score in aligned positions. We took different sequence identity intervals and observed high correlations of nextPARS scores (median Pearson correlation >0.4) when the sequence percent identity is above ~50% (Figure 1A). This value is significantly higher than expected by chance, as correlations between randomly shuffled scores were consistently below 0.2 (Supplementary Figure S2). We compare orthologous transcripts found using MSA methods (see Methods) so they are conserved at the sequence level. 50% is a very low identity for nucleotides, and here, we are interested in looking for a low boundary from which this relationship is lost. Levels of structural conservation across orthologous pairs reflected phylogeny (Gabaldón et al., 2016), where CTG *Candida* species (*C. albicans, C. parapsilosis, C. tropicalis*) and post-WGD species (*C. glabrata* and *S. cerevisiae*) belong to two distant clades (Figure 1B).

Moreover, we compared the correlations of the nextPARS score in aligned positions but using multiple sequence alignments from orthologs in the five species, using each species as a seed (Supplementary Figure S1). On average, we observed a positive correlation between the nextPARS score in aligned positions (Supplementary Figure S3).

Investigating alterations in transcript secondary structure induced by temperature changes is crucial, as it helps elucidate how Candida pathogens dynamically regulate their gene expression in response to temperature shifts encountered during infection processes. Thus, to investigate the alterations in transcript secondary structure prompted by temperature changes in *C. parapsilosis* mRNAs, we performed nextPARS experiments at three different temperatures: 23°C, 37°C and 55°C. We noticed differences on average in the nextPARS score at different temperatures (Supplementary Figure S4A), where the significant differences are between 23°C compared with 37°C and 55°C. We also measured the temperature stability at different temperatures but found no significant difference when comparing conserved positions against non-conserved ones (Supplementary Figure S4B).

Exploring structural dynamics in *Candida glabrata* mRNA regions in coding sequences vs. UTRs and the impact of SNPs

Exploring the structural characteristics of coding and untranslated regions in *C. glabrata* mRNAs is pivotal for understanding the intricacies of gene expression regulation. Thus, we investigate the structure in different regions within mRNAs. We compared the average nextPARS score across the coding sequence (CDS) and untranslated regions (UTRs) in *C. glabrata*. CDS exhibit significantly higher nextPARS scores than 5′ and 3’ UTRs (Figure 2A; p-value <1.7 e-06 and p-value <0.00053, respectively). These findings agree with the results observed in other species, such as *Saccharomyces cerevisiae* and *Arabidopsis thaliana* (Kertesz et al., 2010; Li et al., 2012). However, opposite results have been observed in humans, *Drosophila melanogaster* and *Caenorhabditis elegans,* in which CDS are slightly more single-stranded than the UTRs (Li et al., 2012; Wan et al., 2014).

We then delved into the impact of secondary structure on sequence variation, analyzing SNPs in UTRs from *C. glabrata*. We used variant calling data obtained from CandidaMine (see Footnote 4) – an integrative data warehouse for *Candida* yeasts. We compared nextPARS scores in loci with SNPs to those in loci with no reported SNPs. Our results show no differences between those two groups for any of the four *Candida* species (Supplementary Figure S5). Then, we performed a similar analysis using UTRs from *C. glabrata* – as these features remain unannotated in the other *Candida species*. Changes in those regions are associated with deregulation in gene expression at transcriptional and post-transcriptional levels (Lawless et al., 2009). When we compared SNP’s positions in UTRs, we observed that SNPs in 5’ UTR tended to localize to less structured positions (Figure 2B; p-value <0.02976), shedding light on potential regulation in these critical regions. However, in the 3’ UTR, a difference in means was noted, albeit with limited statistical significance (p-value = 0.2491), due to the small number of values in SNP positions. This observation underscores the dynamic interplay between sequence variations and structural features, adding depth to our understanding of the intricate regulatory landscape within *Candida* yeasts.

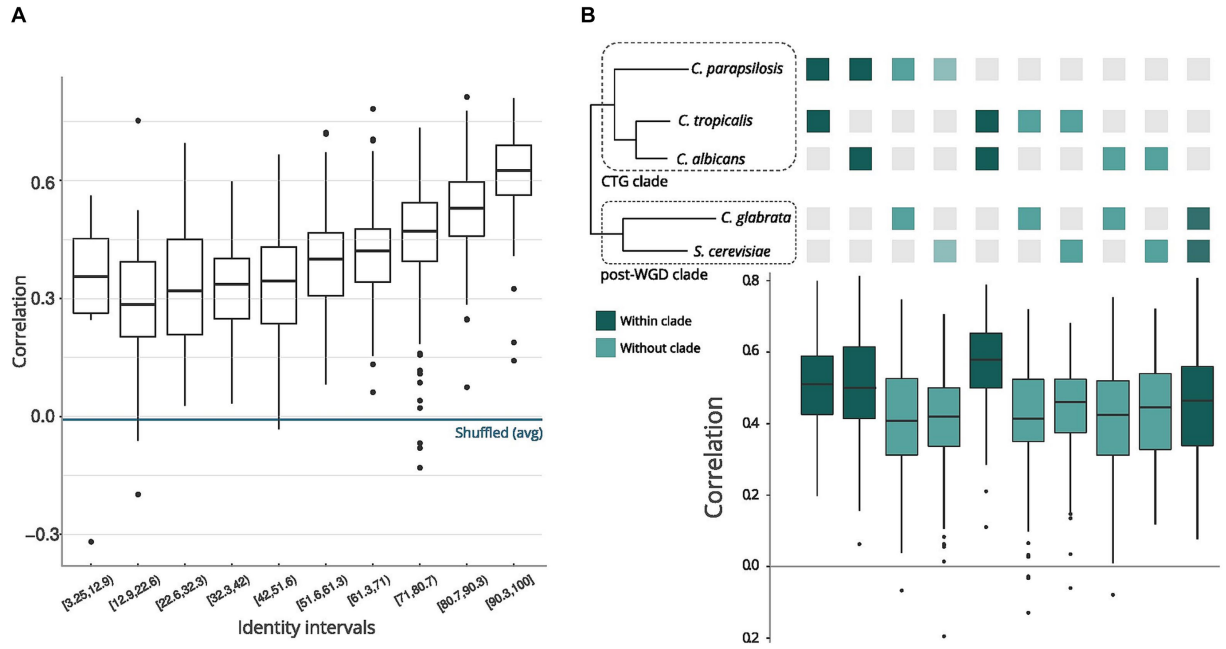

FIGURE 1
Comparison of sequences and structures across yeast species. **(A)** Pair sequence alignments by intervals. The X-axis represents sequence identity intervals as a percentage, while the Y-axis illustrates the correlations of nextPARS scores across various identity intervals. **(B)** Correlation between sequence and structure in orthologs. The orthologs represent pairs of genes from the species depicted in the tree above the graph. The tree's topology and evolutionary distances are adapted from Gabaldón et al. (2016).

Structural characterization of *Candida* lncRNAs using experimental information

Catalogues of long non-coding RNAs (lncRNAs) in major *Candida* species have only been recently characterized (Hovhannisyan and Gabaldón, 2021). We used our nextPARS data to characterize the secondary structure of inferred lncRNAs in the four *Candida* species considered in this study. Given the generally low expression levels of lncRNAs, we could only detect a few lncRNAs with sufficient confidence (> five average counts per position) (Supplementary Table S1). These nevertheless provide the first structural insights of lncRNAS in these relevant pathogens. Furthermore, it is crucial to highlight that our characterization of lncRNAs involved a unique approach. Leveraging the experimental insights provided by our nextPARS data, we employed these data as constraints during the RNA folding process of lncRNAs. This approach allowed us to guide *in-silico* secondary structure predictions with the experimental data. We noticed significant differences (threshold below 6) using similarity scores of RNAsmc (Wang et al., 2023) for certain lncRNAs compared to predictions made without these constraints (see Methods, Supplementary Figure S6). Moreover, we observed that most lncRNAs folded in potential independent subdomains with fewer long-range interactions compared to our dataset's average percentage (22.6%) of long-range interactions (see Methods, Supplementary Figure S7). This is in accordance with previous studies on human lncRNAs (Somarowthu et al., 2015; Chorostecki et al., 2021; Ziv et al., 2021).

Next, we focused on a specific lncRNA known as MSTRG.4167.1, which displayed upregulation in *C. albicans* during epithelial cell infection, as reported by Hovhannisyan and Gabaldón (2021). To investigate its structural characteristics, we folded the lncRNA using data from nextPARS experiments as constraints (Figure 3A). We examined this particular lncRNA in detail based on the experimental data we obtained through our nextPARS experiments. We compared predicted RNA structures using data from nextPARS as constraints for experiments conducted at different temperatures for this particular lncRNA (Figure 3B). By utilizing similarity scores of the structures, we observed that significant differences emerged when comparing MSTRG.4167.1 at 23 and 37 degrees, with an even more substantial distinction between 37 and 55 degrees. The most pronounced difference was when comparing the lncRNA at 23 and 55 degrees (Supplementary Table S2). As expected, the similarity score decreases as the temperature difference increases. These results imply that the lncRNA folds differently at distinct temperatures, which could result in varied functions or regulations. Furthermore, we noticed a slight trend of fewer long-range interactions for MSTRG.4167.1 as the temperature increased (Supplementary Figure S8).

Comparative analysis of hairpin-like structures reveals enrichment in lncRNAs across *Candida species*

To further explore the structural landscape of lncRNAs in *Candida* yeast pathogens, we delved into the presence of stem-loop structures essential for the functionality of characterized lncRNAs in diverse species. We noticed that some lncRNAs structures, as determined using nextPARS, presented stem-loop structures. These hairpin-like structures are essential for function in characterized lncRNAs in other species, such as Xist (Maenner et al., 2010), NORAD (Chorostecki et al., 2021; Ziv et al., 2021), LINC00152 (Reon et al., 2018),

10.3389/fmicb.2024.1362067

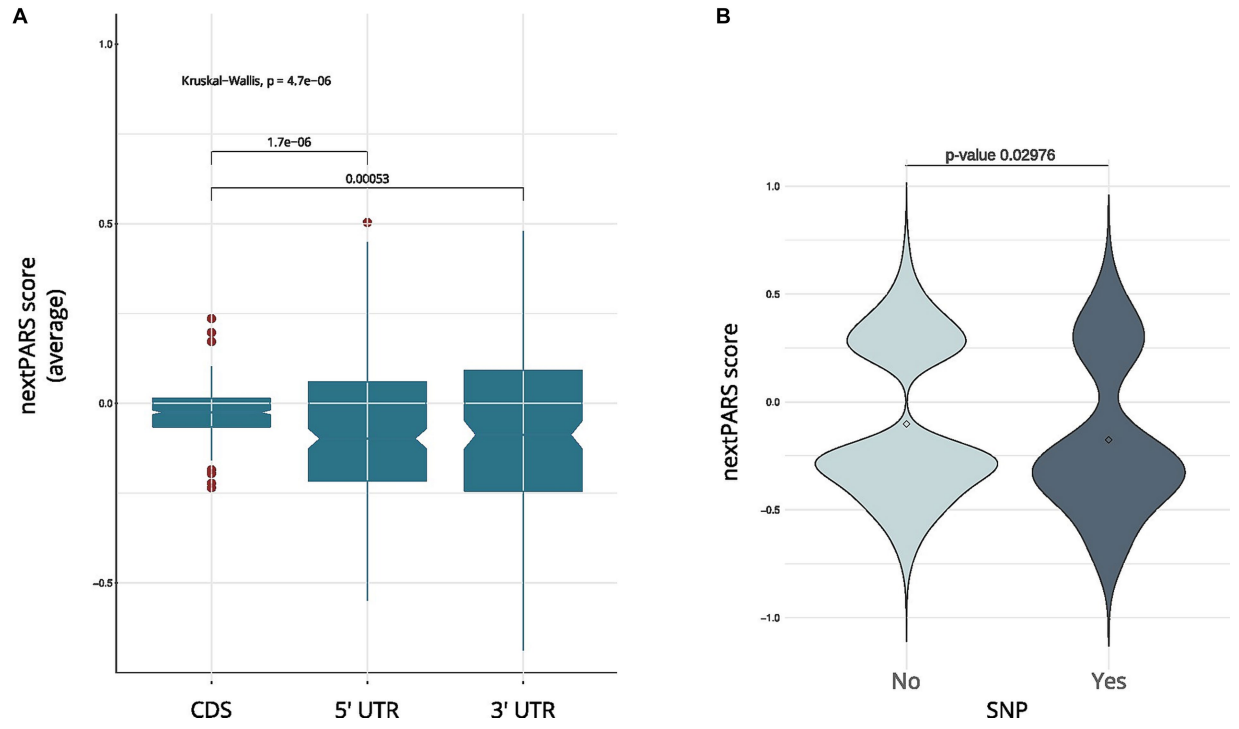

FIGURE 2
RNA structure comparative analysis in *C. glabrata*. **(A)** A boxplot showing a comparison between coding regions in *C. glabrata* and UTRs. **(B)** The nextPARS score in SNP vs. no SNP positions in *C glabrata* 's 5' UTR.

TCONS_00202587 (Song et al., 2020), roX1 and roX2 (Ilik et al., 2013) among others. We compared the proportion of hairpins in computationally predicted structures of lncRNAs and protein-coding genes in the four *Candida* species (Figure 4A). For this, we scanned for hairpins similar to those previously described on lncRNAs. We observed that the number of hairpin structures normalized by the number of molecules and sequence length was higher in lncRNAs than in protein-coding genes in the four *Candida* species (p-value <0.0016; Figure 4B). We performed the same approach on humans and *Arabidopsis thaliana* (Figure 4A), and we observed similar results in lncRNAs and protein-coding genes of hairpin structures (Figure 4B). These results suggest that hairpin enrichment in lncRNAs may be a conserved feature of eukaryotes.

Discussion

In this study, we performed nextPARS experiments to investigate the structural landscape of mRNAs and lncRNAs in the four major *Candida* pathogens. Our findings provide valuable insights into the relationship between sequence conservation and structural features and the impact of secondary structure on sequence variation. One of the key findings of our study is the correlation between sequence and structure similarity in orthologous mRNAs coding regions. We observed high correlations of nextPARS scores when the sequence identity was above 50%, indicating significant conservation of structural features in highly similar sequences. This correlation was consistent with the phylogenetic relationship of the species, with CTG *Candida* species and post-WGD species belonging to distinct clades. There's not always a direct relation between sequence and structure for some RNAs. Compensatory mutations in MSA may occur, wherein a substitution on one side of a base pair is compensated by a substitution on the other side, which is expected to restore base pairing. This may result in RNA that is not well conserved in the primary sequence but is conserved in the secondary structure. In the case of certain tRNA molecules, they can maintain their structural integrity and function despite variations in their primary sequences across different species. Here, we also characterized some lncRNAs, and it was shown they could form complex secondary structures, which tend to be more conserved than primary sequences and are thought to mediate their molecular functions (Novikova et al., 2013). These results highlight the importance of considering both sequence and structure conservation in understanding the functional implications of RNA molecules.

Furthermore, we explored the impact of SNPs on RNA secondary structure. Surprisingly, we found no significant differences between the structural propensities of loci with SNPs and those without SNPs in the coding regions of *Candida* species, contrary to what has been observed for humans (Pegueroles and Gabaldón, 2016). In addition, we performed a more detailed analysis of SNPs within coding regions. Specifically, we separated SNPs into those causing amino acid changes (missense variants) and those that are silent (synonymous variants). After comparing these two subsets, we did not observe any statistically significant differences to see the impact on nextPARS scores (Supplementary Figure S9). We also compared these subsets to positions without SNPs for all four *Candida* species we analyzed. Our results still indicate no significant differences in the structural propensities between loci with and without SNPs in the coding regions of Candida species. However, in the UTRs of *C. glabrata*, SNPs tended

10.3389/fmicb.2024.1362067

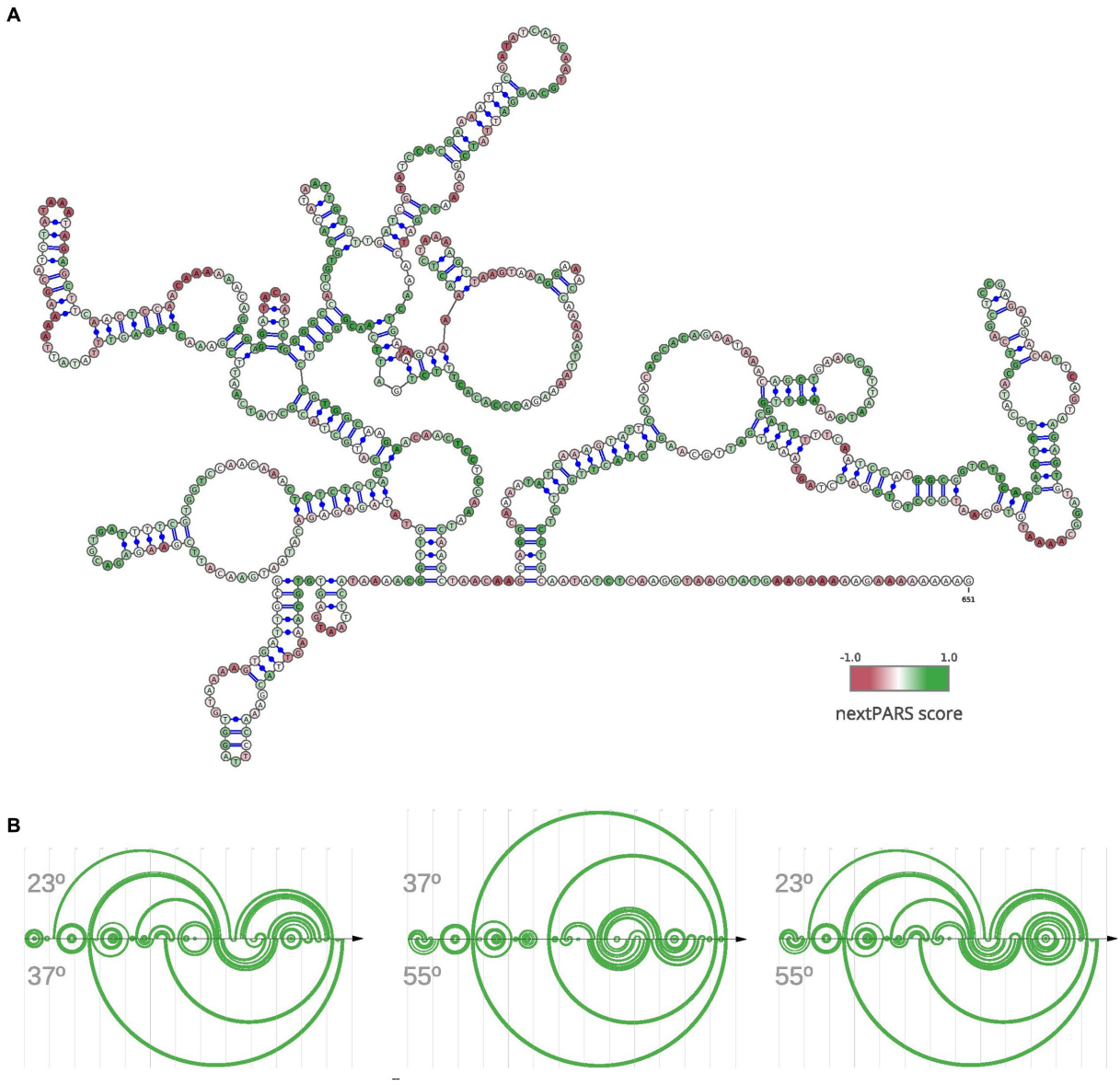

FIGURE 3
Structural Dynamics of lncRNA MSTRG.4167.1 This figure illustrates the structure of a specific lncRNA, MSTRG.4167.1. **(A)** RNA secondary structure of MSTRG.4167.1, where the folding was performed using nextPARS experiments as constraints. The color scale indicates the nextPARS score, ranging from red (indicating single-stranded regions) to green (indicating double-stranded regions). **(B)** Three different arc diagrams (R-Chie) are presented. On the left is a comparison of the same MSTRG.4167.1 lncRNA, using nextPARS experiments at 23 degrees (above) and 37 degrees (below). In the middle, the folding with nextPARS experiments at 37 degrees above and 55 degrees below. The last one shows the folding with nextPARS at 23 degrees above and 55 degrees below.

to be in less structured positions within the 5' UTR. This suggests that structural constraints play a role in maintaining the stability and functionality of UTRs, potentially affecting gene expression regulation. These discrepancies in humans and *Candida* RNAs may be attributed to species-specific variations in RNAs and by the use of high-resolution structural data provided by nextPARS experiments in our analysis, allowing us to uncover subtle structural constraints on sequence variation that might not have been discernible.

Comparing the structural characteristics of coding regions and UTRs, we observed that coding regions in *C. glabrata* exhibited significantly higher nextPARS scores than the 5′ and 3' UTRs. This finding is consistent with previous studies in other species, such as *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. However, it contradicts observations in humans, *Drosophila melanogaster*, and *Caenorhabditis elegans*, where UTRs were reported to be more structured than coding regions. These differences may reflect species-specific variations in RNA structure–function relationships and emphasize the need for further investigations in diverse organisms. It would be worthwhile to explore whether metazoans possess a higher abundance of RNA-binding proteins that interact with CDS, potentially influencing their structural characteristics. Examining this phenomenon using comprehensive annotations could provide valuable insights into the evolution of RNA structures and their regulatory roles in different organisms.

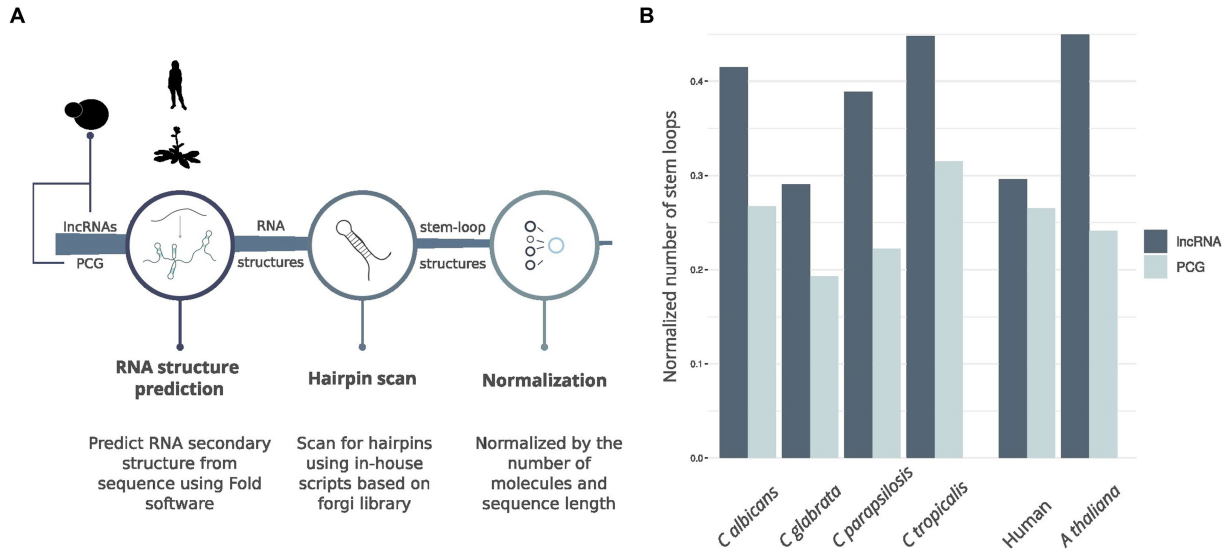

FIGURE 4
Hairpin comparison between lncRNAs and protein-coding genes. **(A)** Workflow of the pipeline to obtain hairpin structures. **(B)** Barplot showing the comparison between hairpins found in lncRNAs and PCG for the four *Candida* species, *A. thaliana* and Human.

Moreover, our study provides the first experimental characterization of lncRNA structures in the four major *Candida* species. The experimental data revealed the intricate secondary structures of lncRNAs, highlighting their potential independent subdomains and minimal long-range interactions. Using nextPARS data as constraints during RNA folding allowed us to uncover significant differences in certain lncRNAs compared to predictions made without these experimental constraints. Building on this foundational characterization, we focused on the specific lncRNA, MSTRG.4167.1, revealing temperature-dependent folding variations. The observed differences in structural conformations at different temperatures shed light on this particular lncRNA. These findings contribute to our understanding of *Candida* lncRNAs and underscore the importance of incorporating experimental data to enhance the accuracy of *in-silico* predictions in studying RNA structural landscapes. Interestingly, we identified hairpin-like structures in lncRNAs, which are functionally important in other species. Comparing the proportion of hairpin structures between lncRNAs and protein-coding genes, we found that lncRNAs were enriched in hairpin structures in *Candida* species, as well as in humans and *Arabidopsis thaliana*. One possible explanation for this phenomenon could be that these hairpin structures in lncRNAs contribute to their stability and function as structural scaffolds, aiding in interactions with other molecules or proteins within the cell. These structural features might play crucial roles in processes such as RNA-protein interactions, localization, or regulation of gene expression, highlighting the importance of further investigation into the functional implications of these hairpin-like structures in lncRNAs.

Our study comprehensively analyzes the structural landscape of *Candida* mRNAs and lncRNAs, revealing important insights into the relationship between sequence, structure, and function. The observed correlations between sequence and structure conservation and the differences in structure between coding regions and UTRs highlight the intricate interplay between RNA sequence and structure in gene expression regulation. The observed enrichment in hairpin structures of lncRNAs suggests their potential functional significance in diverse organisms. Fever is a potent defense against fungal infections, impacting *Candida* by exposing it to temperatures between 37°C and 42°C, influencing its characteristics and drug resistance. We think temperature-induced alterations in RNA secondary structures during infection may affect gene expression, offering insights for combating Candida infections. Further investigations into RNA structure and evolution will deepen our understanding of the intricate gene expression mechanisms and provide possible therapeutic strategies against *Candida* infections.

Data availability statement

The datasets presented in this study are deposited in the NCBI database under accession numbers PRJNA714002 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA714002) and PRJNA838569 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA838569).

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

UC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software,

Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ES: Methodology, Writing – review & editing. TG: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. TG group acknowledges support from the Spanish Ministry of Science and Innovation for grant PGC2018-099921-B-I00, cofounded by the European Regional Development Fund (ERDF); from the Catalan Research Agency (AGAUR) SGR423; from the European Union's Horizon 2020 research and innovation programme (ERC-2016-724173); from the Gordon and Betty Moore Foundation (Grant GBMF9742); from the "La Caixa" foundation (Grant LCF/PR/HR21/00737), and the Instituto de Salud Carlos III (IMPACT Grant IMP/00019 and CIBERINFEC CB21/13/00061 – ISCIII-SGEFI/ERDF). UC was partly funded through MICINN (IJC2019-039402-I) and MICINN (RYC2021-032641-I).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1362067/full#supplementary-material

References

Alastruey-Izquierdo, A. (2022). *WHO fungal priority pathogens list to guide research, development and public health action* Organización Mundial de la Salud (OMS).

Bergman, A., and Casadevall, A. (2010). Mammalian endothermy optimally restricts fungi and metabolic costs. *MBio* 1:e00212-10. doi: 10.1128/mBio.00212-10

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4:165rv13. doi: 10.1126/scitranslmed.3004404

Bugnon, L. A., Edera, A. A., Prochetto, S., Gerard, M., Raad, J., Fenoy, E., et al. (2022). Secondary structure prediction of long noncoding RNA: review and experimental comparison of existing approaches. *Brief. Bioinform.* 23:bbac205. doi: 10.1093/bib/bbac205

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Chorostecki, U., Molina, M., Pryszcz, L. P., and Gabaldón, T. (2020). MetaPhOrs 2.0: integrative, phylogeny-based inference of orthology and paralogy across the tree of life. *Nucleic Acids Res.* 48, W553–W557. doi: 10.1093/nar/gkaa282

Chorostecki, U., Saus, E., and Gabaldón, T. (2021). Structural characterization of NORAD reveals a stabilizing role of spacers and two new repeat units. *Comput. Struct. Biotechnol. J.* 19, 3245–3254. doi: 10.1016/j.csbj.2021.05.045

Chorostecki, U., Willis, J. R., Saus, E., and Gabaldon, T. (2021). Profiling of RNA structure at single-nucleotide resolution using nextPARS. *Methods Mol. Biol.* 2284, 51–62. doi: 10.1007/978-1-0716-1307-8_4

Chursov, A., Walter, M. C., Schmidt, T., Mironov, A., Shneider, A., and Frishman, D. (2011). Sequence–structure relationships in yeast mRNAs. *Nucleic Acids Res.* 40, 956–962. doi: 10.1093/nar/gkr790

Consortium OPATHY, Gabaldón, T. (2019). Recent trends in molecular diagnostics of yeast infections: from PCR to NGS. *FEMS Microbiol. Rev.* 43, 517–547. doi: 10.1093/femsre/fuz015

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974–1975. doi: 10.1093/bioinformatics/btp250

Delaunay, S., and Frye, M. (2019). RNA modifications regulating cell fate in cancer. *Nat. Cell Biol.* 21, 552–559. doi: 10.1038/s41556-019-0319-0

Diekema, D., Arbefeville, S., Boyken, L., Kroeger, J., and Pfaller, M. (2012). The changing epidemiology of healthcare-associated candidemia over three decades. *Diagn. Microbiol. Infect. Dis.* 73, 45–48. doi: 10.1016/j.diagmicrobio.2012.02.001

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., et al. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39, W13–W17.

Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., and Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700. doi: 10.1038/nature12756

Gabaldón, T., Naranjo-Ortíz, M. A., and Marcet-Houben, M. (2016). Evolutionary genomics of yeast pathogens in the Saccharomycotina. *FEMS Yeast Res.* 16:fow064. doi: 10.1093/femsyr/fow064

Hinnebusch, A. G., Ivanov, I. P., and Sonenberg, N. (2016). Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416. doi: 10.1126/science.aad9868

Hovhannisyan, H., and Gabaldón, T. (2021). The long non-coding RNA landscape of Candida yeast pathogens. *Nat. Commun.* 12:7317. doi: 10.1038/s41467-021-27635-4

Ilik, I. A., Quinn, J. J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., et al. (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in drosophila. *Mol. Cell* 51, 156–173. doi: 10.1016/j.molcel.2013.07.001

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* 10, 833–844. doi: 10.1038/nrg2683

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., et al. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107. doi: 10.1038/nature09322

Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., et al. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841. doi: 10.1371/journal.pgen.1002841

Lawless, C., Pearson, R. D., Selley, J. N., Smirnova, J. B., Grant, C. M., Ashe, M. P., et al. (2009). Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics* 10:7. doi: 10.1186/1471-2164-10-7

Leppek, K., Das, R., and Barna, M. (2018). Author correction: functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* 19:673. doi: 10.1038/s41580-018-0055-5

Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., et al. (2012). Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* 1, 69–82. doi: 10.1016/j.celrep.2011.10.002

Li, F., Zheng, Q., Vandivier, L. E., Willmann, M. R., Chen, Y., and Gregory, B. D. (2012). Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. *Plant Cell* 24, 4346–4359. doi: 10.1105/tpc.112.104232

Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., et al. (2011). Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* 108, 11063–11068. doi: 10.1073/pnas.1106501108

Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., et al. (2010). 2-D structure of the a region of Xist RNA and its implication for PRC2 association. *PLoS Biol.* 8:e1000276. doi: 10.1371/journal.pbio.1000276

Mortimer, S. A., Kidwell, M. A., and Doudna, J. A. (2014). Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* 15, 469–479. doi: 10.1038/nrg3681

Novikova, I. V., Hennelly, S. P., and Sanbonmatsu, K. Y. (2013). Tackling structures of long noncoding RNAs. *Int. J. Mol. Sci.* 14, 23672–23684. doi: 10.3390/ijms141223672

Papon, N., Courdavault, V., Clastre, M., and Bennett, R. J. (2013). Emerging and emerged pathogenic Candida species: beyond the *Candida albicans* paradigm. *PLoS Pathog.* 9:e1003550. doi: 10.1371/journal.ppat.1003550

Pegueroles, C., and Gabaldón, T. (2016). Secondary structure impacts patterns of selection in human lncRNAs. *BMC Biol.* 14:60. doi: 10.1186/s12915-016-0283-0

Ponting, C. P., Oliver, P. L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136, 629–641. doi: 10.1016/j.cell.2009.02.006

Reon, B. J., Karia, B. T. R., Kiran, M., and Dutta, A. (2018). LINC00152 promotes invasion through a 3′-hairpin structure and associates with prognosis in glioblastoma. *Mol. Cancer Res.* 16, 1470–1482. doi: 10.1158/1541-7786.MCR-18-0322

Reuter, J. S., and Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11:129. doi: 10.1186/1471-2105-11-129

Saus, E., Willis, J. R., Pryszcz, L. P., Hafez, A., Llorens, C., Himmelbauer, H., et al. (2018). nextPARS: parallel probing of RNA structures in Illumina. *RNA* 24, 609–619. doi: 10.1261/rna.063073.117

Shabalina, S. A., Ogurtsov, A. Y., and Spiridonov, N. A. (2006). A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 34, 2428–2437. doi: 10.1093/nar/gkl287

Shapiro, R. S., Robbins, N., and Cowen, L. E. (2011). Regulatory circuitry governing fungal development, drug resistance, and disease. *Microbiol Mol Biol Rev* 75, 213–267. doi: 10.1128/MMBR.00045-10

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., and Pyle, A. M. (2015). HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58, 353–361. doi: 10.1016/j.molcel.2015.03.006

Song, Y., Chen, P., Liu, P., Bu, C., and Zhang, D. (2020). High-temperature-responsive poplar lncRNAs modulate target gene expression via RNA interference and act as RNA scaffolds to enhance heat tolerance. In. *Int. J. Mol. Sci.* 21:6808. doi: 10.3390/ijms21186808

Tsybulskyi, V., Mounir, M., and Meyer, I. M. (2020). R-chie: a web server and R package for visualizing cis and trans RNA–RNA, RNA–DNA and DNA–DNA interactions. *Nucleic Acids Res.* 48:e105. doi: 10.1093/nar/gkaa708

Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., et al. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001. doi: 10.1038/nmeth.1529

Vicens, Q., and Kieft, J. S. (2022). Thoughts on how to think (and talk) about RNA structure. *Proc. Natl. Acad. Sci. USA* 119:e2112677119. doi: 10.1073/pnas.2112677119

Wan, Y., Qu, K., Ouyang, Z., and Chang, H. Y. (2013). Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* 8, 849–869. doi: 10.1038/nprot.2013.045

Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., et al. (2012). Genome-wide measurement of RNA folding energies. *Mol. Cell* 48, 169–181. doi: 10.1016/j.molcel.2012.08.008

Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709. doi: 10.1038/nature12946

Wang, H., Lu, X., Zheng, H., Wang, W., Zhang, G., Wang, S., et al. (2023). RNAsmc: a integrated tool for comparing RNA secondary structure and evaluating allosteric effects. *Comput. Struct. Biotechnol. J.* 21, 965–973. doi: 10.1016/j.csbj.2023.01.007

Ziv, O., Farberov, S., Lau, J. Y., Miska, E., Kudla, G., and Ulitsky, I. (2021). Structural features within the NORAD long noncoding RNA underlie efficient repression of Pumilio activity. *bioRxiv*. [Epub ahead of preprint] doi: 10.1101/2021.11.19.469243. abstract