Processing of ribosomal RNA

Ralf Bundschuh

Department of Physics
Department of Chemistry and Biochemistry
Division of Hematology, Department of Internal Medicine
Center for RNA Biology
Interdisciplinary Biophysics Graduate Program

Ohio State University

July 23, 2024

Overview

1. The ribosome
2. To H1 or not to H1
3. Leader-Trailer helices

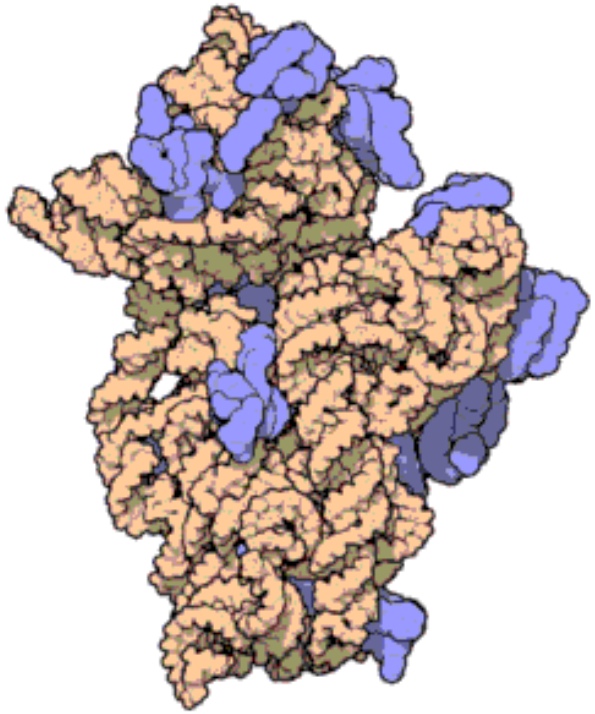
Overview

1. The ribosome
2. To H1 or not to H1
3. Leader-Trailer helices

Ribosome background

Background on the ribosome

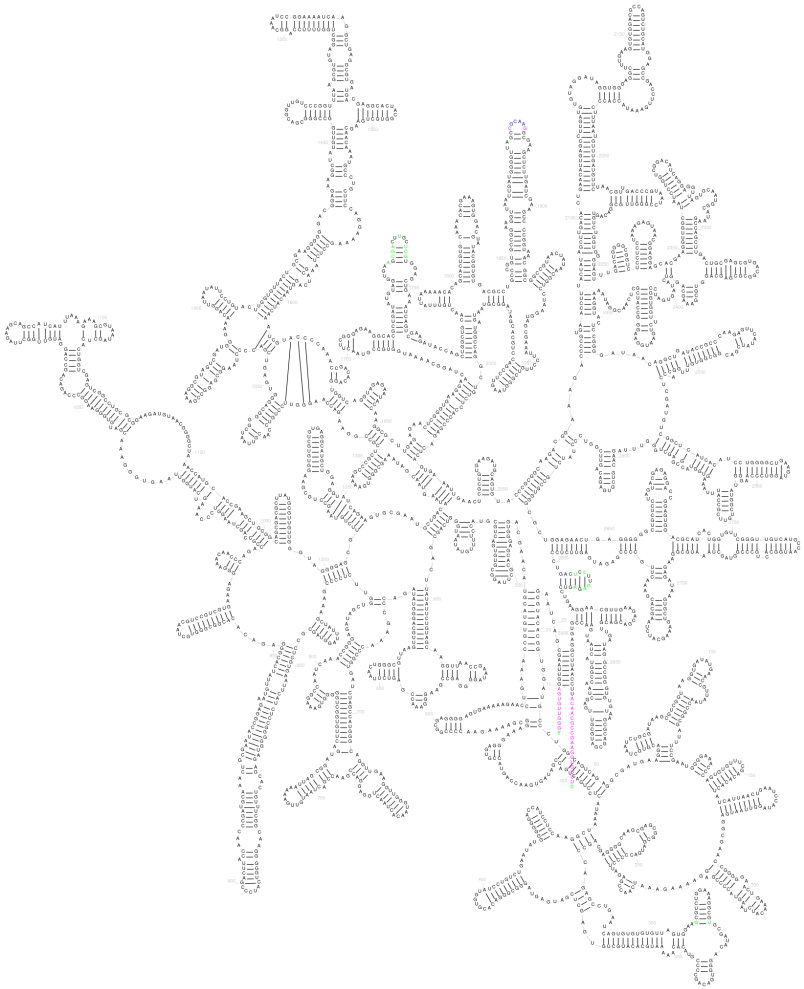
- The ribosome makes all proteins in living systems
- It has two subunits (in bacteria 30S and 50S)
- Both subunits are separately assembled and join on the transcript to be translated
- 30S subunit contains 16S rRNA, 50S subunit 23S and 5S
- All three rRNAs are transcribed together in an operon
- A lot of processing has to happen to make a mature ribosomal subunit

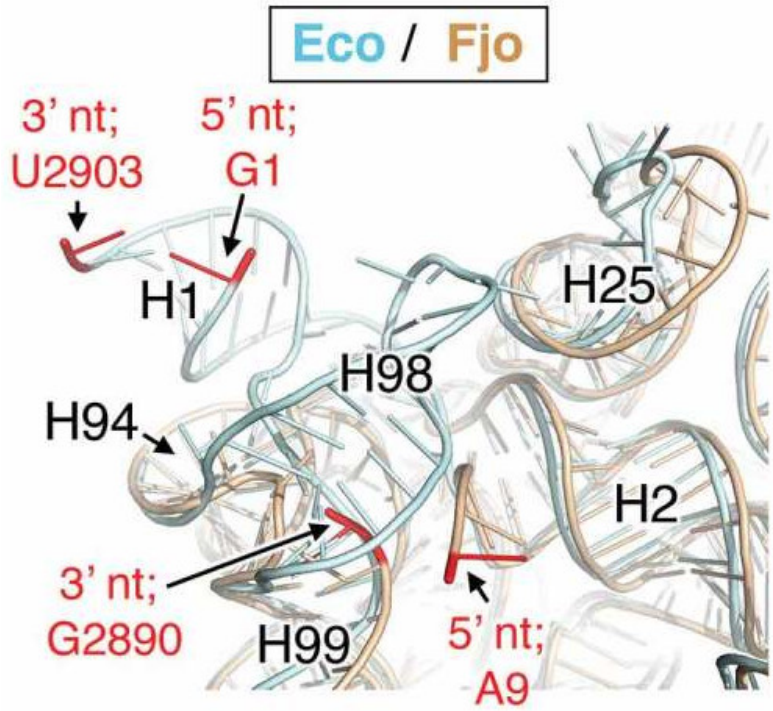


Overview

1. The ribosome
2. To H1 or not to H1
   - The problem
   - Approach #1
   - Approach #2
   - Conclusions on helix H1
3. Leader-Trailer helices

23S rRNA secondary structure from RNACentral and helix H1



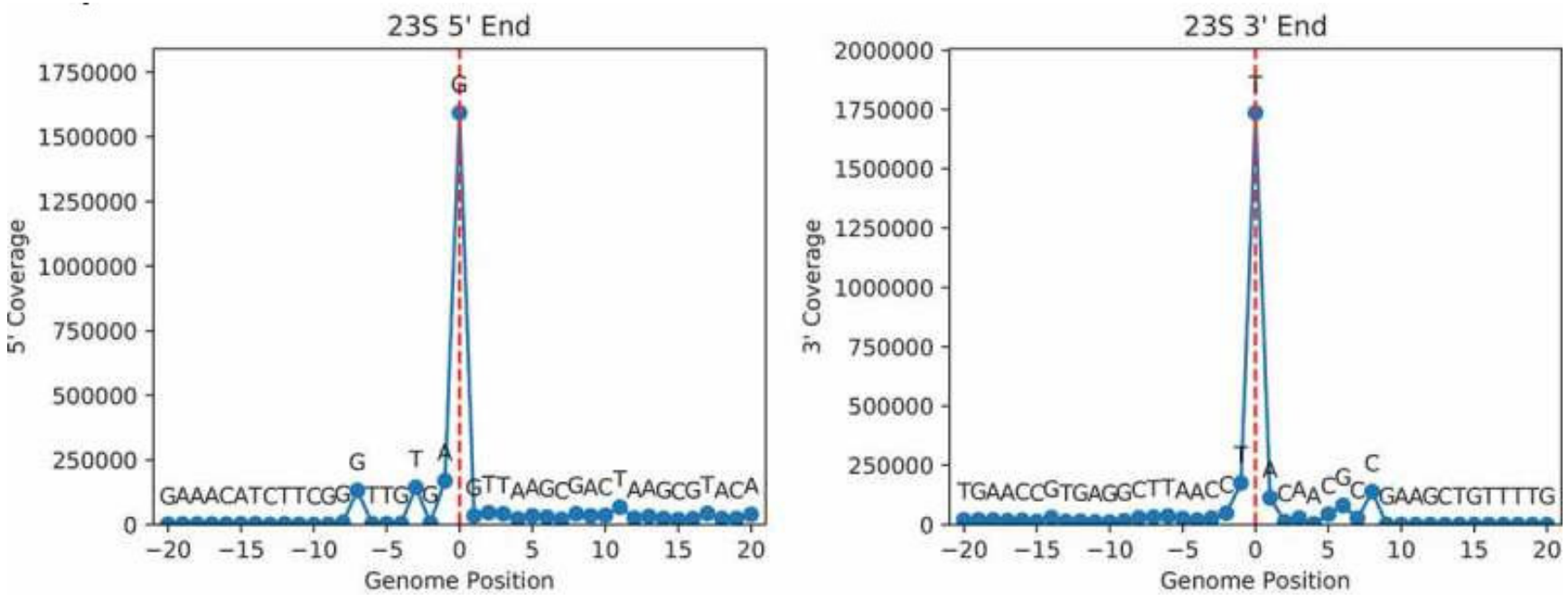


Helix H1 exists in *E. coli* but not in *F. johnsohniae*

Specialized sequencing for a few organisms

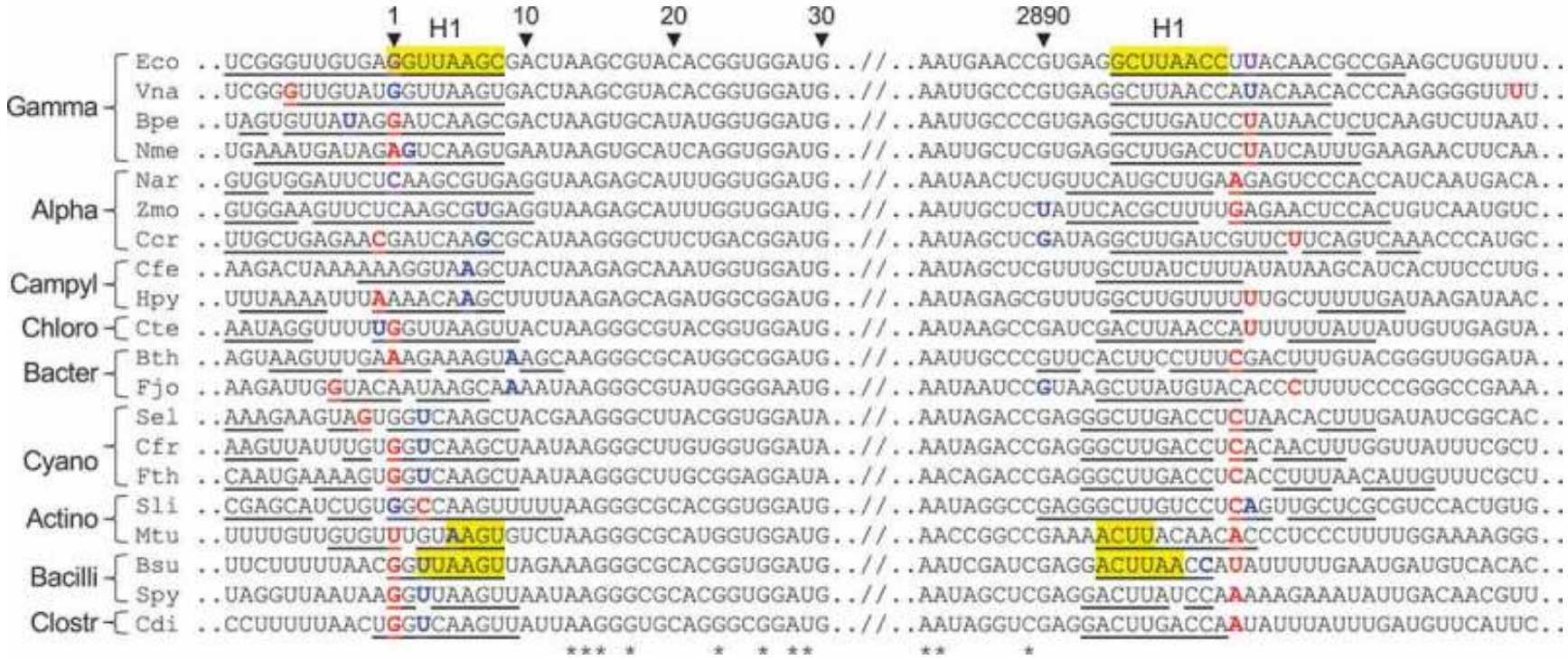
End-specific sequencing

- Find all prokaryotic dRNA-seq, TSS-seq, Term-seq, and Rend-seq data sets
- Yields data sets for 19 prokaryotic species
- Map to 23S rRNA and record ends



Results from approach #1

23S findings from specialized sequencing



- Several organisms are missing H1
- rRNA ends are horrendously misannotated
- Organisms mostly either have the full H1 or no H1 at all

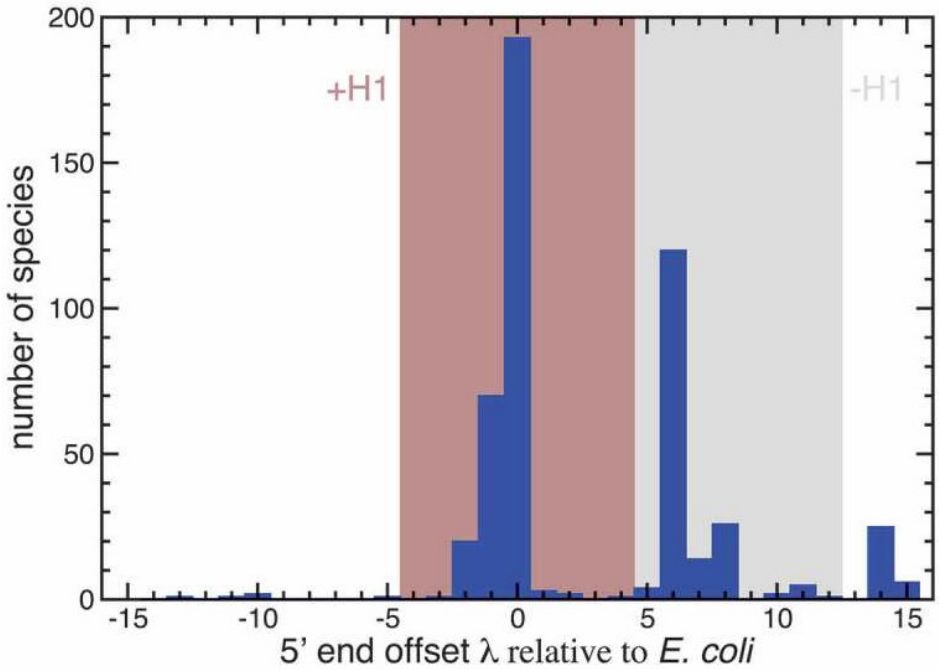
Trash to treasures - more organisms

Large scale 5' end reannotation

- Obtained metatranscriptomic data set from TARA Oceans consortium
- TARA Oceans used Clontech SMART-seq kit following RiboZero
- SMART-seq adds non-encoded GGG to the 5' end of an RNA before adding the template switch oligo
- Identified 10^8 reads that start with GGG and include a conserved motif from the 5' end of 23S rRNA
- Align these reads to a database of prokaryotic 23S rRNA sequences
- Target identifies species; location identifies 5' end
- Obtained 5' annotations for 23S rRNA in 498 species

Again either full or no H1

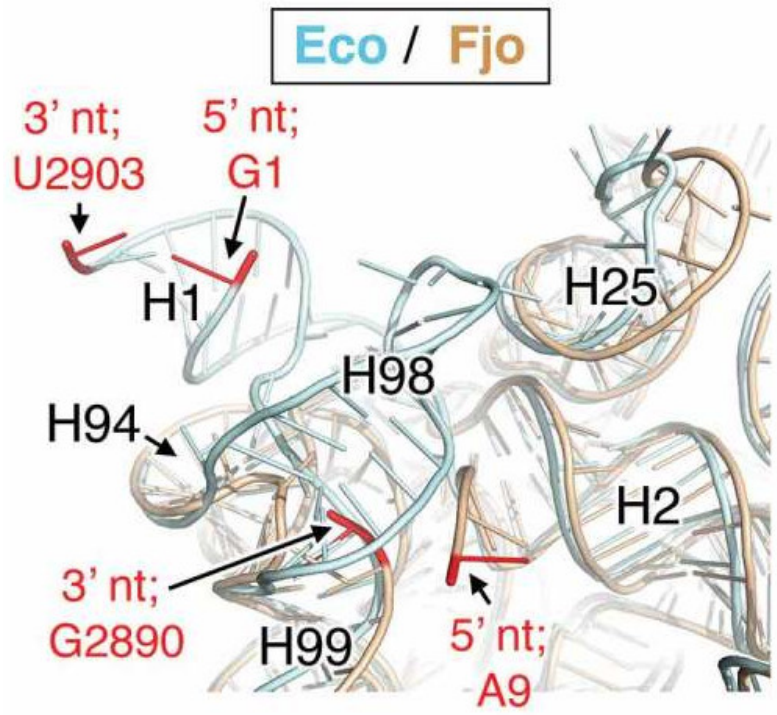
Histogram of 5' end positions



- Species either do or don't have H1
- About half of the species have H1

Keeping H1 aligns with phylogeny

Phylogenetic distribution of H1



- Retention of H1 strongly correlates with phylogeny
- Retention of H1 also strongly correlates with existence of H98

Conclusions

Conclusions

- Helix H1 is removed in about half of bacteria
- Absence of H1 correlates strongly with absence of H98
- Helix H1 is likely excised during subunit maturation
- SMART-seq allows 5' end annotation *en masse*

Future work

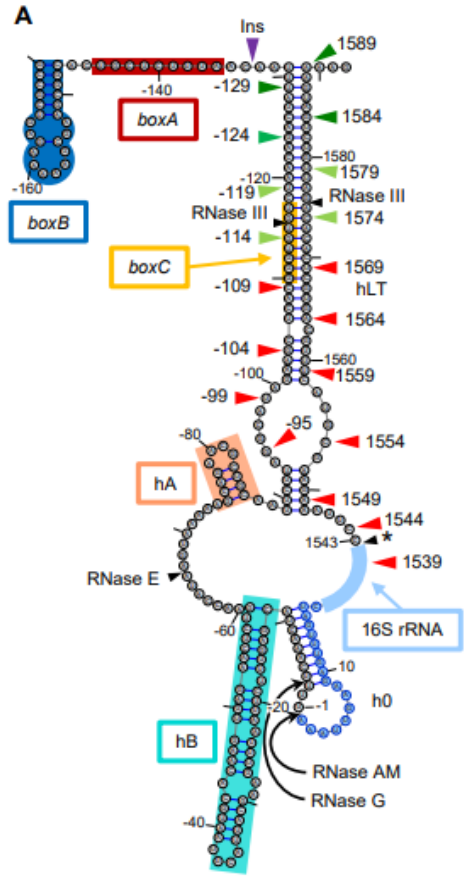
- Apply SMART-seq trick more broadly to improve 5' end annotations

Overview

1. The ribosome
2. To H1 or not to H1
3. Leader-Trailer helices
   - The problem
   - Measuring leader-trailer helices
   - Approach #1
   - Approach #2
   - Results
   - Conclusions on LT helices

Leader-trailer helices

- *E. coli* ribosome subunits assemble from processed rRNAs *in vitro*
- However, this process is slow and requires non-physiological conditions
- In *E. coli*, rRNA are flanked by leader-trailer helices
- In *E. coli*, leader-trailer helices are essential for *in vivo* assembly



The question

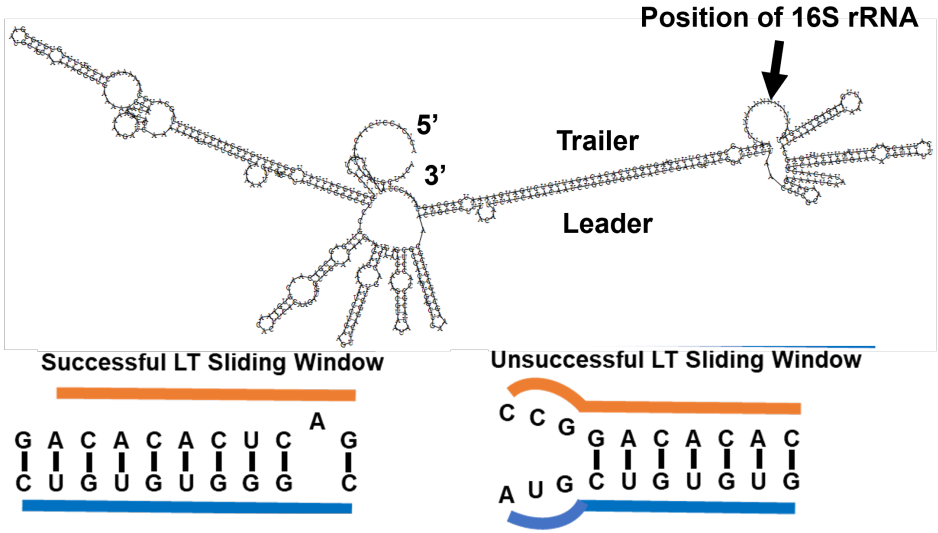
The question

- If LT helices are that important, they should be universal
- But prior work (Saito *et al.*, 2000) found no LT helices in *Deinococcus radiodurans* and some others after analyzing a handfull of species
- Question: Who has leader-trailer helices?

Quantitating LT helices

Counting successful sliding windows (SSWs)

- Concatenate leader and trailer
- Predict structure
- Successful Sliding Window (SSW): A window of length 10 that has at least 8 base pairs with a window of size 10 in the opposite strand

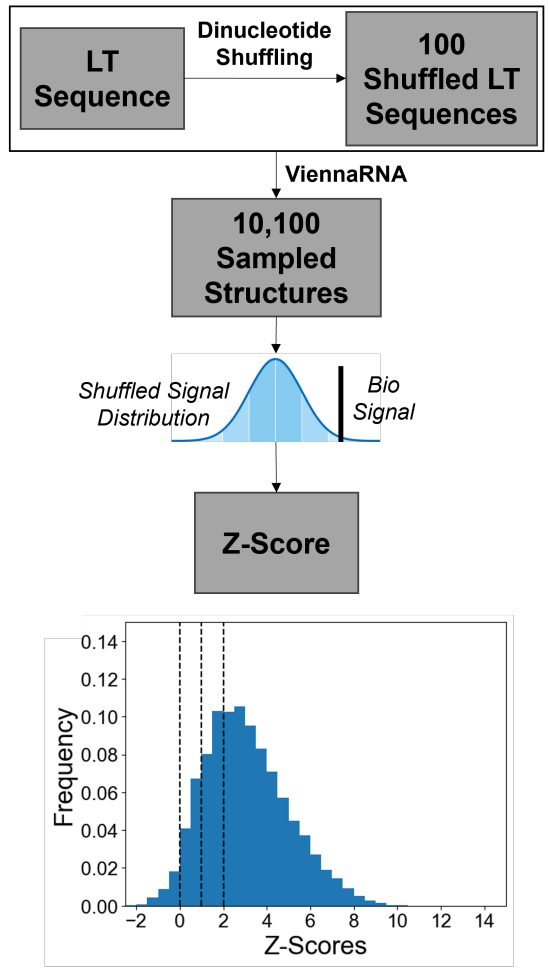


- Quantify LT "helicity" by counting SSWs

Shuffling method

The shuffling method

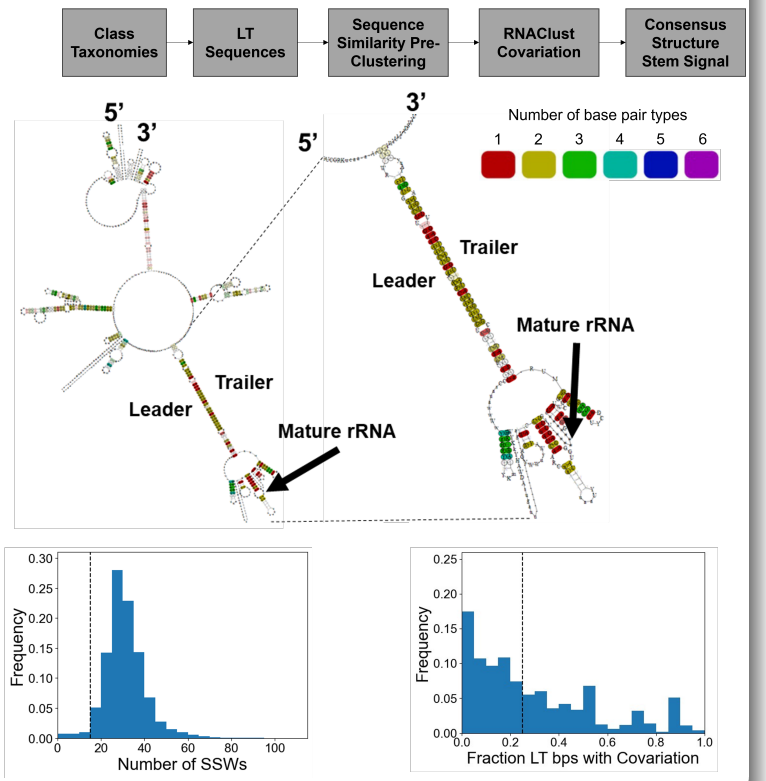
- Take leader and trailer up to next annotated gene (max 500/300nts)
- Average #SSWs for each sequence over 100 randomly sampled structures
- Redo for 100 dinucleotide shuffled sequences
- Calculate z-score
- Distribution skews heavily positive
- Downside: false negatives due not including enough nts



Covariation method

The covariation method

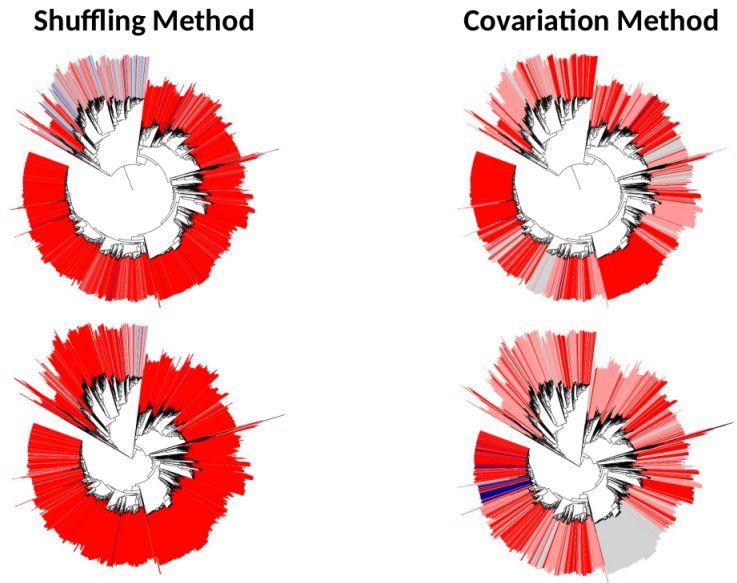
- Take 500nt leader and 300nt trailer
- Divide sequences by phylogenetic class
- Pre-cluster sequences
- Run RNAClust
- Get LocARNA consensus structure
- Count SSWs and look at covariation
- Upside: Irrelevant parts of sequence "disappears"
- Downside: missing predictions due to clusters of one



Approaches complement each other

Comparison of approaches

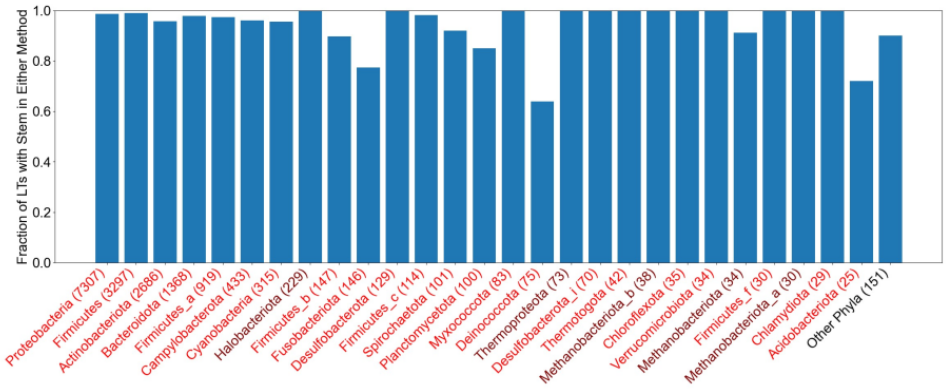
- Show on phylogenetic tree
- Here: Proteobacteria
- Red: has LT helix
  Blue: no LT helix
  Gray: no prediction
- Most species have a LT helix in at least one of the two methods
- Actually, most have one in both



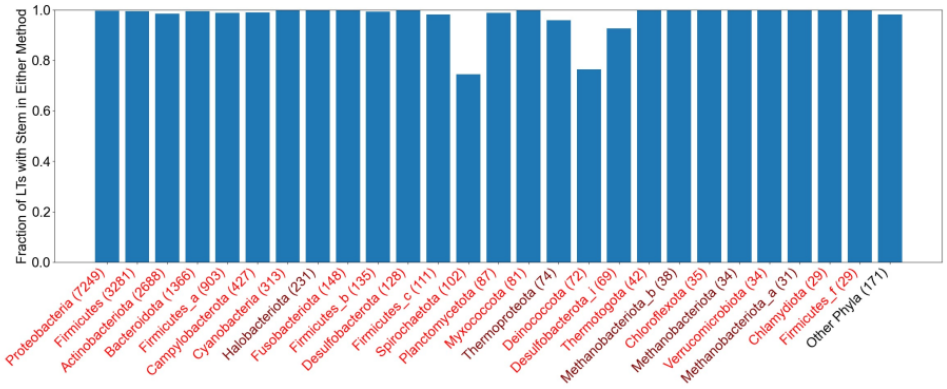
LT helices are ubiquitous

Statistics for all bacterial and archaeal phyla

16S



23S

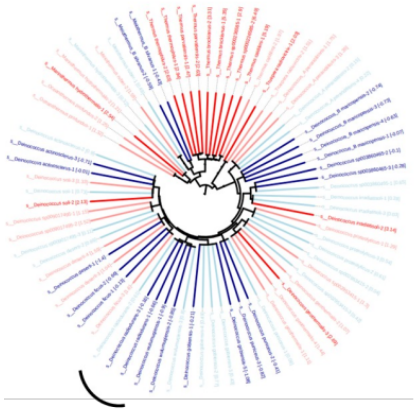


- LT helices are ubiquitous
- Manual inspection of "missing" cases still finds LT helices

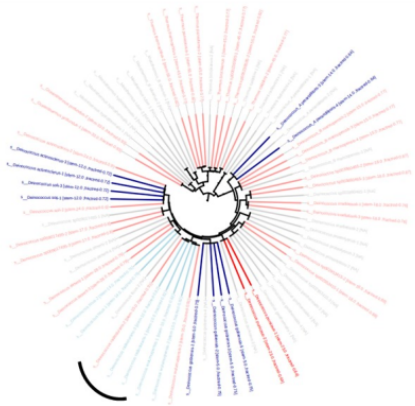
Deinoccocota

Situation in Deinoccocota

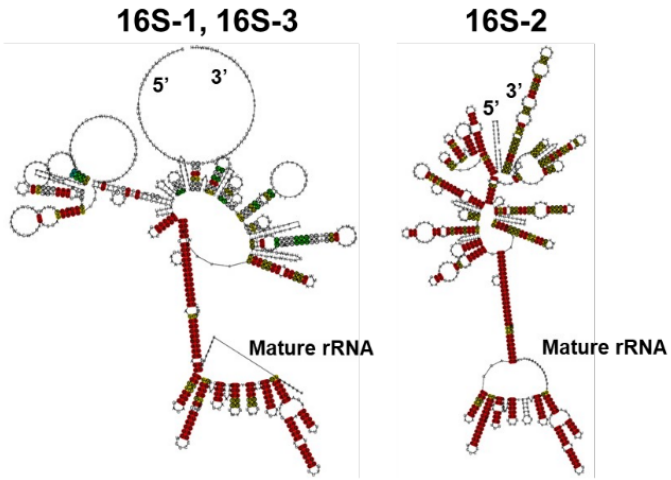
Shuffling Method



Covariation Method



D. radiodurans Consensus Structures



- Even Deinococcota have clear LT helices
- They are interrupted by bigger bulges/internal loops

Conclusions

Conclusions

- LT helices are ubiquitous in bacteria and archaea
- Quantifying "helicity" is actually not trivial
- One always has to calculate ensemble quantities

Future work

- Look at TT and LT structures systematically, which are more diverse
- Look at operon-to-operon differences within a species

Overview

1. The ribosome
2. To H1 or not to H1
3. Leader-Trailer helices

The people

Bryan Gemler

Elan Shatoff

Ben Warner

Kurt Fredrick

$$$

- National Science foundation
  - MCB-2029502, MCB-2344534
- National Institutes of Health
  - R01 GM072528