Functionally annotating the *Candida auris* genome

Skyler Sung[1,2], Diana Proctor[1], Sean Conlan[1], Clayton Deming[1], Shih-Queen Lee-Lin[1], Julia A. Segre[1]

[1] Microbial Genomics Section, Translational and Functional Genomics Branch, NHGRI, NIH, Bethesda, MD, [2] Amherst College, Amherst, MA
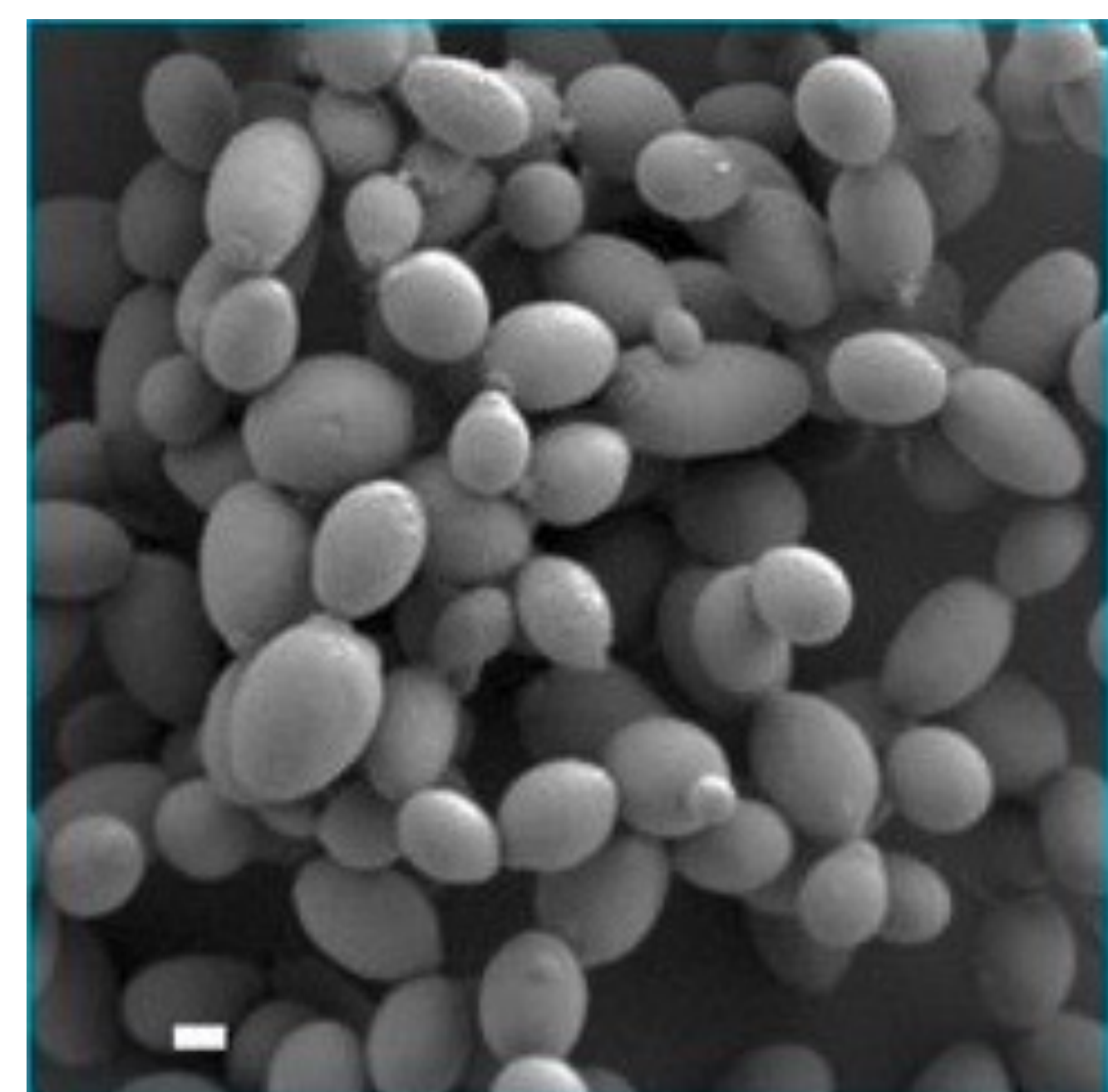
BACKGROUND

Why study *Candida auris*?
- 39% mortality rate
- Hospital outbreaks
- Persists on human skin & healthcare surfaces
- Pan-resistant

Chicago Outbreak
- First cases in 2016
- Colonized at high prevalence
- Clade IV

Horton *et al.* mSphere vol. 5,1. 2020

GOALS

1. Test performance of Funannotate (against NISC)
2. Assess whether RNA-seq data improves annotations
3. Compare our annotations to those in GenBank

METHODS

Sample Preparation

Genome long-read nanopore sequencing

Funannotate Pipeline

Mask highly repetitve regions

Method 1: BUSCO training set → *Ab initio* gene prediction → Functional Annotation

Method 2: RNA-seq data → RNA-seq assembly → *Ab initio* gene prediction → Add untranslated regions → Functional Annotation

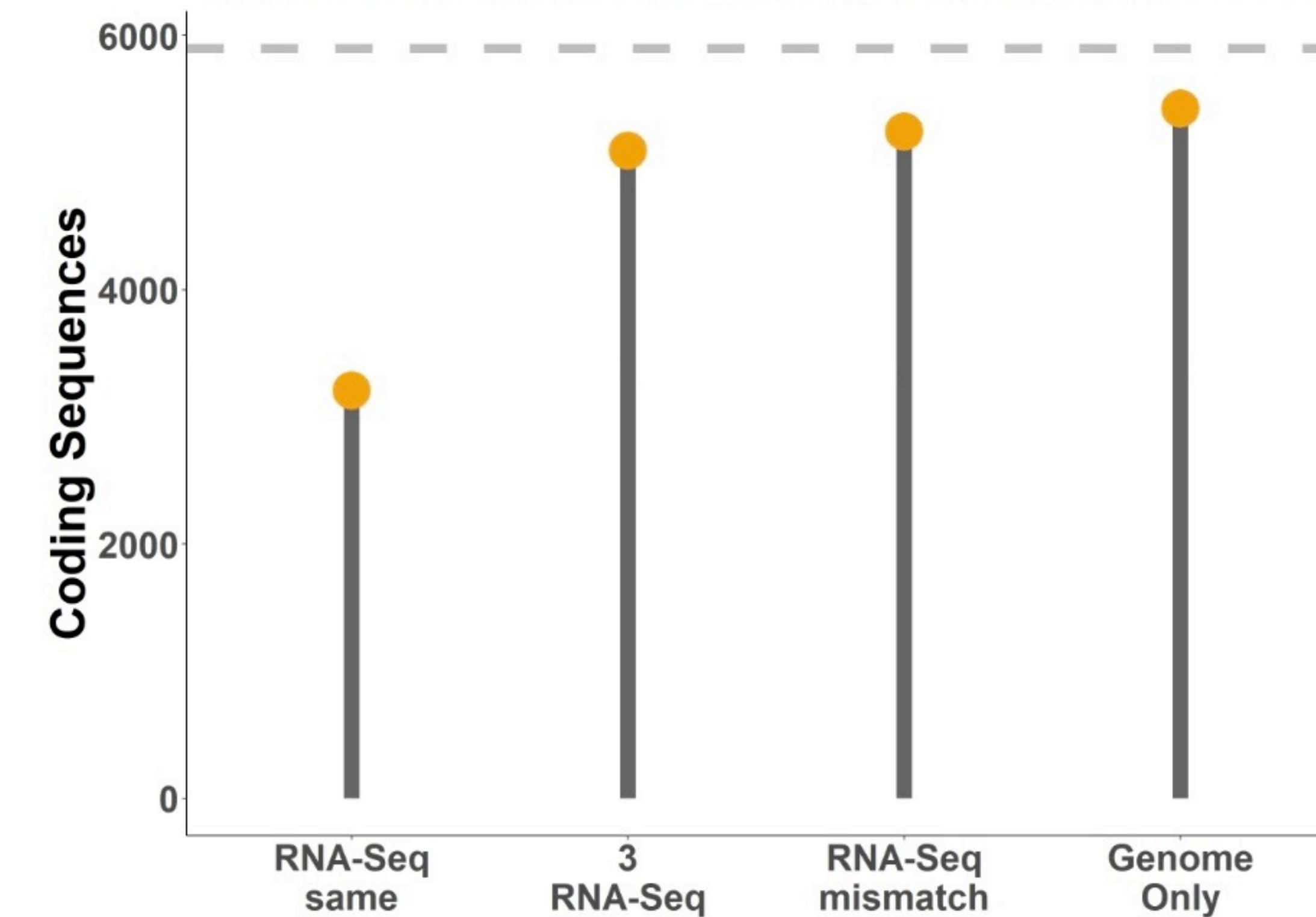**Method 3**: Pull annotations from NCBI GenBank

Visualization

Annotation GFF3 analysis

RESULTS

1) Funannotate underestimates CDS's & 2) RNA-Seq data undercuts performance
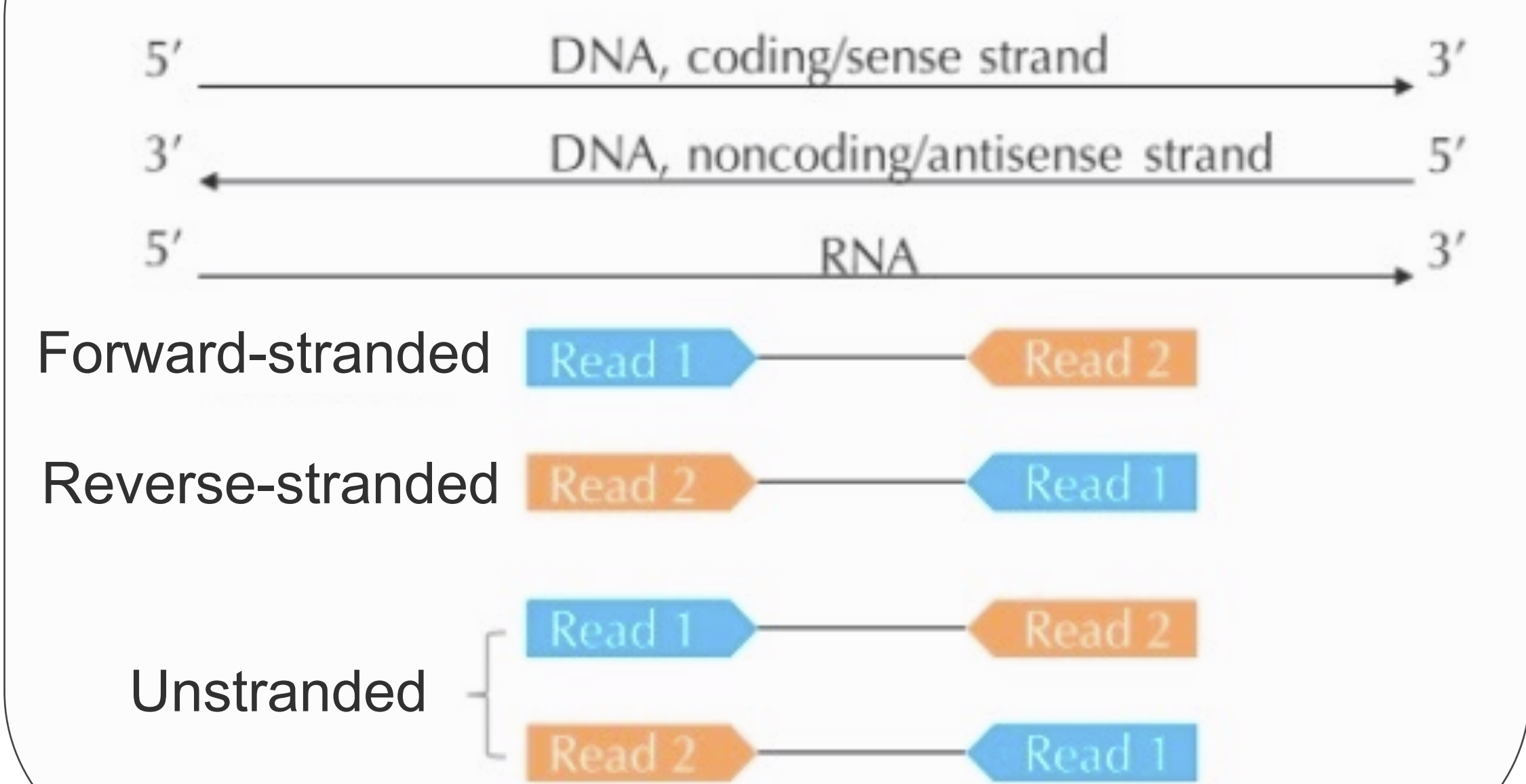

NISC vs Funannotate Annotations: CDS
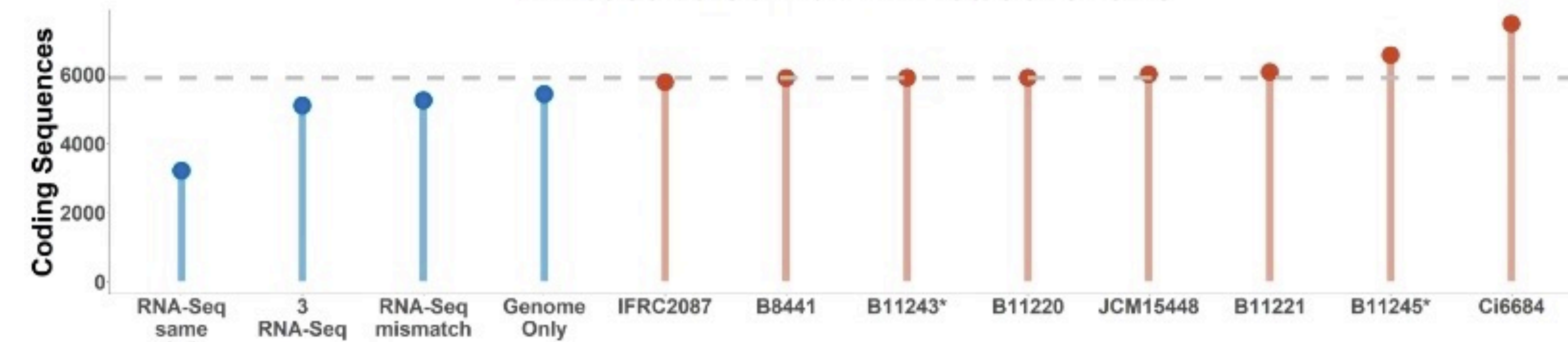
Troubleshooting

Readmapping → RNA-seq strandedness parameter → RNA-seq strandedness
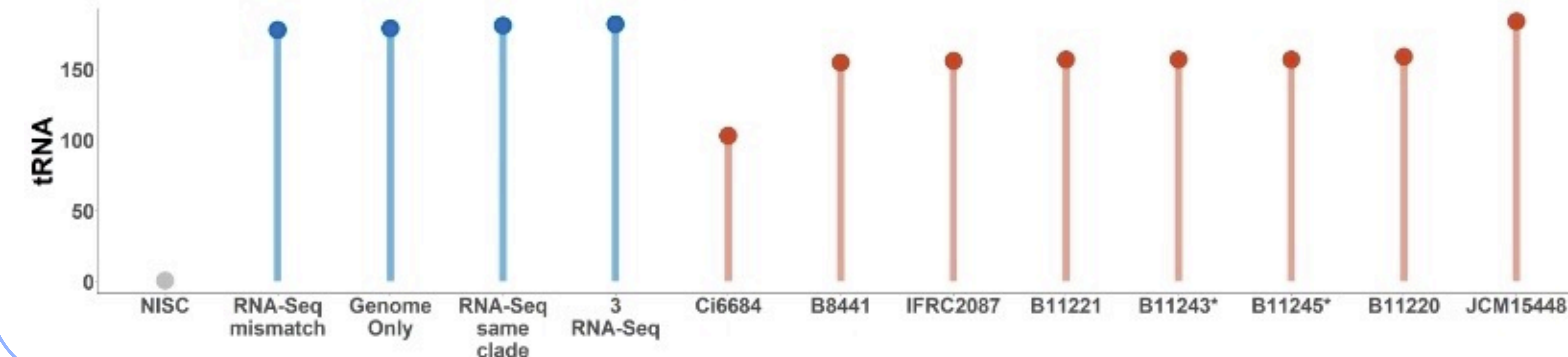
3) Funannotate annotations inconsistent with GenBank's

In-house vs GenBank Annotations: CDS

In-house vs GenBank Annotations: tRNA

FUTURE DIRECTIONS

- Annotations with RNA-seq data from different conditions
- Post scripts for future fungal annotation studies