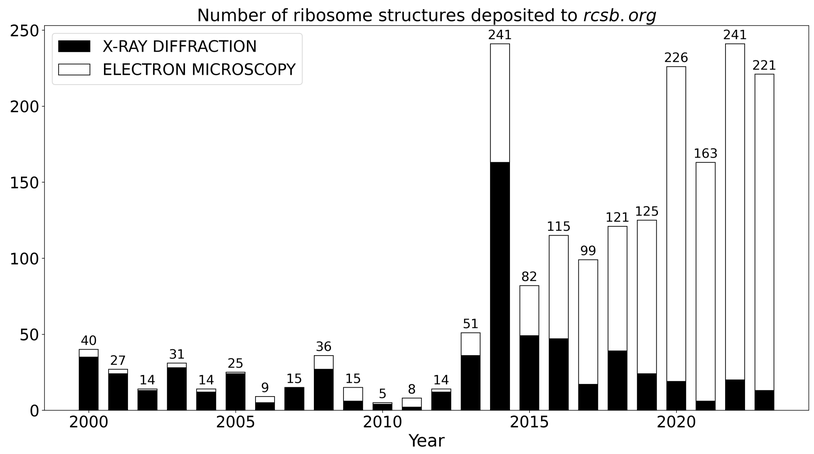
Computational methods for unraveling ribosome structural heterogeneity

Artem Kushner[1], Shiqi Yu[2], Aryan Tajmir Riahi[3], Simcha Srebnik[2], and *Khanh Dao Duc*[*,1,3]

[1] Department of Mathematics, [2] Department of Chemical and Biological Engineering, [3] Department of Computer Science, University of British Columbia, Vancouver, BC V6T1Z2, Canada

Background

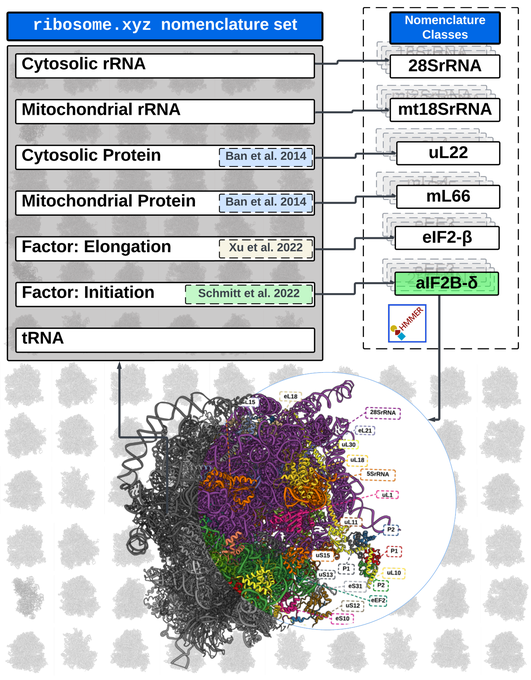


- Advances in Cryo-EM led to a surge of new ribosome structures accounting for conformational and compositional heterogeneity (Poitevin *et al.*, Molecules 2020).
- Challenge for comparative analysis and structural heterogeneity study: inconsistent annotation of chain units and nomenclature across databases.
- We developed bioinfomatic tools and computational methods to study ribosome heterogeneitiy from 3D structures and EM maps.

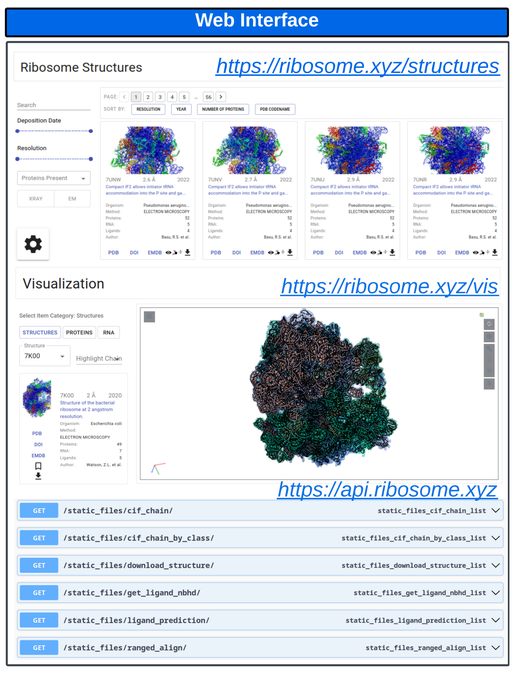
Comprehensive database and visualization of ribosome structures

- **RiboXYZ**: Database and web-application from PDB structures (Kushner *et al.*, NAR 2022)

Classification framework for chain units



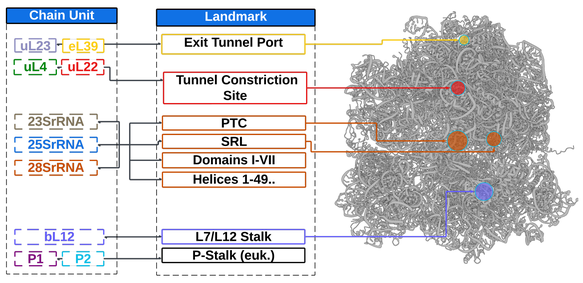
https://ribosome.xyz



- The database contains **1679 structures** (894 full structures, 422 (363) large (small) subunits) at ≤ 4 Å resolution with consistent annotation of components (e.g. RNA, proteins...).

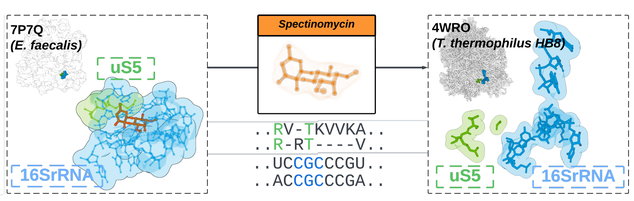
Examples of applications and tools

- Identification of landmark sites



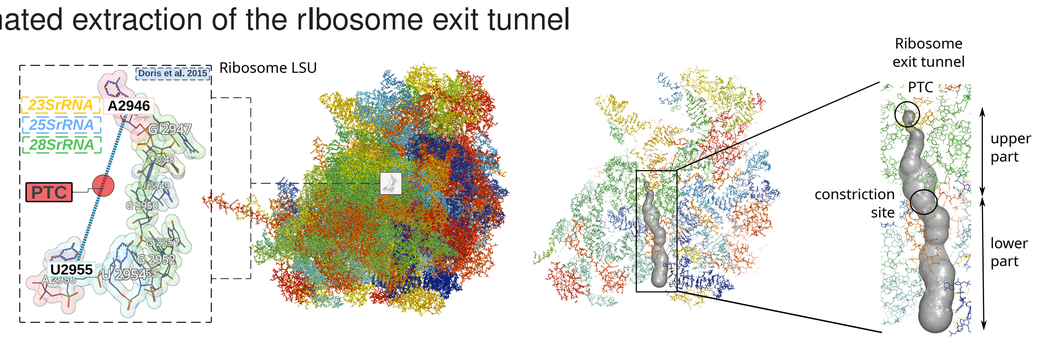
- Standardized nomenclature enables automated extraction of structural "landmarks" (including functional clusters or specific loci)

- Binding site prediction



- Prediction of Spectinomycin binding site in *T. Thermophilus* based on its neighborhood residues in *E. faecalis*.
- Database contains 1260 other unique Drugbank-registered compounds.

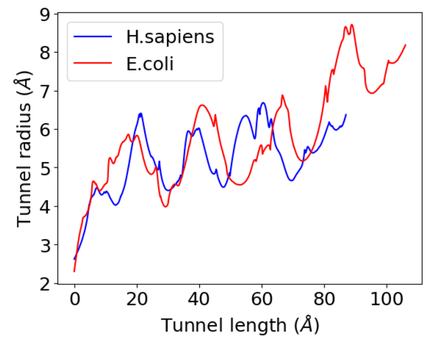
- Automated extraction of the rIbosome exit tunnel



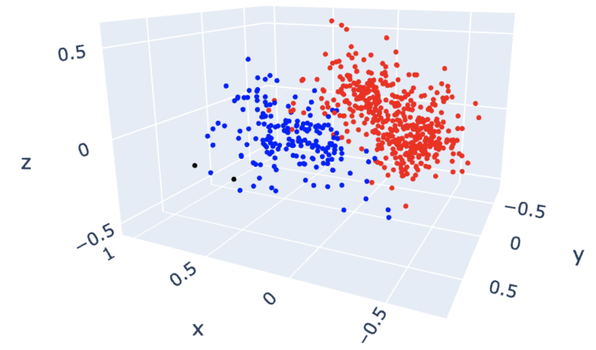
- Tunnel search algorithm (Sehnal *et al.*, NAR 2012) constrained by PTC and constriction site landmarks

Ribosome exit tunnel heterogeneity

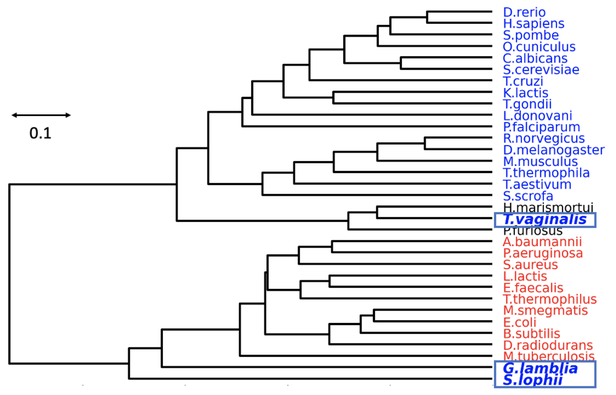
Extracting tunnel radial variations along its centerline



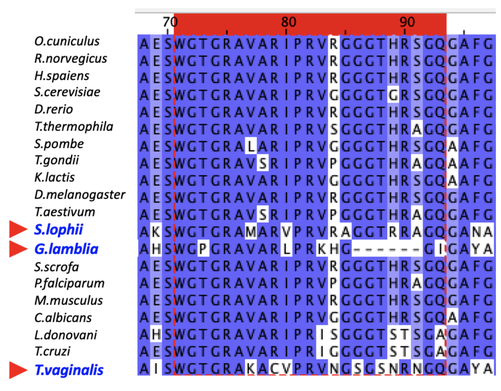
Multi-Dimensional Scaling (MDS) from radial plots (**759** tunnel structures / **31** species)

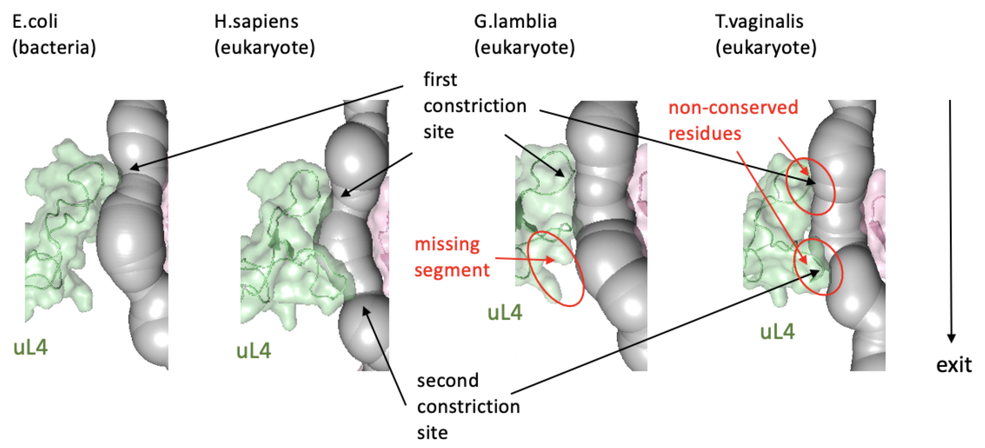


Hierarchical clustering of tunnels



Multiple sequence alignment of uL4



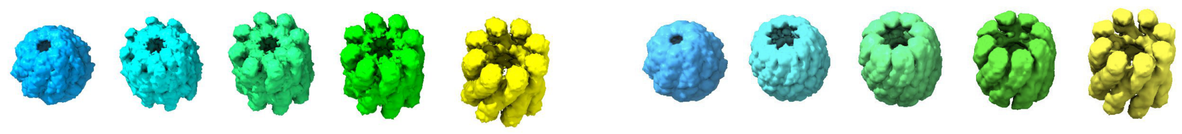


- Data driven study of the tunnels enables discovering geometric differences across domains of life and species.
- Functional and evolutionary consequences for small ORFs (Yu et al. Biophysical J., 2023)

Transport based methods for studying heterogeneity from EM maps

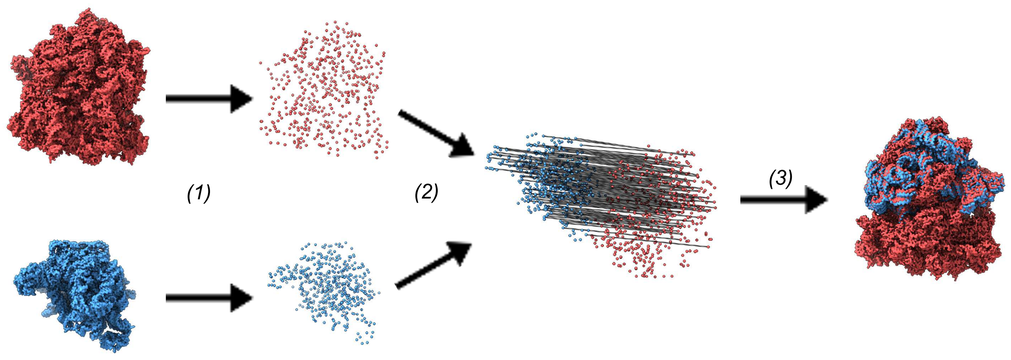
- Optimal Transport (OT) theory $\longrightarrow$ non-linear metric for comparing EM maps
  (Tajmir Riahi *et al.*, IEEE/ACM Transcactions on Computational Biology and Bioinformatics, 2023)
- Application: Improved interpolation (Ecoffet et al., Bioinformatics 2020)

Linear morphing (UCSF Chimera) OT-based morphing (MorphOT)



- Partial 3D alignment (registration) of EM maps (Tajmir Riahi *et al.*, MLSB NeuriPS 2023)

Alignment of large ribosomal subunit to global structure



- EMPOT procedure: (1) Point cloud conversion (2) transport based matching (3) rigid body alignment
- *Applications* (in progress): Ribosomal assembly landscape, atomic model fitting

References

1. K. Dao Duc, S. Batra, N. Bhattacharya, J.H.D. Cate and Y.S. Song (2019) Differences in the path to exit the ribosome across the three domains of life, Nucleic Acids Research, gkz106.
2. A. Kushner, A.S. Petrov, K. Dao Duc, (2022) RiboXYZ: A comprehensive database for ribosome structures, Nucleic Acids Research, gkac939
3. A. Tajmir Riahi, C. Zhang, J.Chen, A. Condon, K. Dao Duc (2023), EMPOT: partial alignment of density maps and atomic model fitting using unbalanced Gromov-Wasserstein divergence, NeurIPS workshop on Machine Learning in Structural Biology (accepted)

(*) **Contact:** kdd@math.ubc.ca

Research supported in part by NSERC Discovery Grants (PG 22R3468 and PG 11R02115), UBC STAIR grant and NFRFE-2019-00486 grant.