DFG Deutsche Forschungsgemeinschaft

UNIVERSITÄT
LEIPZIG

Class 2 CCA-adding Enzymes in Archaeal Lineages

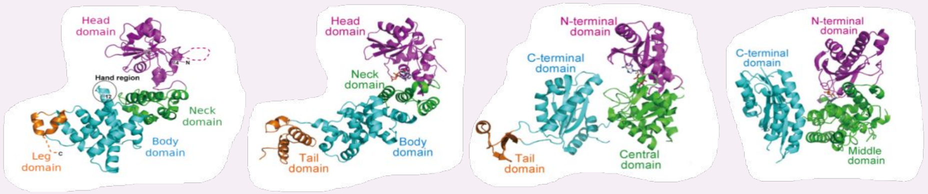
Bled 2025
Ella Cassidy
Computation EvoDevo
Supervisor Sonja Prohaska

What are tRNA-nucleotidyltransferases (CCA-adding enzymes)

CCA-adding enzymes

Superfamily

- class 2 (**CCA2**)
  - PAP
  - CCA
- class 1 (**CCA1**)
  - CCA
  - PAP



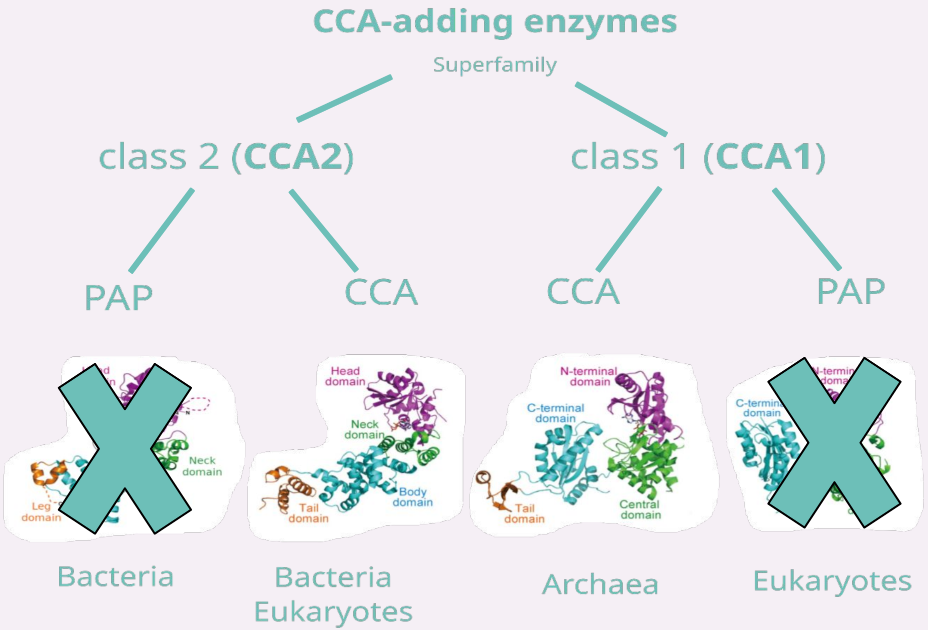
Bacteria

Bacteria
Eukaryotes

Archaea

Eukaryotes

What are tRNA-nucleotidyltransferases (CCA-adding enzymes)



CCA-adding enzymes

Superfamily

class 2 (**CCA2**)

class 1 (**CCA1**)

PAP

CCA

CCA

PAP

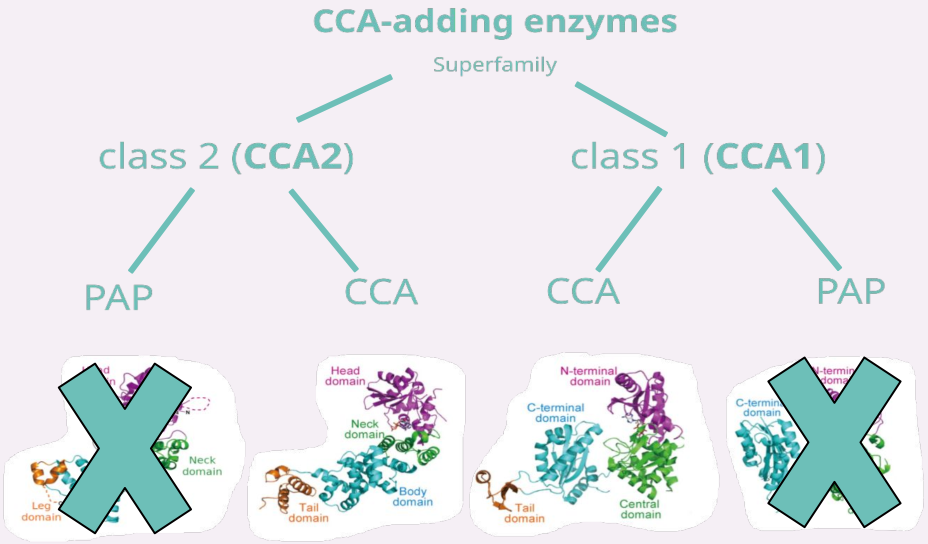
Bacteria

Bacteria
Eukaryotes

Archaea

Eukaryotes

What are tRNA-nucleotidyltransferases (CCA-adding enzymes)



CCA-adding enzymes

Superfamily

class 2 (**CCA2**)

class 1 (**CCA1**)

PAP

CCA

CCA

PAP

Bacteria

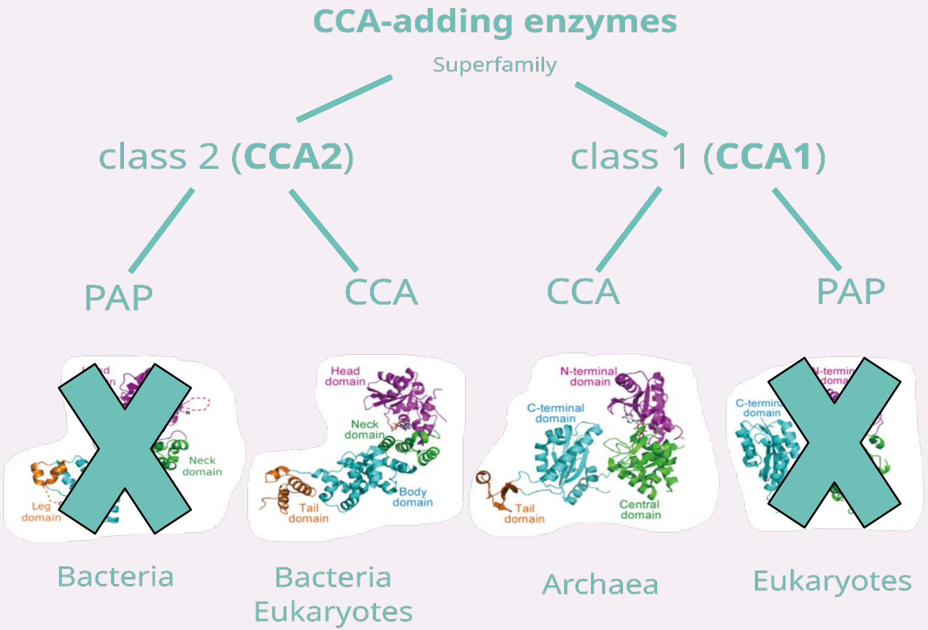
Bacteria
Eukaryotes

Archaea

Eukaryotes

Low sequence identity between class 1 and class 2

What are tRNA-nucleotidyltransferases (CCA-adding enzymes)



CCA-adding enzymes

Superfamily

class 2 (**CCA2**)

class 1 (**CCA1**)

PAP

CCA

CCA

PAP

Bacteria

Bacteria
Eukaryotes

Archaea

Eukaryotes

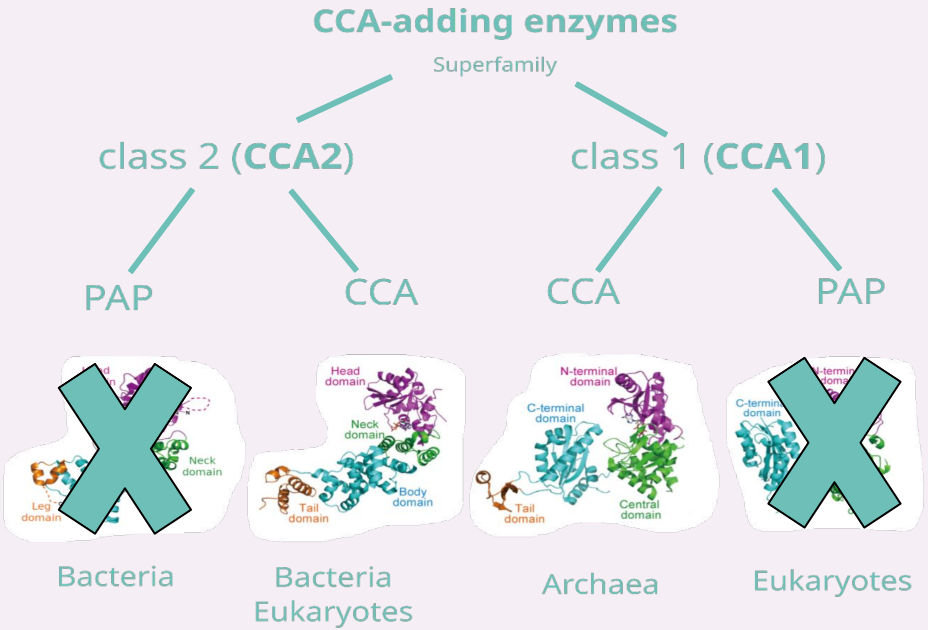
Low sequence identity between class 1 and class 2

1. Essential for tRNA maturation

2. Quality control mechanisms

→ Abundant in all living organisms and **essential**

What are tRNA-nucleotidyltransferases (CCA-adding enzymes)



CCA-adding enzymes

Superfamily

class 2 (**CCA2**)

class 1 (**CCA1**)

PAP

CCA

CCA

PAP

Bacteria

Bacteria
Eukaryotes

Archaea

Eukaryotes

Low sequence identity between class 1 and class 2

1. Essential for tRNA maturation
2. Quality control mechanisms

→ Abundant in all living organisms and **essential**

DO WE FIND CCA2 IN ARCHAEAL LINEAGES?

Why study Archaeal evolution?

→ first discovery only 40 years ago

→ understanding the tree of life

→ not been directly linked to cause human disease **but** they're currently being studied in their possible implications in microbiome studies, existing on skin, tracts, guts (so not just extreme environments) → **regulatory effects?**

→ Important in **hydrosphere**, greenhouse gases (methane) solely metabolic product of methanogenic archaea (some produce, some consume), Ammonia-oxidizing archaea (AOA) nitrification

Why study Archaeal evolution?

→ first discovery only 40 years ago

→ understanding the tree of life

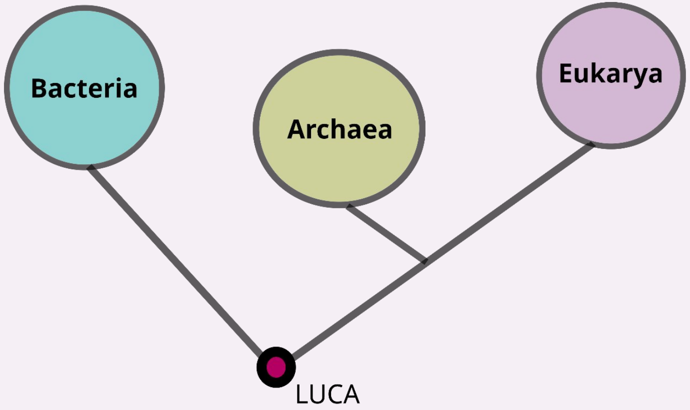
→ not been directly linked to cause human disease but they're currently being studied in their possible implications in microbiome studies, existing on skin, tracts, guts (so not just extreme environments) → regulatory effects?

→ Important in hydrosphere, greenhouse gases (methane) solely metabolic product of methanogenic archaea (some produce, some consume), Ammonia-oxidizing archaea (AOA) nitrification

Difficulties

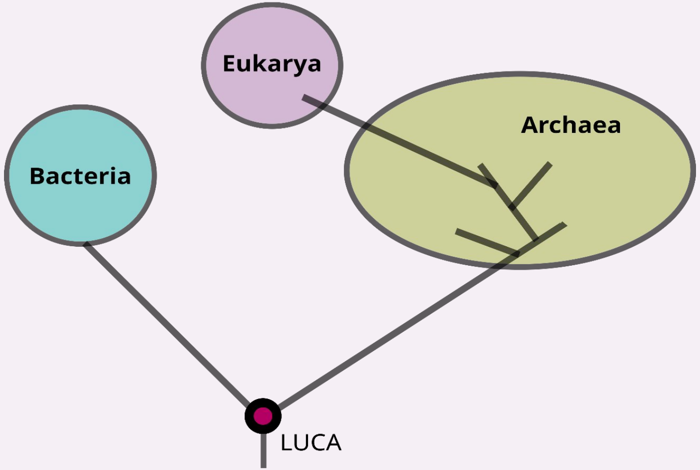
- Culturing challenges (difficult to grow, unknown nutrients, slow growth rates)
- Missing phylogeny/not well resolved
- Severely underrepresented in databases

Why study Archaeal evolution?



Three-domain tree

Bacteria, Archaea, and Eukarya represent three distinct domains of life. archaea and eukaryotes split from a common ancestor, forming two distinct lineages.

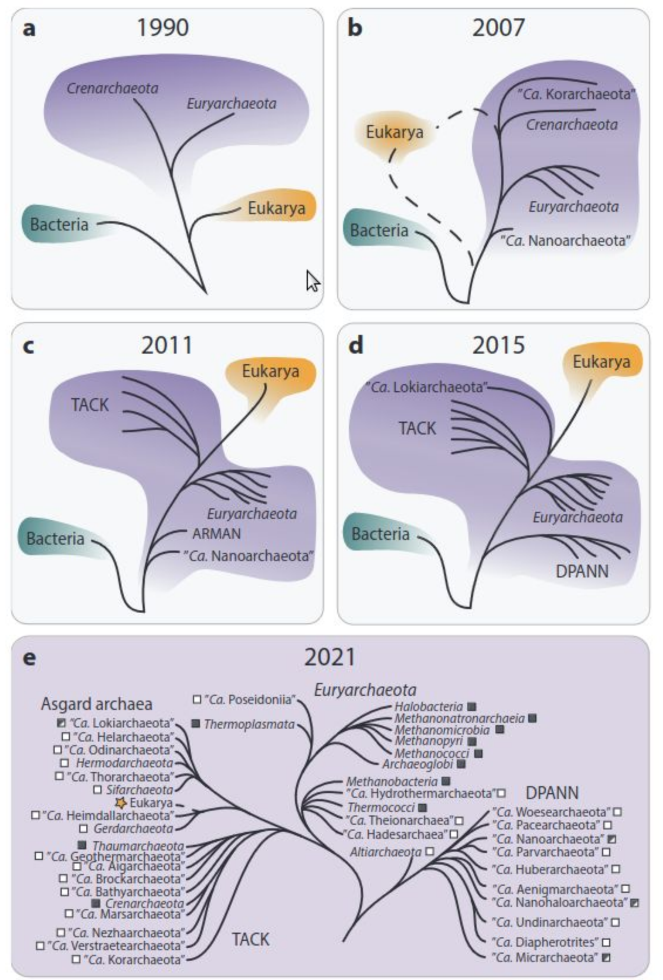


Two-domain tree

Eukaryotes emerge from within Archaea, meaning they are just a special type of Archaeon that later incorporated bacterial features (like mitochondria)

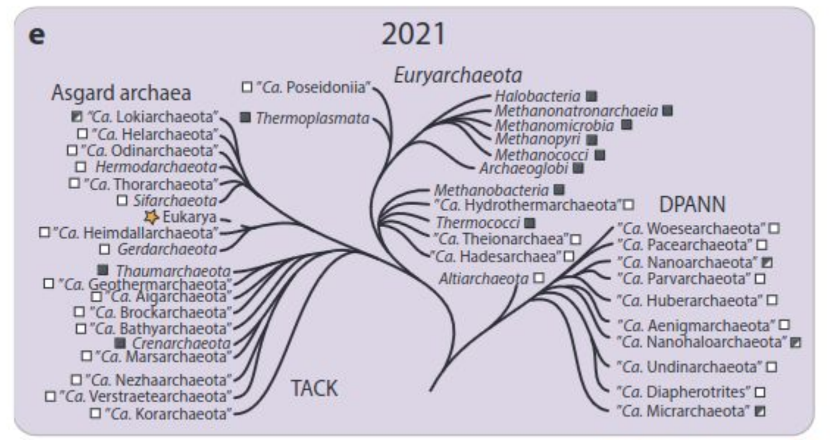
Archaeal current phylogeny

→ expansion of archaeal phylogenies



Archaeal current phylogeny

→ expansion of archaeal phylogenies

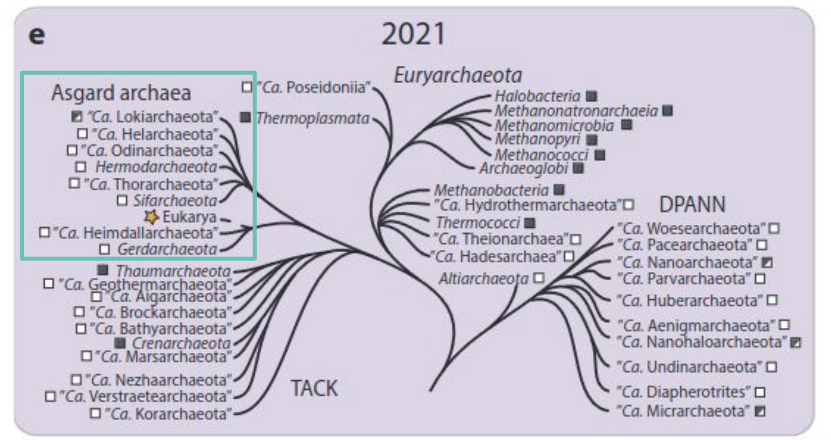


Archaeal current phylogeny

→ expansion of archaeal phylogenies

→ ESPs found in Asgards

- **Endomembrane systems**
- **Small GTPases**
- **Membrane trafficking machineries**
- **Vesicle biogenesis proteins**
- **Post-translational protein modification systems**
- **Cytoskeletal components**



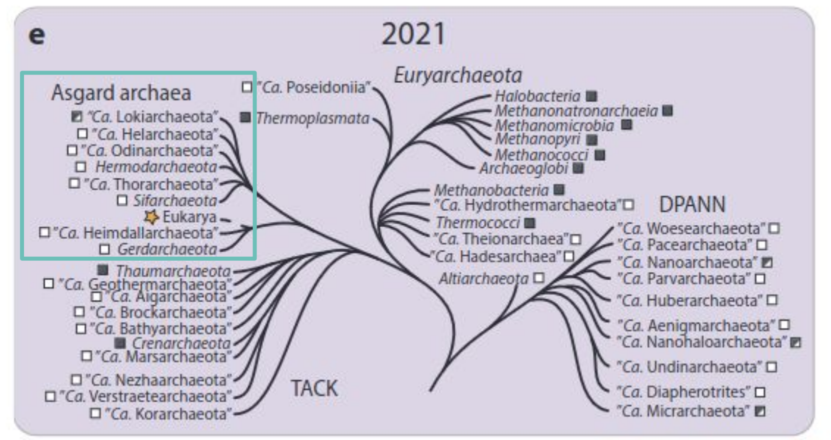
Archaeal current phylogeny

→ expansion of archaeal phylogenies

→ ESPs found in Asgards

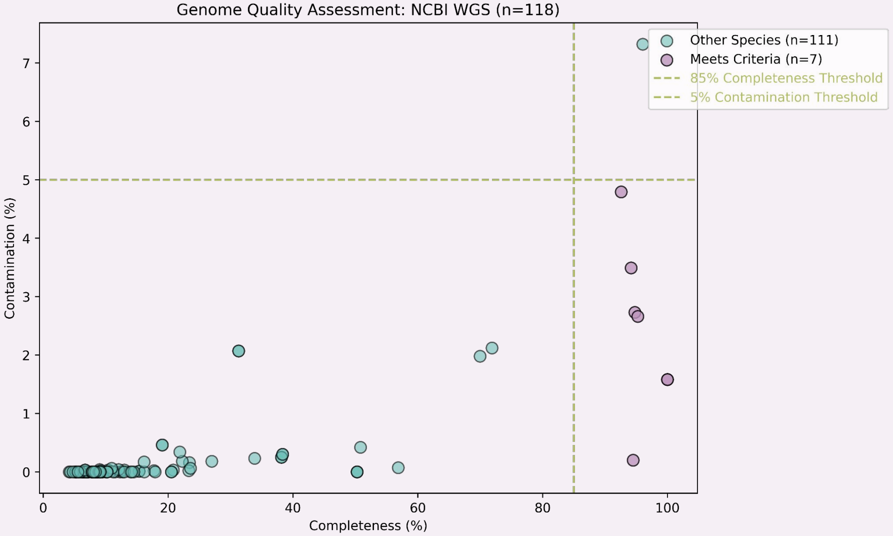
- Endomembrane systems
- Small GTPases
- Membrane trafficking machineries
- Vesicle biogenesis proteins
- Post-translational protein modification systems
- Cytoskeletal components

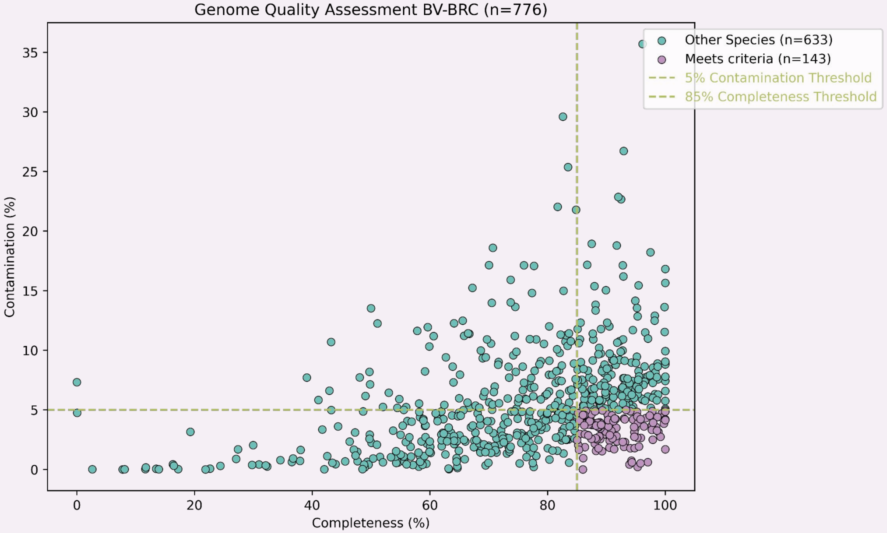
→ Asgard are phyla that eukaryotes originated from in the two-domain theory, specifically the **Heimdallarchaeota** family



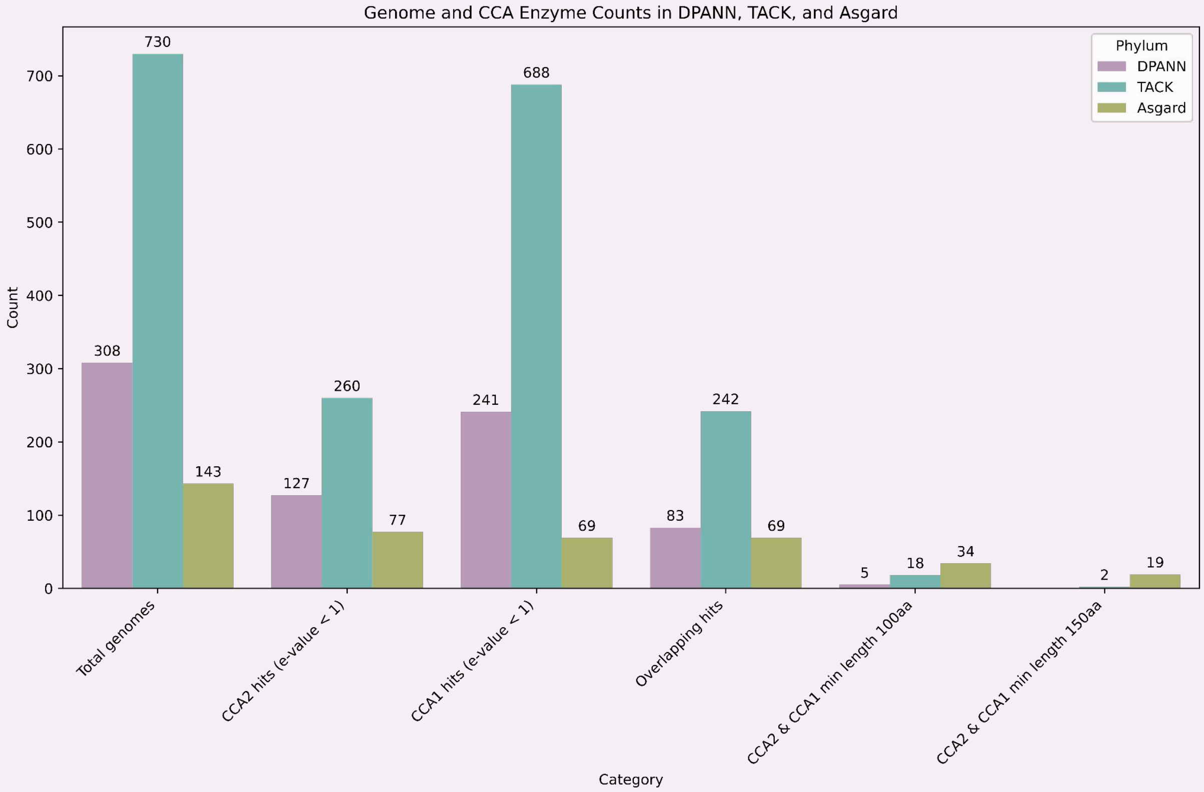
Challenges: NCBI WGS vs. BV-BRC genome quality

- Highlights the low/poor quality of **MAGs** from NCBI database
- NCBI data, only 118 sequences >100kbp for TACK, DPANN and Asgard
- BV-BRC data for Asgards only





Results! Using only BV-BRC data

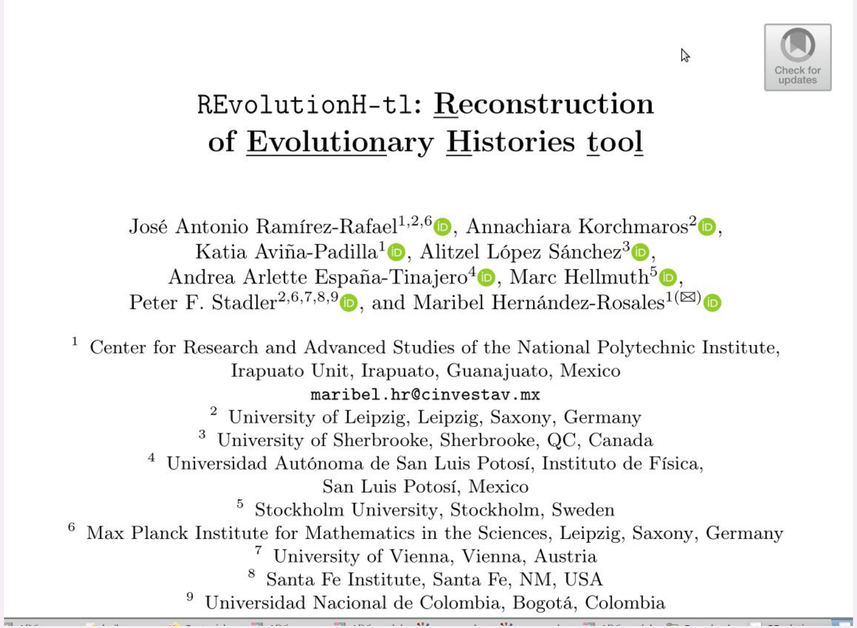


- Blasted CCA1 and CCA2 references against databases created for each phyla
- All sequences containing CCA2 from one family of Asgards
- CCA2 bacterial/eukaryotes ~400aa

→ **Heimdallarchaeota family**

Molecular phylogeny construction

- tool that reconstructs evolutionary histories
- not possible to do this without somewhat complete genomes
- 38 Heimdallarchaeota genomes with >85% complete, and <5% contamination, try to make a molecular phylogeny, without information about CCA1 or CCA2



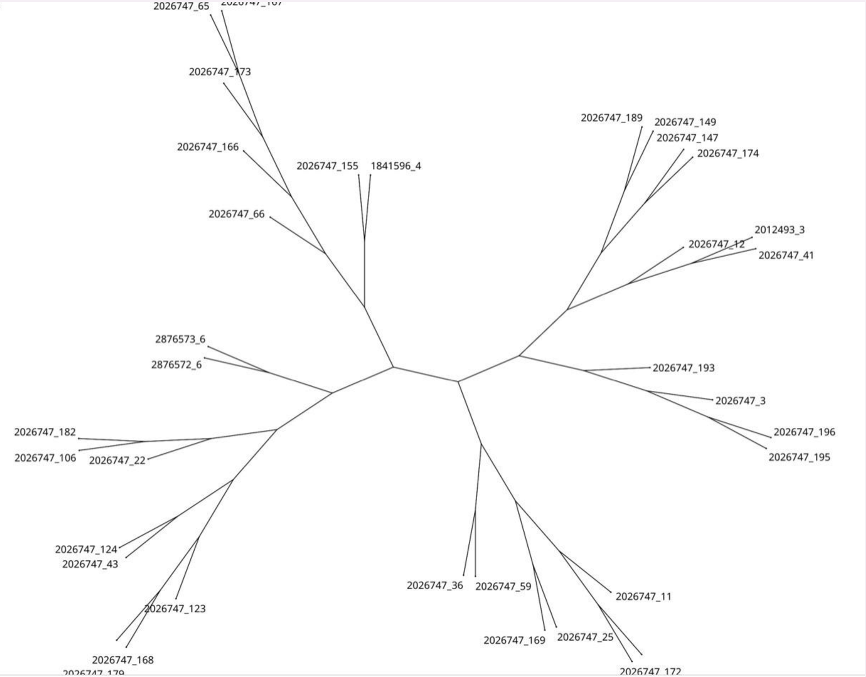
REvolutionH-tl: Reconstruction of Evolutionary Histories tool

José Antonio Ramírez-Rafael[1,2,6], Annachiara Korchmaros[2],
Katia Aviña-Padilla[1], Alitzel López Sánchez[3],
Andrea Arlette España-Tinajero[4], Marc Hellmuth[5],
Peter F. Stadler[2,6,7,8,9], and Maribel Hernández-Rosales[1](✉)

[1] Center for Research and Advanced Studies of the National Polytechnic Institute, Irapuato Unit, Irapuato, Guanajuato, Mexico
maribel.hr@cinvestav.mx
[2] University of Leipzig, Leipzig, Saxony, Germany
[3] University of Sherbrooke, Sherbrooke, QC, Canada
[4] Universidad Autónoma de San Luis Potosí, Instituto de Física, San Luis Potosí, Mexico
[5] Stockholm University, Stockholm, Sweden
[6] Max Planck Institute for Mathematics in the Sciences, Leipzig, Saxony, Germany
[7] University of Vienna, Vienna, Austria
[8] Santa Fe Institute, Santa Fe, NM, USA
[9] Universidad Nacional de Colombia, Bogotá, Colombia

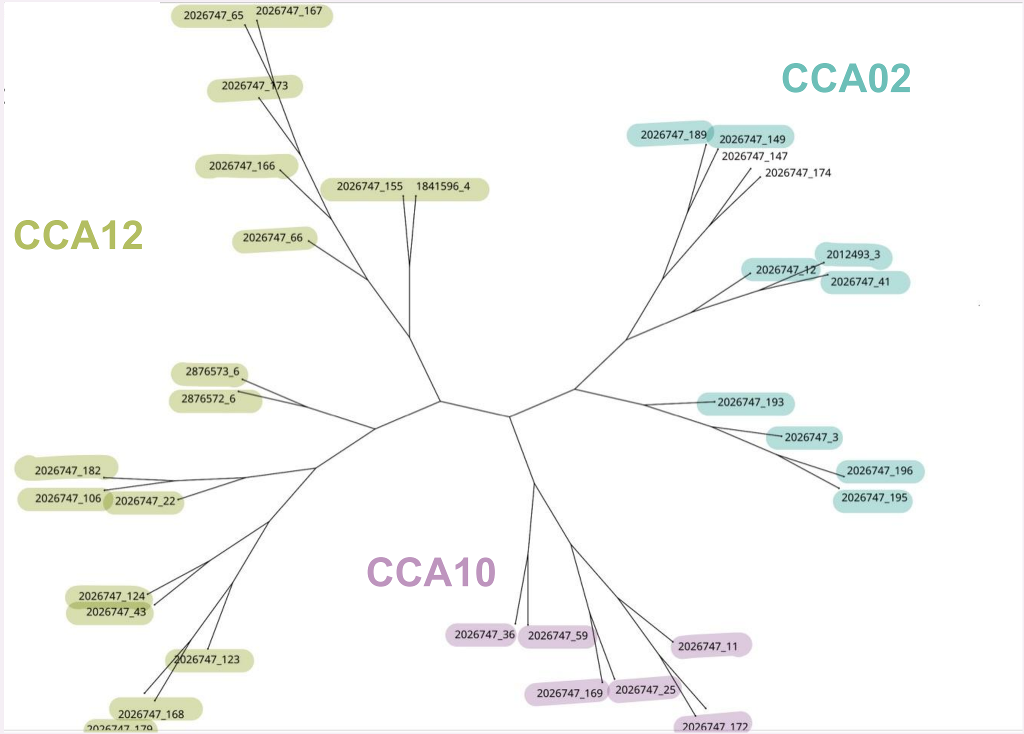
Molecular phylogeny construction

- tool that reconstructs evolutionary histories
- not possible to do this without somewhat complete genomes
- 38 Heimdallarchaeota genomes with >85% complete, and <5% contamination, try to make a molecular phylogeny, without information about CCA1 or CCA2



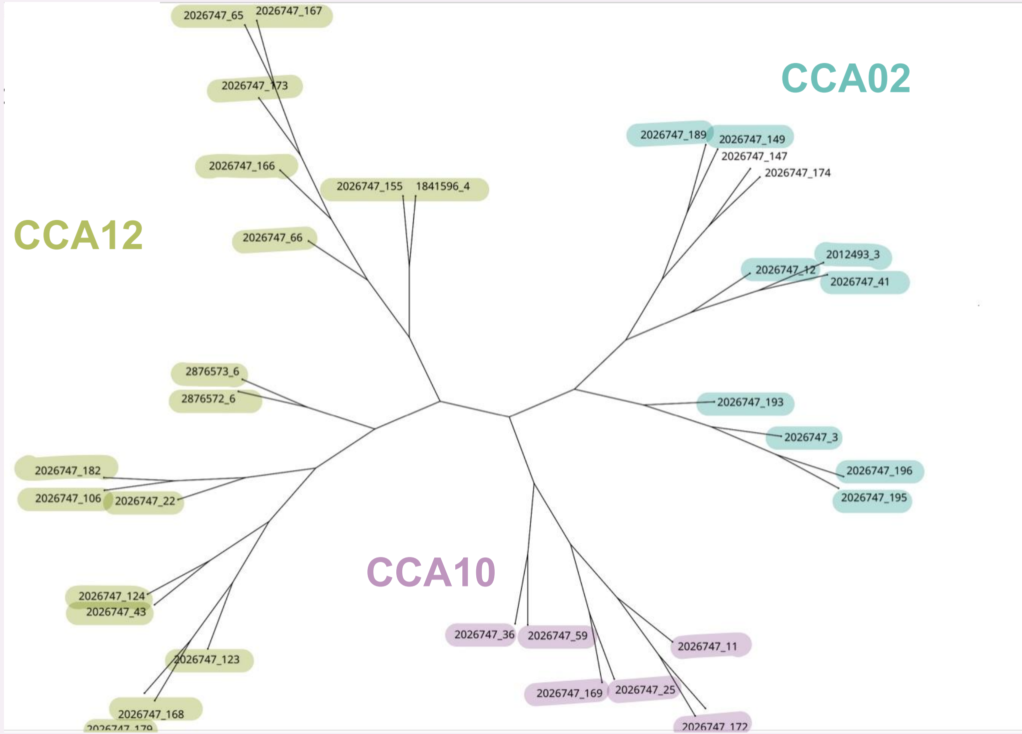
Molecular phylogeny construction

- tool that reconstructs evolutionary histories
- not possible to do this without somewhat complete genomes
- 38 Heimdallarchaeota genomes with >85% complete, and <5% contamination, try to make a molecular phylogeny, without information about CCA1 or CCA2



Molecular phylogeny construction

→ **issue:** no meaningful branch lengths to fully understand relationships



Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories

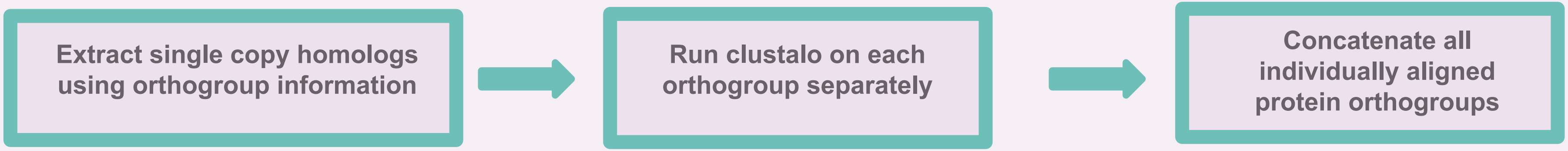
Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories

Extract single copy homologs using orthogroup information → **Run clustalo on each orthogroup separately** → **Concatenate all individually aligned protein orthogroups**

Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories



Extract single copy homologs using orthogroup information

Run clustalo on each orthogroup separately

Concatenate all individually aligned protein orthogroups

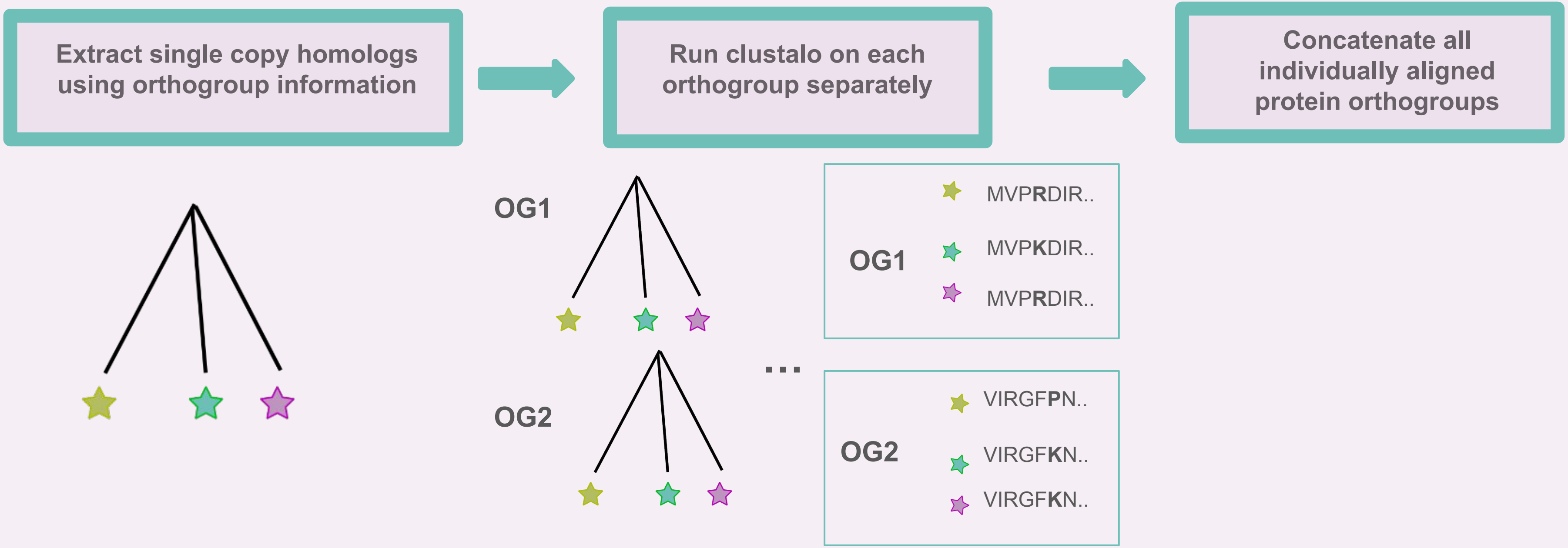
Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories

Extract single copy homologs using orthogroup information → Run clustalo on each orthogroup separately → Concatenate all individually aligned protein orthogroups

Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories



Extract single copy homologs using orthogroup information

Run clustalo on each orthogroup separately

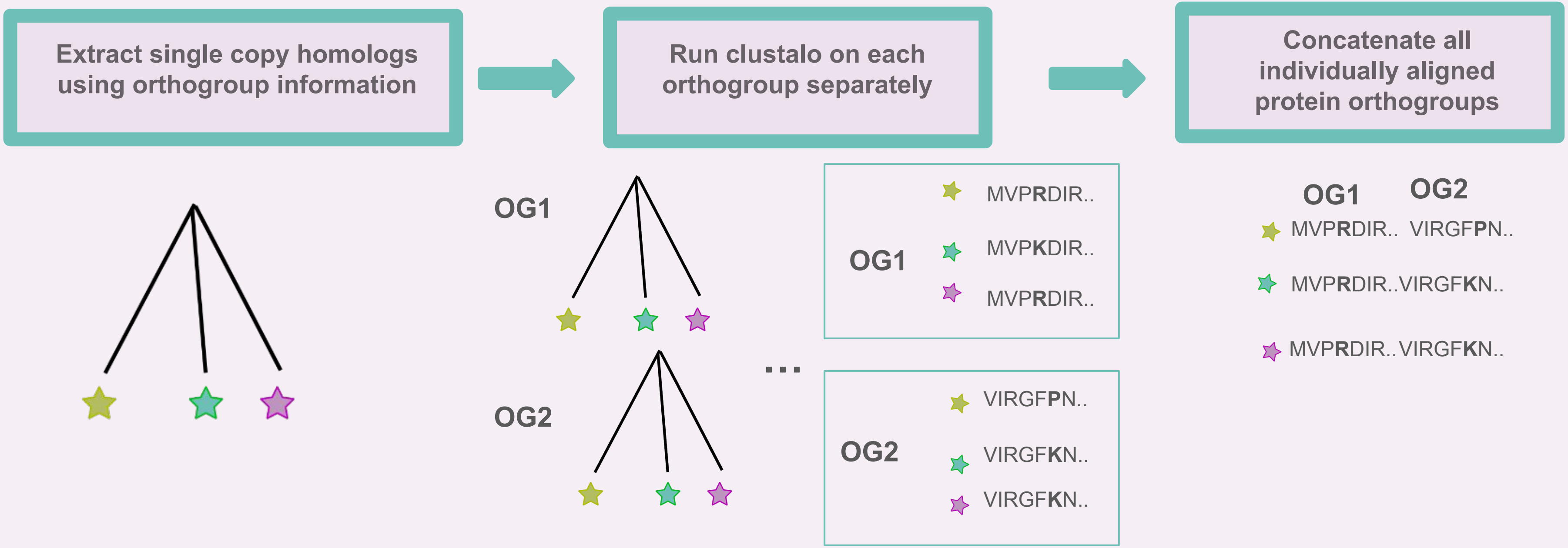
Concatenate all individually aligned protein orthogroups

MVP**R**DIR..
MVP**K**DIR..
MVP**R**DIR..

VIRGF**P**N..
VIRGF**K**N..
VIRGF**K**N..

Obtaining meaningful branch lengths

better understanding of genetic relatedness, especially when conflicting signals or complex evolutionary histories



Extract single copy homologs using orthogroup information

Run clustalo on each orthogroup separately

Concatenate all individually aligned protein orthogroups

OG1

★ MVP**R**DIR..
★ MVP**K**DIR..
★ MVP**R**DIR..

OG2

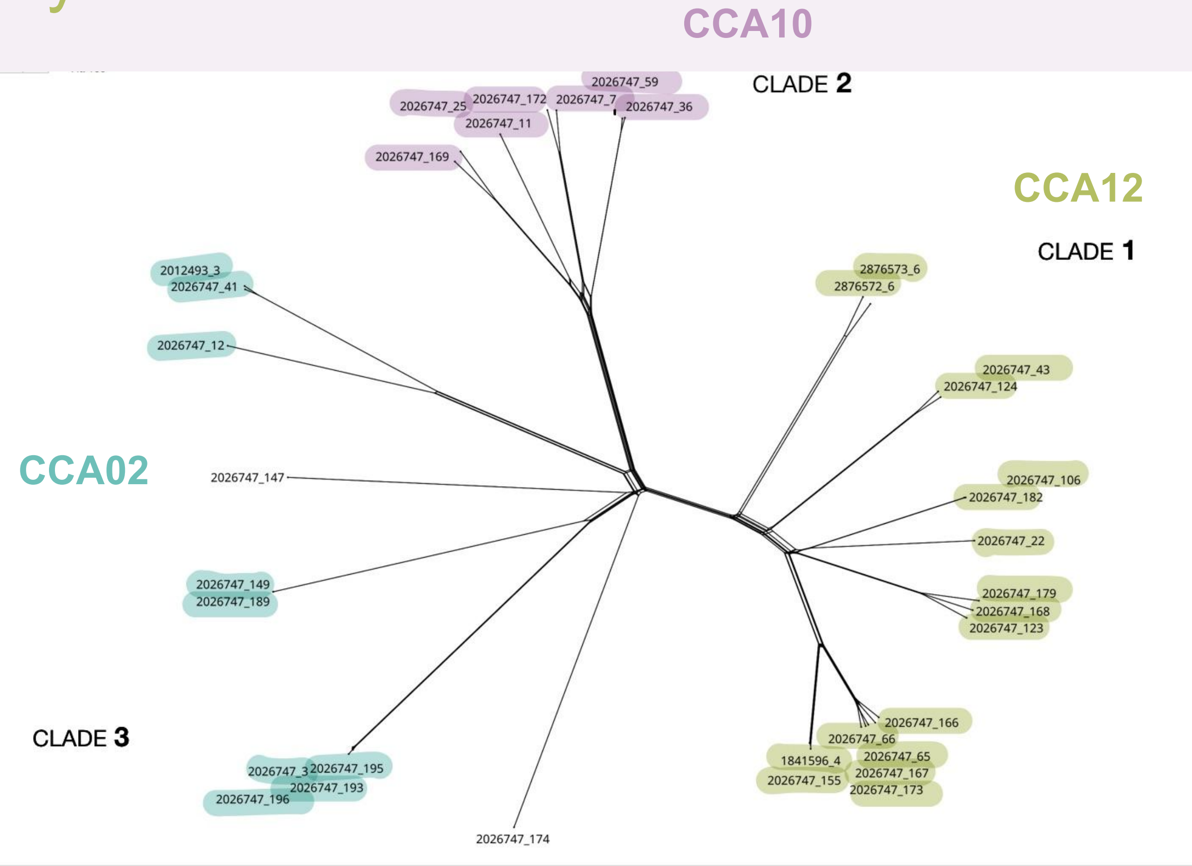
★ VIRGF**P**N..
★ VIRGF**K**N..
★ VIRGF**K**N..

OG1 **OG2**

★ MVP**R**DIR.. VIRGF**P**N..
★ MVP**R**DIR..VIRGF**K**N..
★ MVP**R**DIR..VIRGF**K**N..

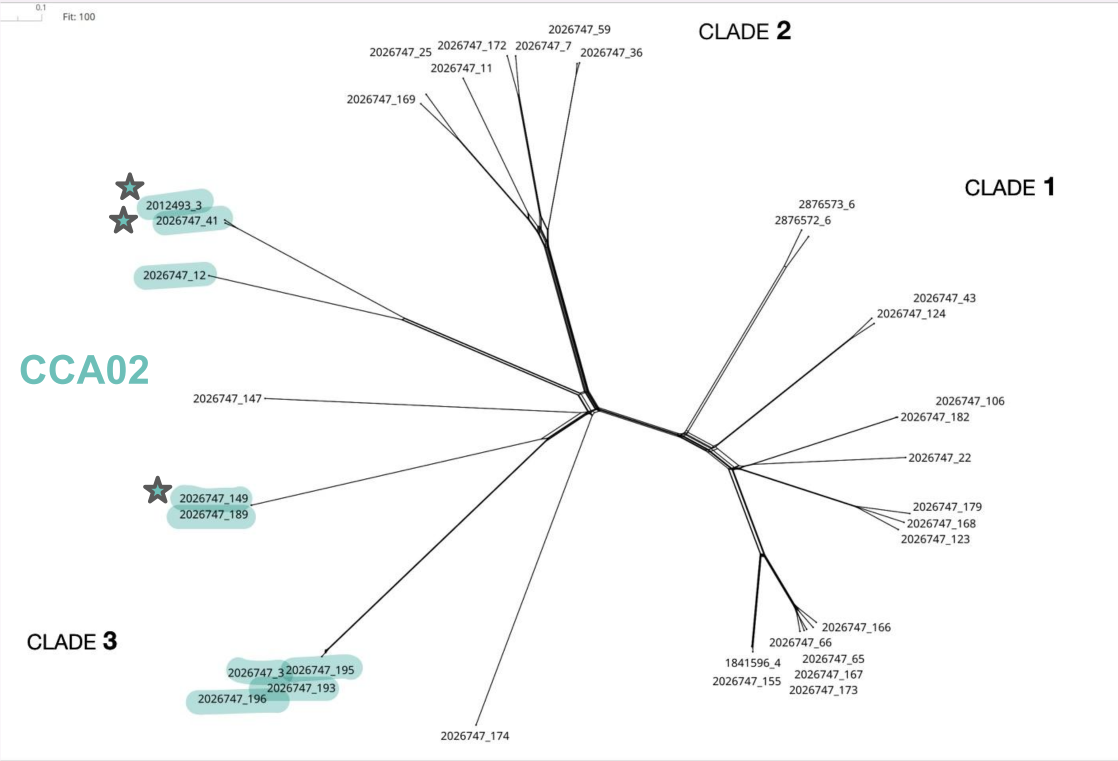
Molecular phylogeny construction

- Constructed phylogeny with meaningful branch lengths
- **But** this is not robust



Finding multiple CCA2 in one clade

- Distinct “CCA2 pairs” found in specific species



Conclusion

Heimdallarchaeota shows a complex evolutionary history of CCA-adding enzymes, where bacterial-like Class II CCA (CCA2) coexists with archaeal Class I CCA (CCA1)

Two explanations:

1. Early Archaeal Acquisition (Supports Two-Domain Hypothesis):

   CCA2 was **present in the archaeal ancestor of eukaryotes** before mitochondria, meaning eukaryotes **inherited** CCA2 rather than acquiring it from bacteria.

2. Later Bacterial HGT (Alternative Hypothesis):

   CCA2 in Heimdallarchaeota came **from bacteria after eukaryotic evolution**, meaning eukaryotes got CCA2 directly from bacteria (possibly during or after symbiosis)

Conclusion

Heimdallarchaeota shows a complex evolutionary history of CCA-adding enzymes, where bacterial-like Class II CCA (CCA2) coexists with archaeal Class I CCA (CCA1)

If Heimdallarchaeota is truly the closest archaeal relative of eukaryotes, then its molecular phylogeny should reflect early steps toward eukaryotic complexity

Conclusion

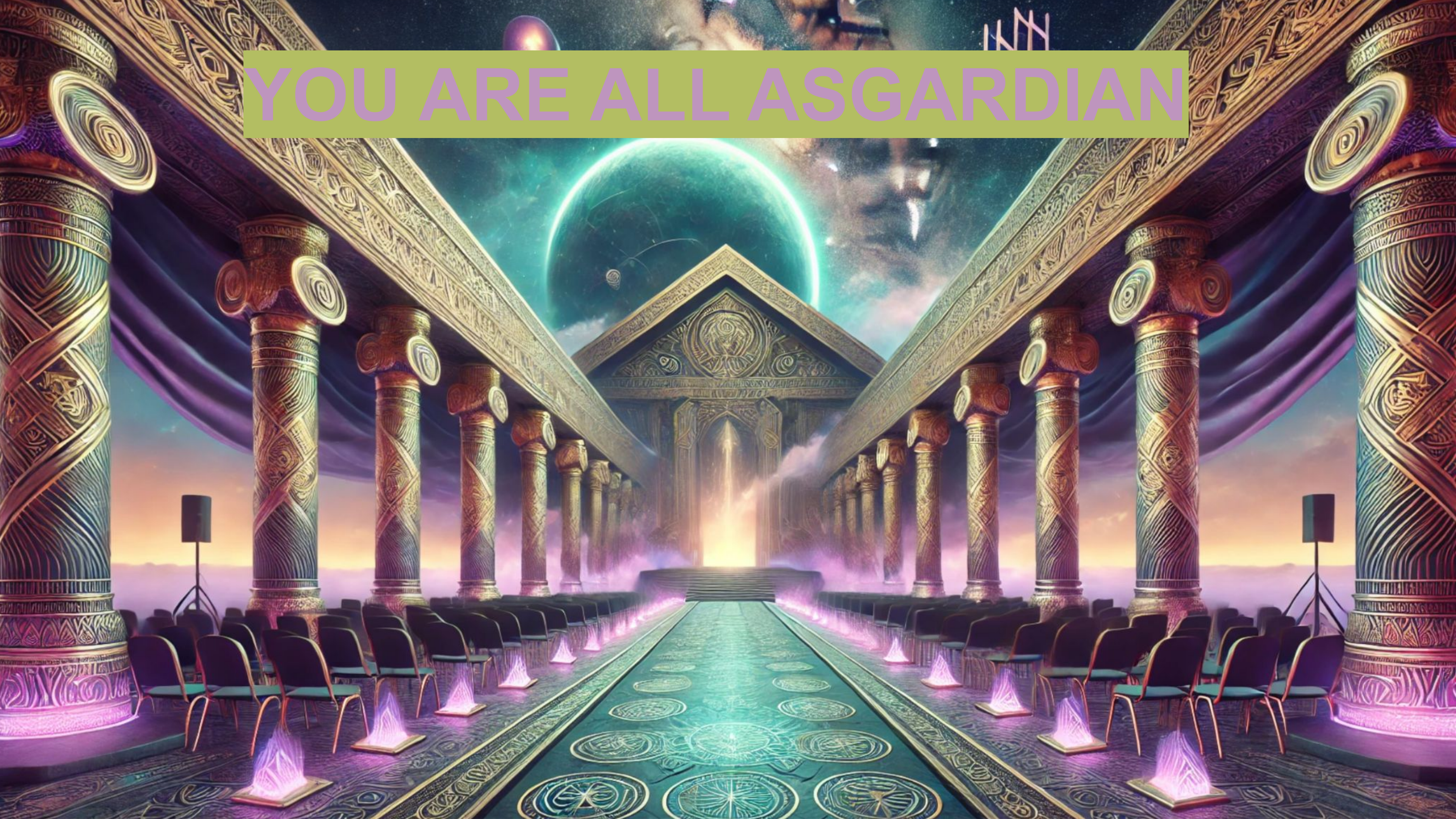
Heimdallarchaeota shows a complex evolutionary history of CCA-adding enzymes, where bacterial-like Class II CCA (CCA2) coexists with archaeal Class I CCA (CCA1)

If Heimdallarchaeota is truly the closest archaeal relative of eukaryotes, then its molecular phylogeny should reflect early steps toward eukaryotic complexity

Next steps

- Wet lab verification that these enzymes (CCA2) are, at least partially, active
- Check if CCA2 are they found in clades that also have more eukaryotic-like features?
- Ascertain whether CCA2's found are closely related to bacteria, are they from a recent HGT event

YOU ARE ALL ASGARDIAN



YOU ARE ALL ASGARDIAN

Thanks to

Prof. Dr. Sonja Prohaska

Biochemistry department collaborators
- Heike Betat
- Mario Mörl

Bierinformatics people <3

Are the CCA2 more related to bacteria?

- The clade 3 CCA2 are more related to CPR's and bacteria
- The clade 1 CCA2 have a distinct long branch
- Different origin?