# something about sRNA in archaea

## or: Processed Small RNAs in Archaea and BHB Elements

Sarah Berkemer

Bioinformatics Vienzig

# Archaea?


Google
scholar
Archaea?
Articles (include patents)
Case law
Stand on the shoulders of giants

# Archaea I

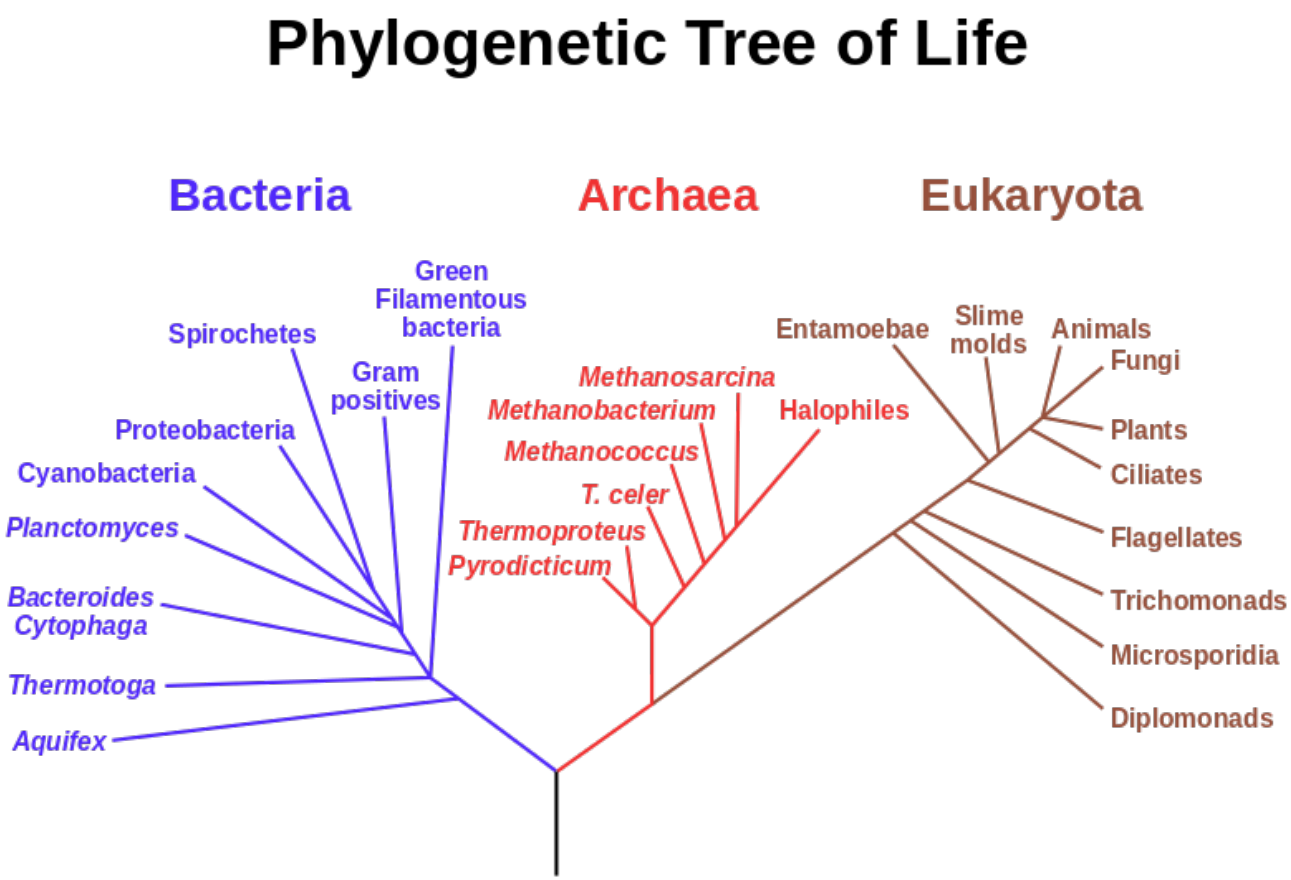



Figure: Phylogenetic Tree (commons.wikimedia.org)

# Archaea II

## Definition [Woese et al., 1990]

Domain Archaea [Greek adjective for ancient, primitive]:

- Cells prokaryotic
- Membrane lipids unique
- Ribosomes containing an archaeal type of rRNA

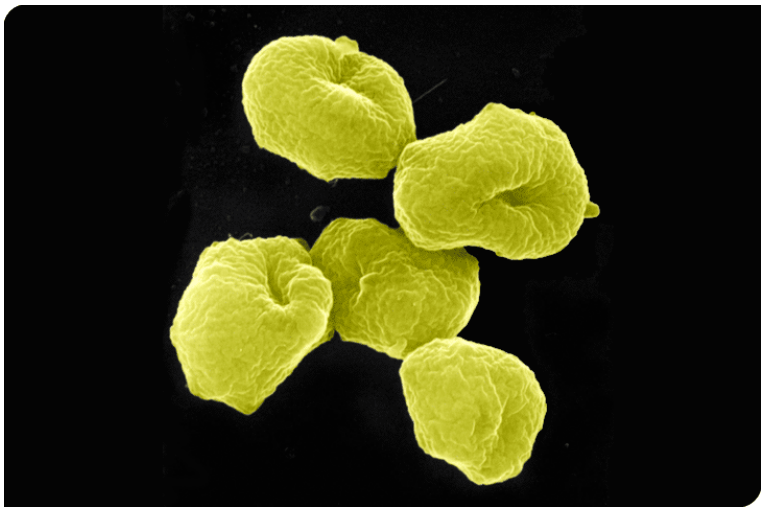


Figure: Archaeal species Sulfolobus (www.microbiologyonline.org.uk)

# Phylogenetic Tree of Life

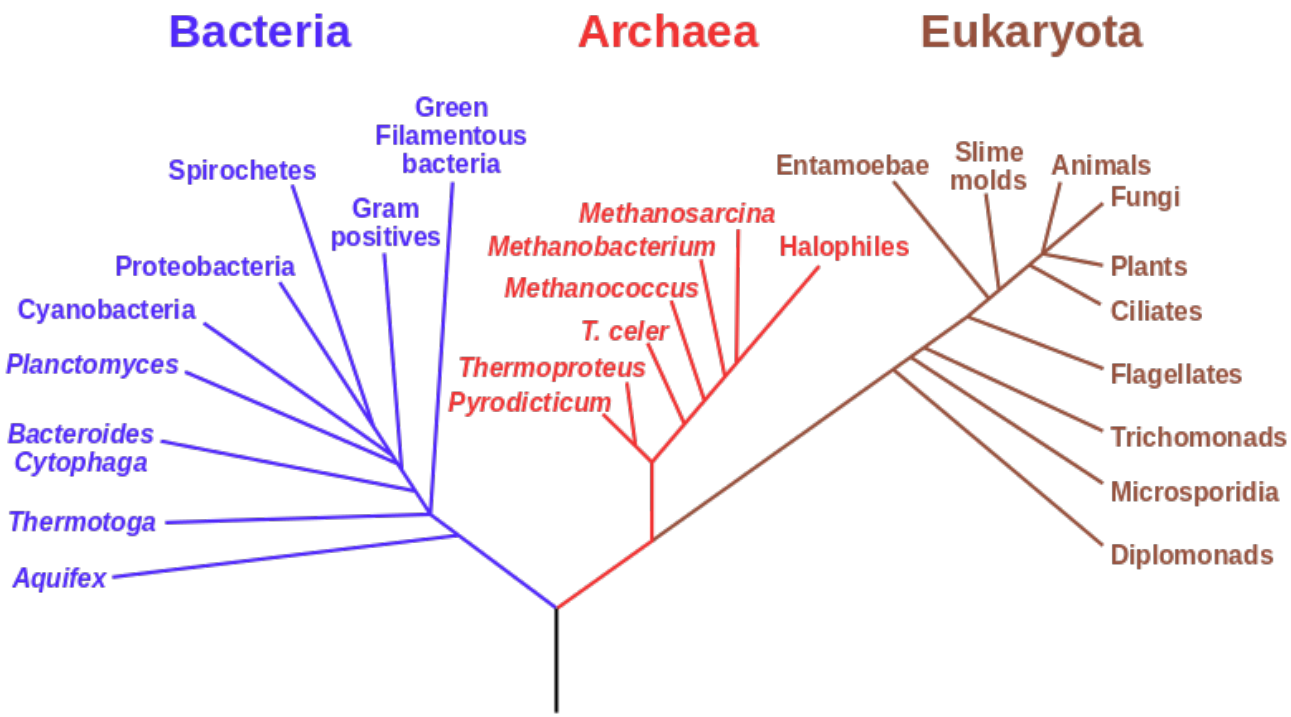
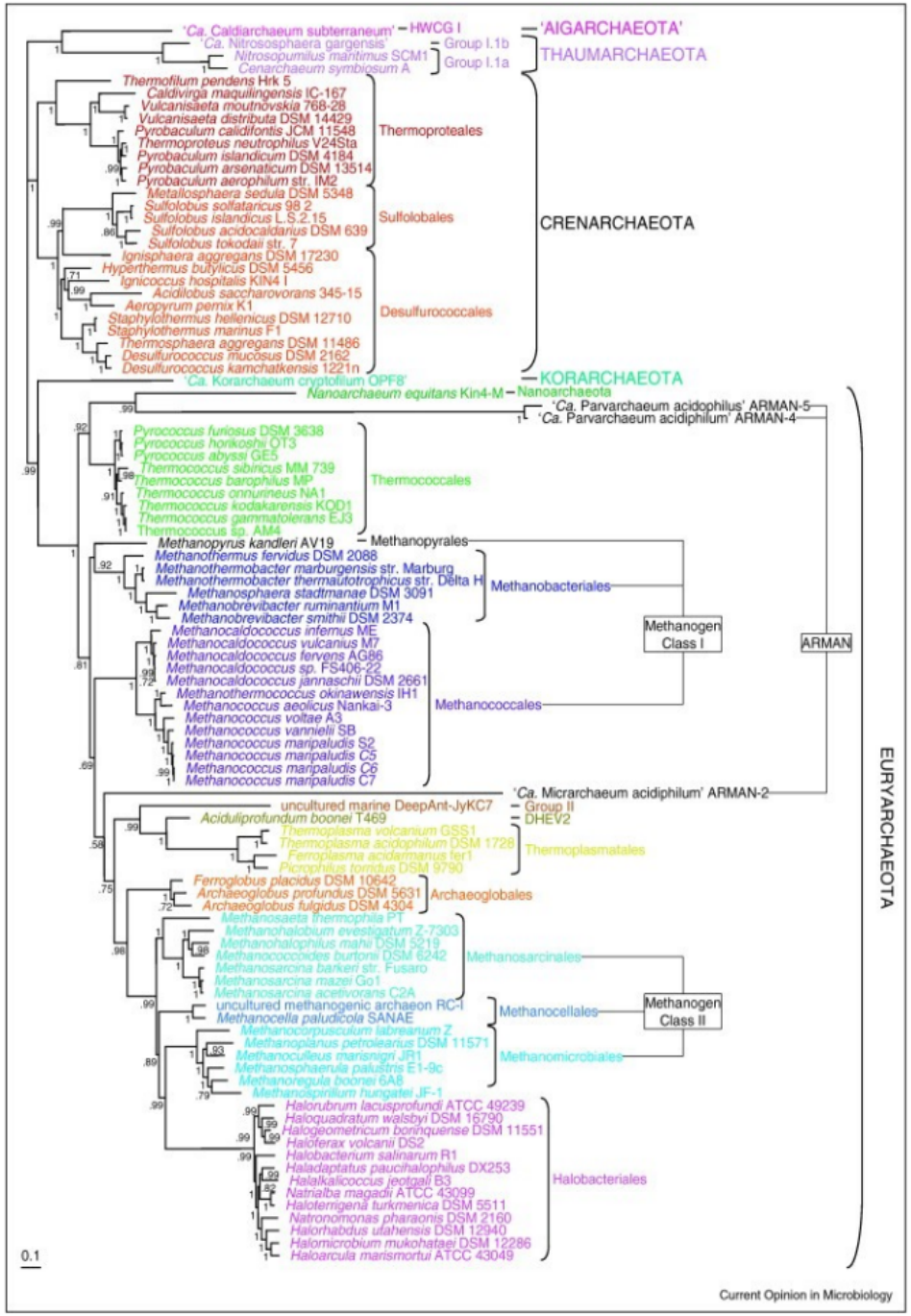



Figure: Phylogenetic Tree (commons.wikimedia.org)


'Ca. Caldiarchaeum subterraneum' – HWCG I
'AIGARCHAEOTA'
'Ca. Nitrososphaera gargensis' – Group I.1b
'Ca. Nitrosopumilus maritimus SCM1' – Group I.1a
THAUMARCHAEOTA
Cenarchaeum symbiosum A
Thermofilum pendens Hrk 5
Caldivirga maquilingensis IC-167
Vulcanisaeta moutnovskia 768-28
Vulcanisaeta distributa DSM 14429
Pyrobaculum calidifontis JCM 11548
Thermoproteus neutrophilus V24Sta
Pyrobaculum islandicum DSM 4184
Pyrobaculum arsenaticum DSM 13514
Pyrobaculum aerophilum str. IM2
Thermoproteales
Metallosphaera sedula DSM 5348
Sulfolobus solfataricus 98 2
Sulfolobus islandicus L.S.2.15
Sulfolobus acidocaldarius DSM 639
Sulfolobus tokodaii str. 7
Sulfolobales
CRENARCHAEOTA
Ignisphaera aggregans DSM 17230
Hyperthermus butylicus DSM 5456
Ignicoccus hospitalis KIN4 I
Acidilobus saccharovorans 345-15
Aeropyrum pernix K1
Staphylothermus hellenicus DSM 12710
Staphylothermus marinus F1
Thermosphaera aggregans DSM 11486
Desulfurococcus mucosus DSM 2162
Desulfurococcus kamchatkensis 1221n
Desulfurococcales
'Ca. Korarchaeum cryptofilum OPF8' – KORARCHAEOTA
Nanoarchaeum equitans Kin4-M – Nanoarchaeota
'Ca. Parvarchaeum acidophilus' ARMAN-5
'Ca. Parvarchaeum acidiphilum' ARMAN-4
Pyrococcus furiosus DSM 3638
Pyrococcus horikoshii OT3
Pyrococcus abyssi GE5
Thermococcus sibiricus MM 739
Thermococcus barophilus MP
Thermococcus onnurineus NA1
Thermococcus kodakarensis KOD1
Thermococcus gammatolerans EJ3
Thermococcus sp. AM4
Thermococcales
Methanopyrus kandleri AV19 – Methanopyrales
Methanothermus fervidus DSM 2088
Methanothermobacter marburgensis str. Marburg
Methanothermobacter thermautotrophicus str. Delta H
Methanosphaera stadtmanae DSM 3091
Methanobrevibacter ruminantium M1
Methanobrevibacter smithii DSM 2374
Methanobacteriales
Methanocaldococcus infernus ME
Methanocaldococcus vulcanius M7
Methanocaldococcus fervens AG86
Methanocaldococcus sp. FS406-22
Methanocaldococcus jannaschii DSM 2661
Methanothermococcus okinawensis IH1
Methanococcus aeolicus Nankai-3
Methanococcus voltae A3
Methanococcus vannielii SB
Methanococcus maripaludis S2
Methanococcus maripaludis C5
Methanococcus maripaludis C6
Methanococcus maripaludis C7
Methanococcales
Methanogen Class I
ARMAN
'Ca. Micrarchaeum acidiphilum' ARMAN-2
uncultured marine DeepAnt-JyKC7 – Group II
Aciduliprofundum boonei T469 – DHEV2
Thermoplasma volcanium GSS1
Thermoplasma acidophilum DSM 1728
Ferroplasma acidarmanus fer1
Picrophilus torridus DSM 9790
Thermoplasmatales
Ferroglobus placidus DSM 10642
Archaeoglobus profundus DSM 5631
Archaeoglobus fulgidus DSM 4304
Archaeoglobales
Methanosaeta thermophila PT
Methanohalobium evestigatum Z-7303
Methanohalophilus mahii DSM 5219
Methanococcoides burtonii DSM 6242
Methanosarcina barkeri str. Fusaro
Methanosarcina mazei Go1
Methanosarcina acetivorans C2A
Methanosarcinales
uncultured methanogenic archaeon RC-I
Methanocella paludicola SANAE
Methanocellales
Methanogen Class II
Methanocorpusculum labreanum Z
Methanoplanus petrolearius DSM 11571
Methanoculleus marisnigri JR1
Methanosphaerula palustris E1-9c
Methanoregula boonei 6A8
Methanolacinia petrolearium IF-1
Methanomicrobiales
Halorubrum lacusprofundi ATCC 49239
Haloquadratum walsbyi DSM 16790
Halogeometricum borinquense DSM 11551
Haloferax volcanii DS2
Halobacterium salinarum R1
Haladaptatus paucihalophilus DX253
Halalkalicoccus jeotgali B3
Natrialba magadii ATCC 43099
Halterrigena turkmenica DSM 5511
Natronomonas pharaonis DSM 2160
Halorhabdus utahensis DSM 12940
Halomicrobium mukohataei DSM 12286
Haloarcula marismortui ATCC 43049
Halobacteriales
EURYARCHAEOTA
0.1
Current Opinion in Microbiology

# 3 domains vs. 2 domains of life

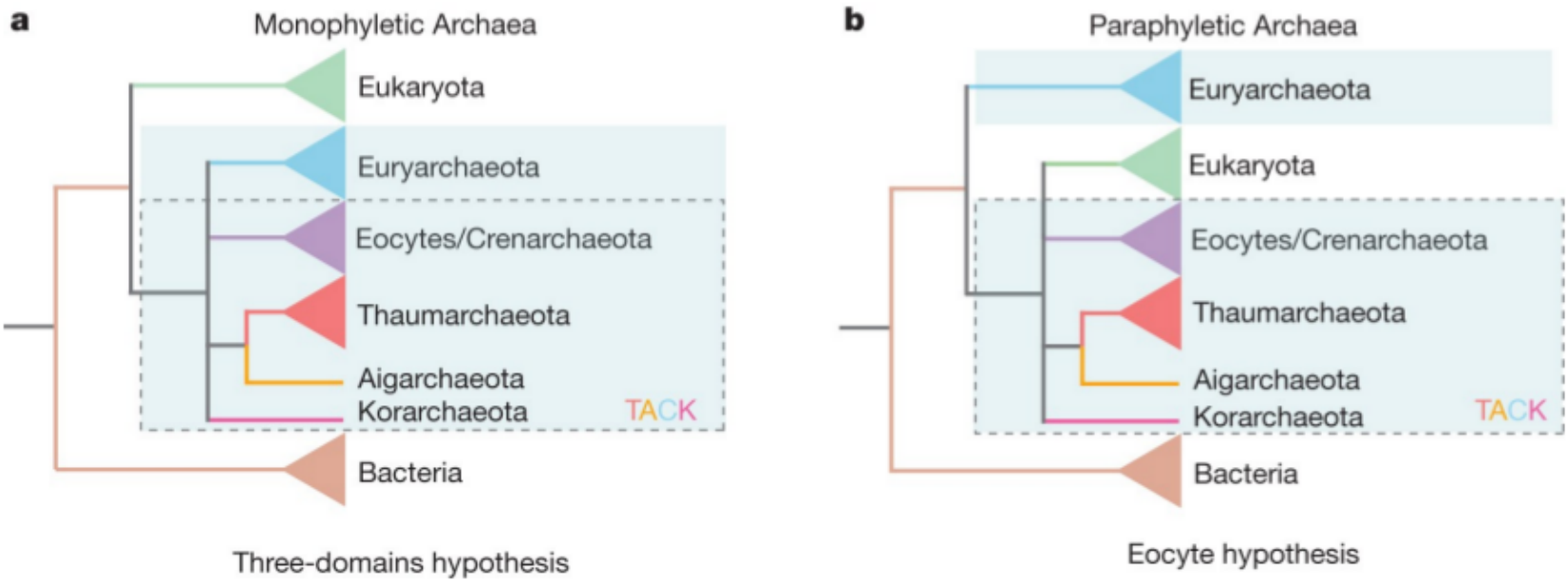
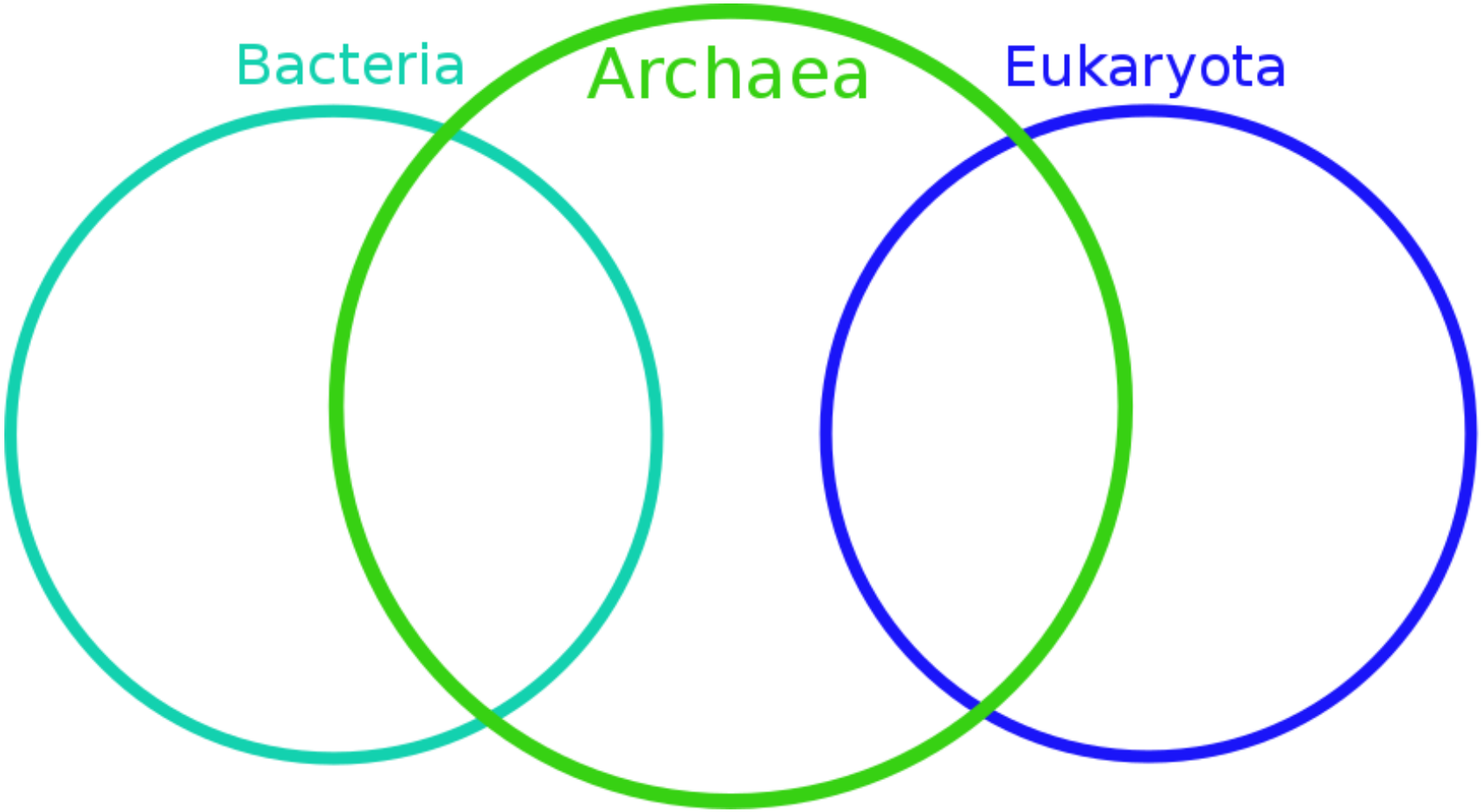
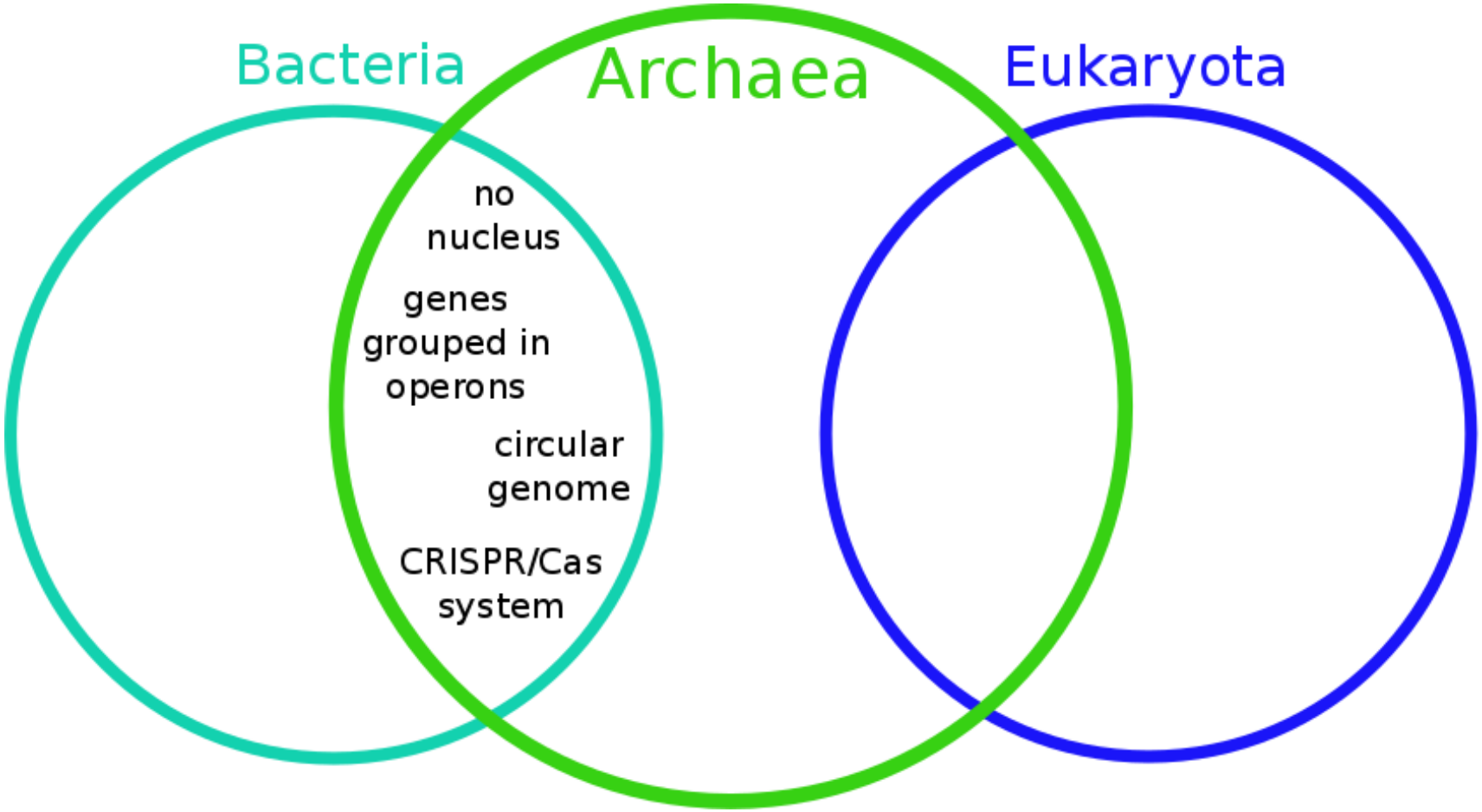



Figure: Possible 'Tree of life' [Williams et al., 2013]

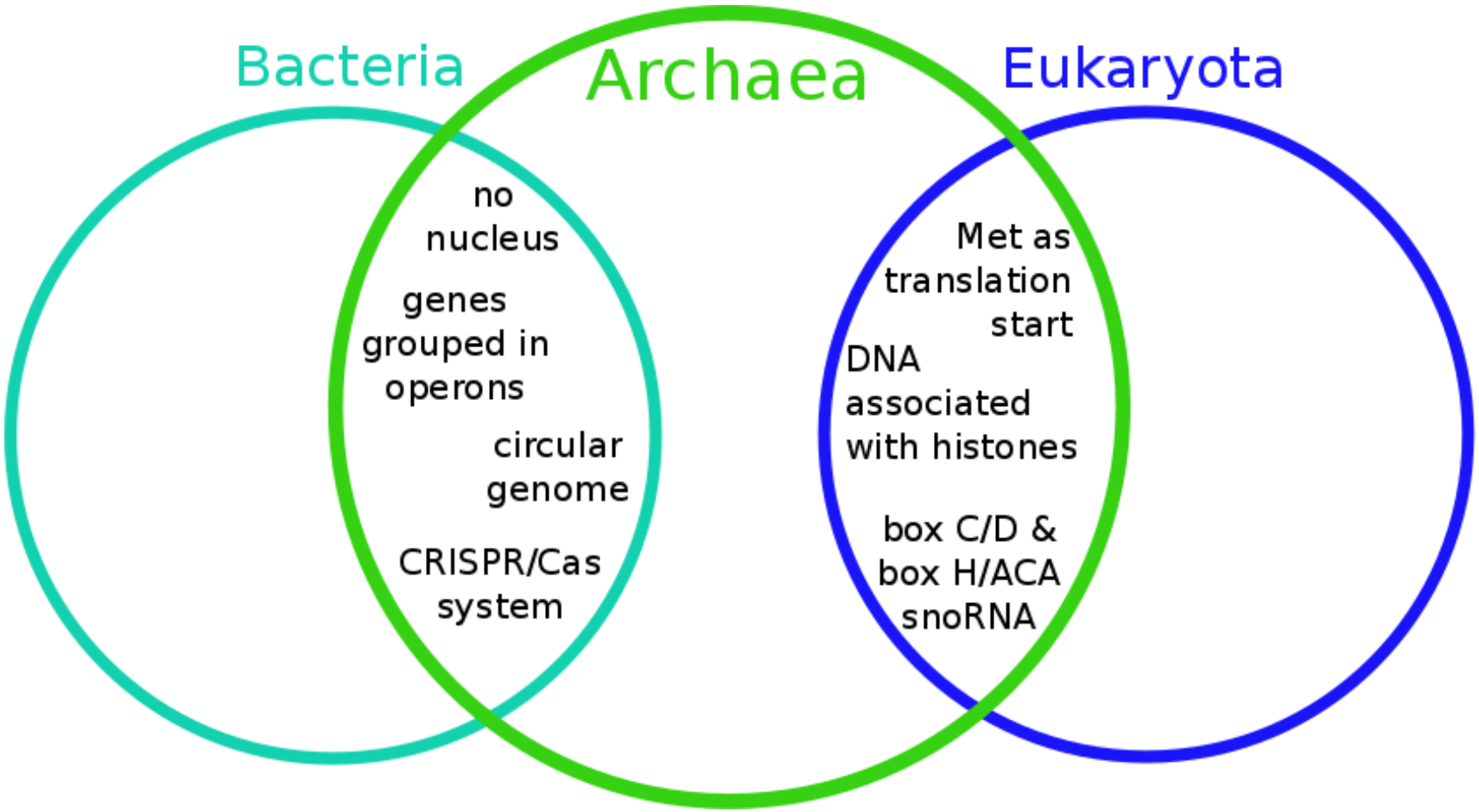
# Archaea III



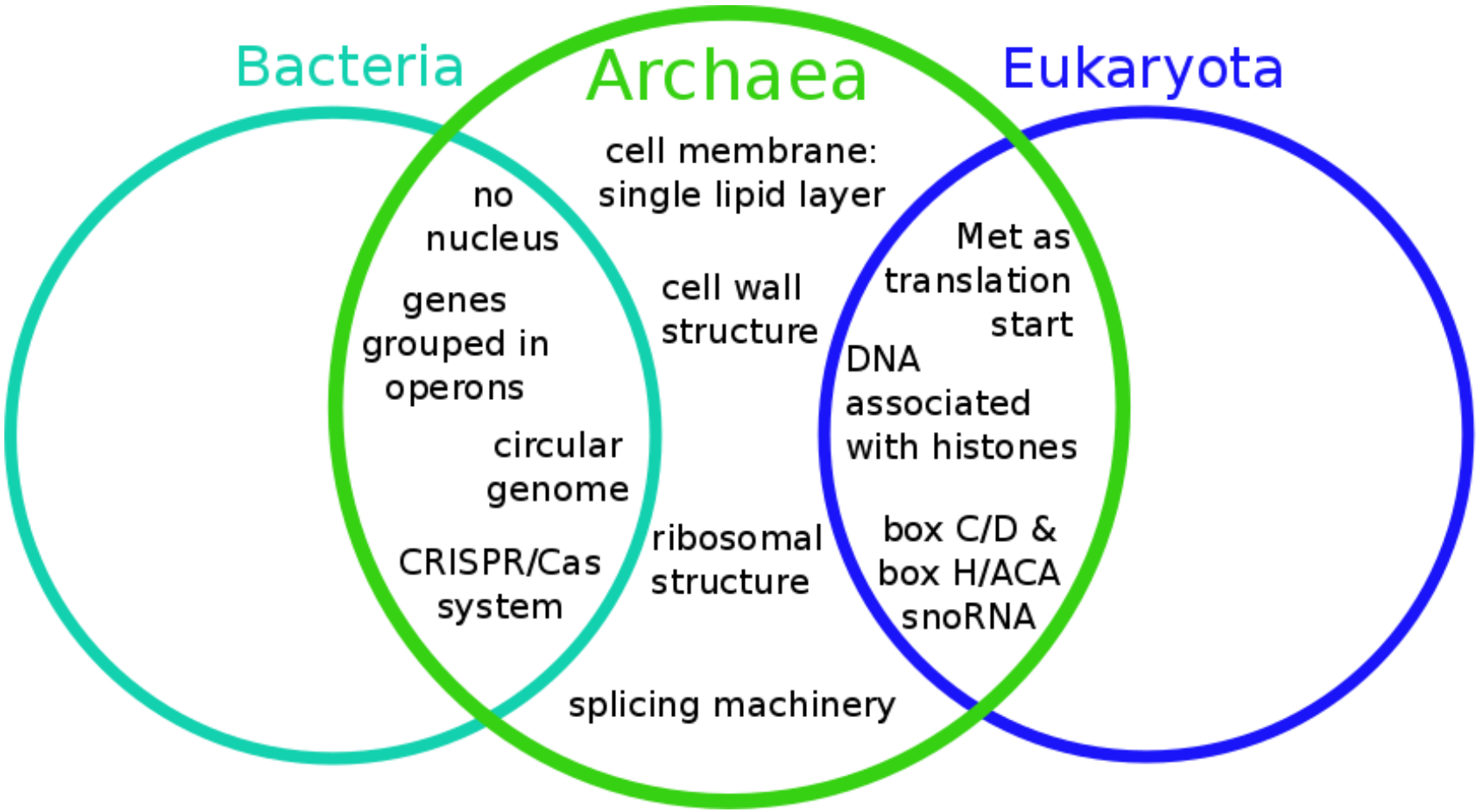
# Archaea III


Bacteria
Archaea
Eukaryota
no nucleus
genes grouped in operons
circular genome
CRISPR/Cas system

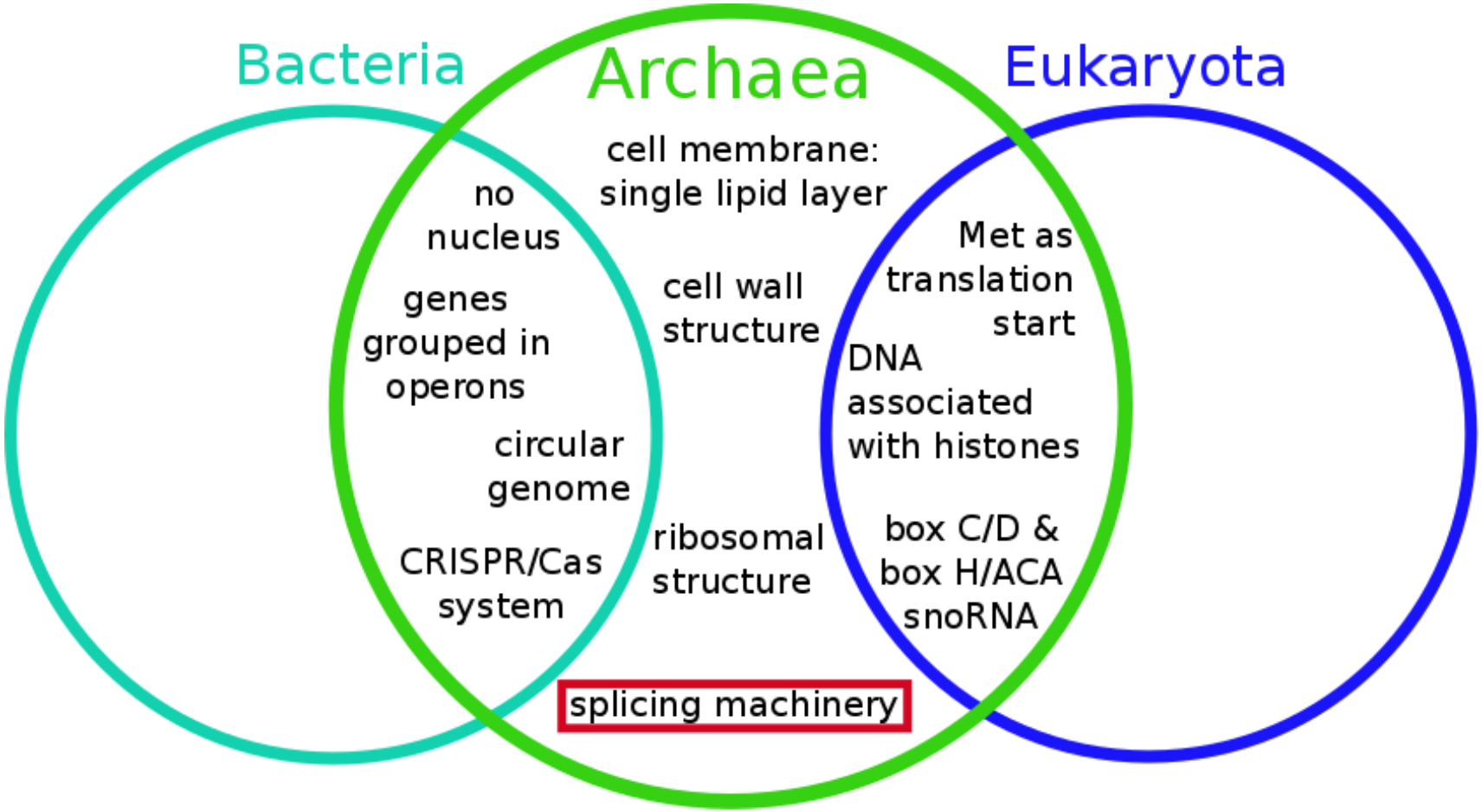
# Archaea III


Bacteria
Archaea
Eukaryota
no nucleus
genes grouped in operons
circular genome
CRISPR/Cas system
Met as translation start
DNA associated with histones
box C/D & box H/ACA snoRNA

# Archaea III


Bacteria
Archaea
Eukaryota
cell membrane:
single lipid layer
no
nucleus
genes
grouped in
operons
circular
genome
CRISPR/Cas
system
cell wall
structure
ribosomal
structure
splicing machinery
Met as
translation
start
DNA
associated
with histones
box C/D &
box H/ACA
snoRNA

# Archaea III


Bacteria
Archaea
Eukaryota
cell membrane:
single lipid layer
no
nucleus
genes
grouped in
operons
circular
genome
CRISPR/Cas
system
cell wall
structure
ribosomal
structure
Met as
translation
start
DNA
associated
with histones
box C/D &
box H/ACA
snoRNA
splicing machinery

# Splicing mechanism

## Splicing

- Eukarya: splicing with spliceosome
- Bacteria: no splicing
- Archaea: splicing with at least two protein enzymes (specific endonuclease and ligase) - BHB (bulge-helix-bulge) element as recognition

# Splicing mechanism

## Splicing

- Eukarya: splicing with spliceosome
- Bacteria: no splicing
- Archaea: splicing with at least two protein enzymes (specific endonuclease and ligase) - BHB (bulge-helix-bulge) element as recognition

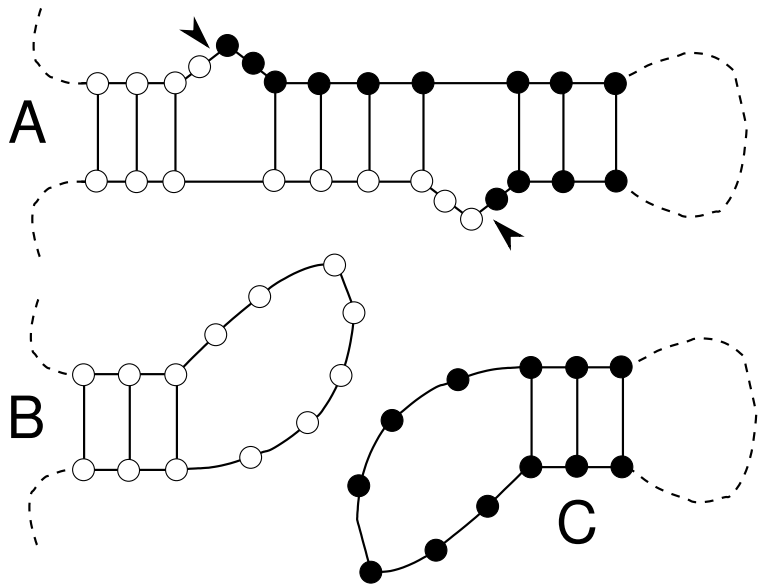



Figure: Scheme for a BHB element

# BHB structural element

```
GCGGGCTTTTAACCGCG...loop...TGCGGGAAGCAACCCG
hhhhhHHHHHBBBggggg...loop...gggggHHHHHBBBhhhhh
(((((((((...(((((...loop...)))))))))...)))))
```

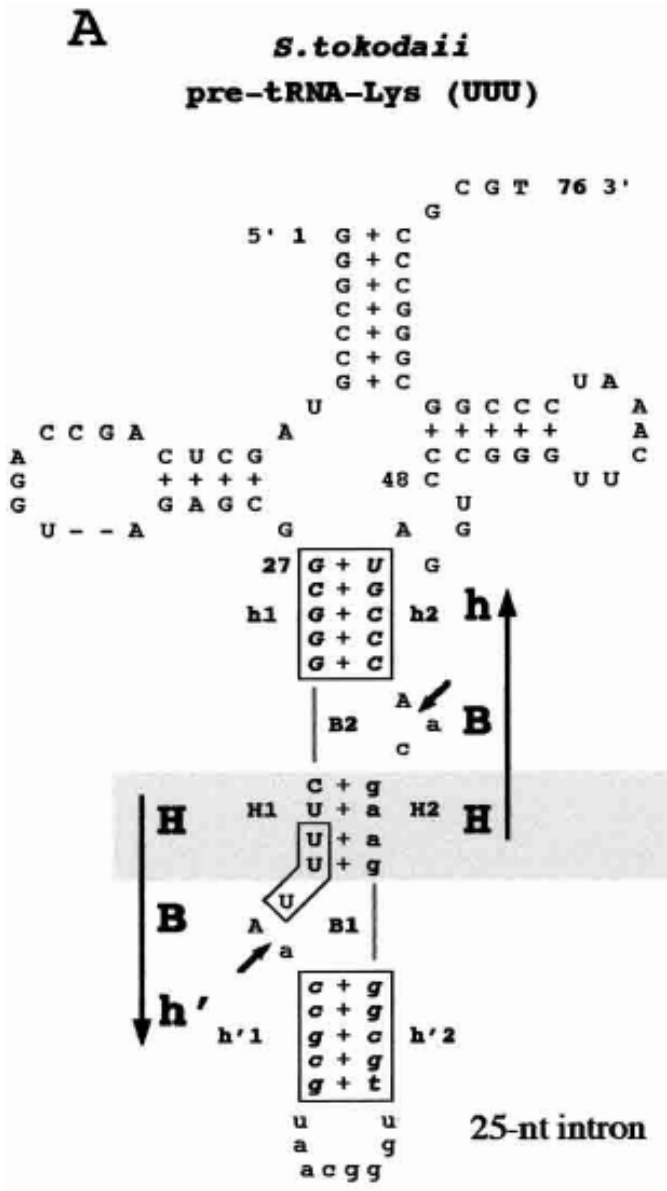
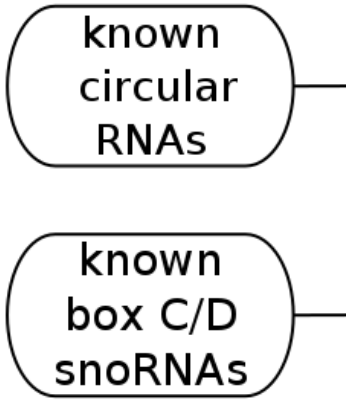
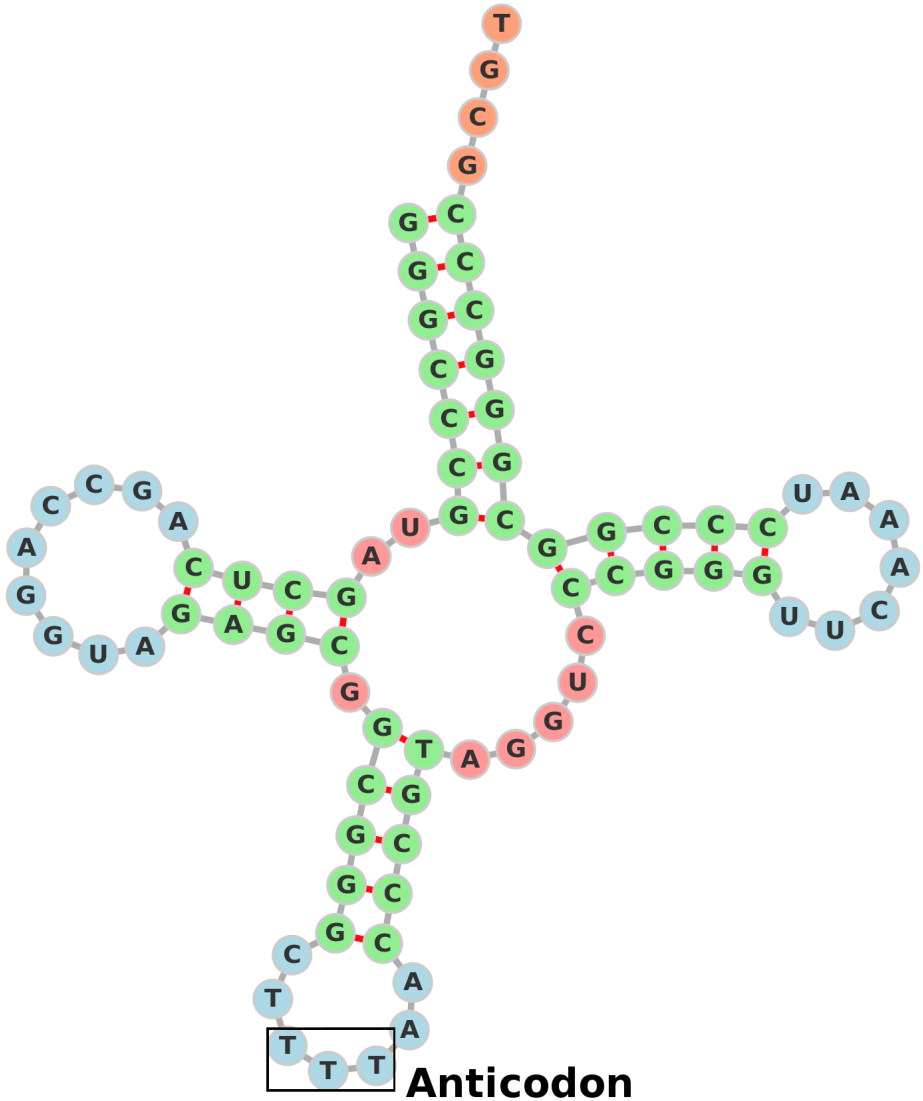



Figure: BHB element, sequence and secondary structure in tRNA Lys of S.tokodaii [Marck & Grosjean, 2003]

# Workflow



# Candidate cRNAs



# tRNA introns

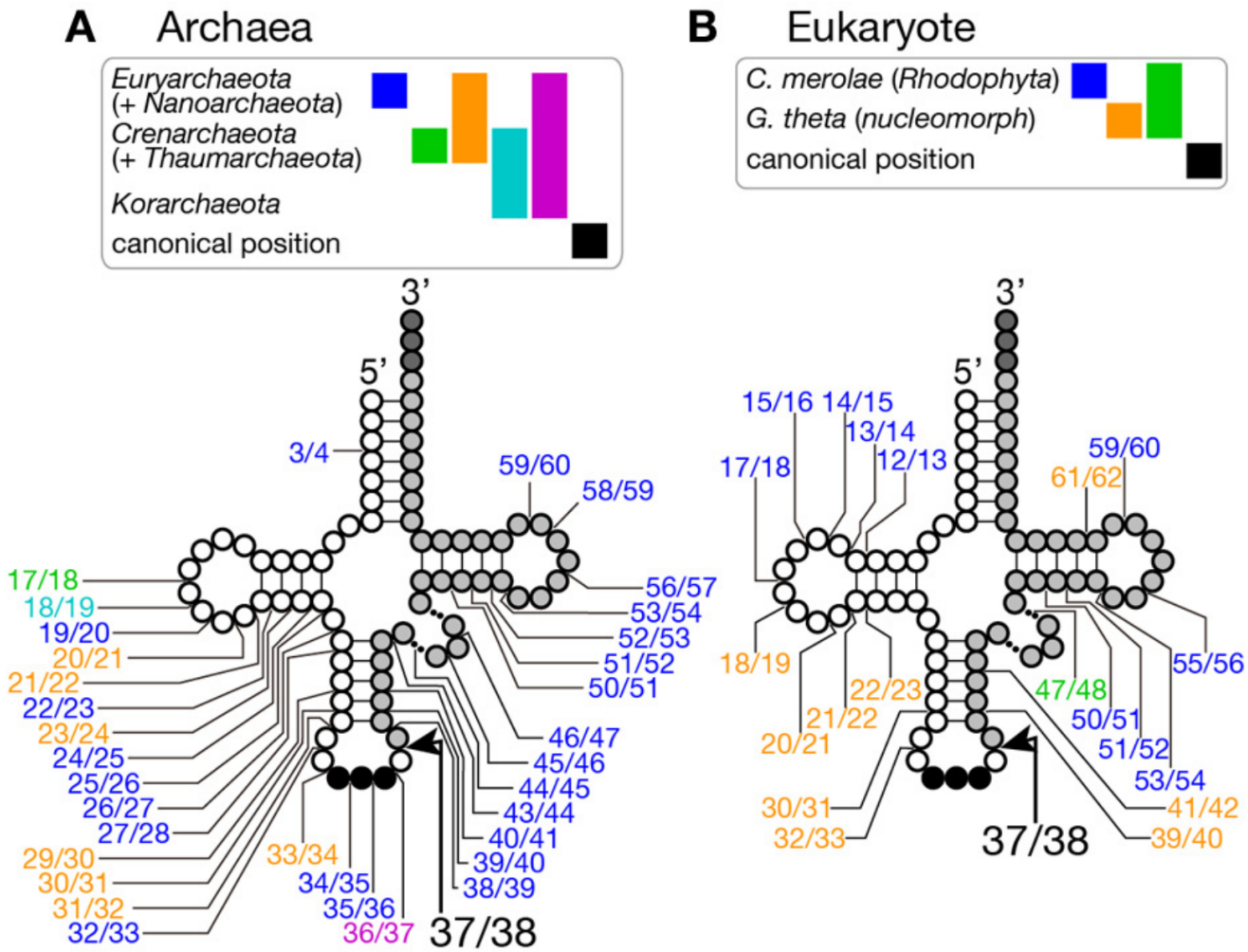
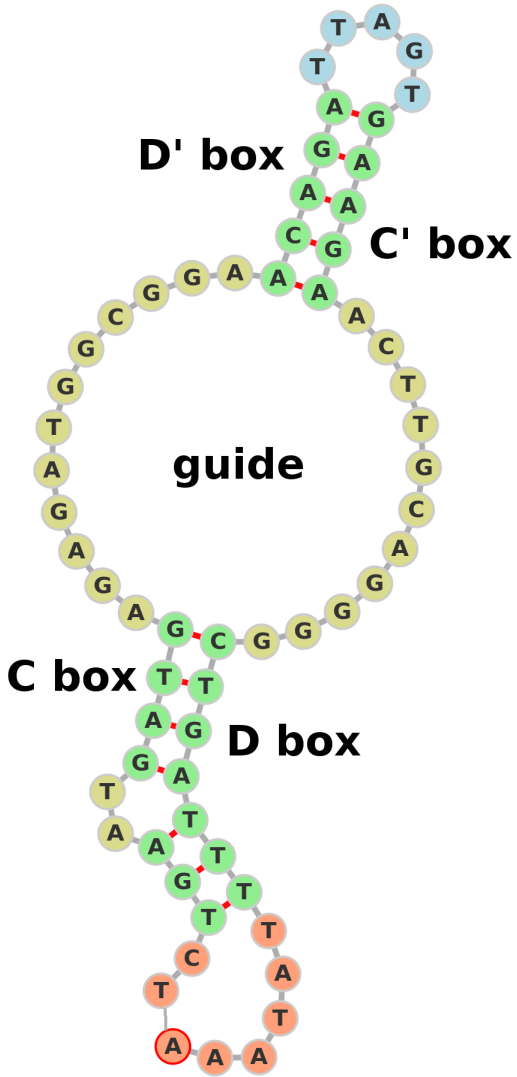
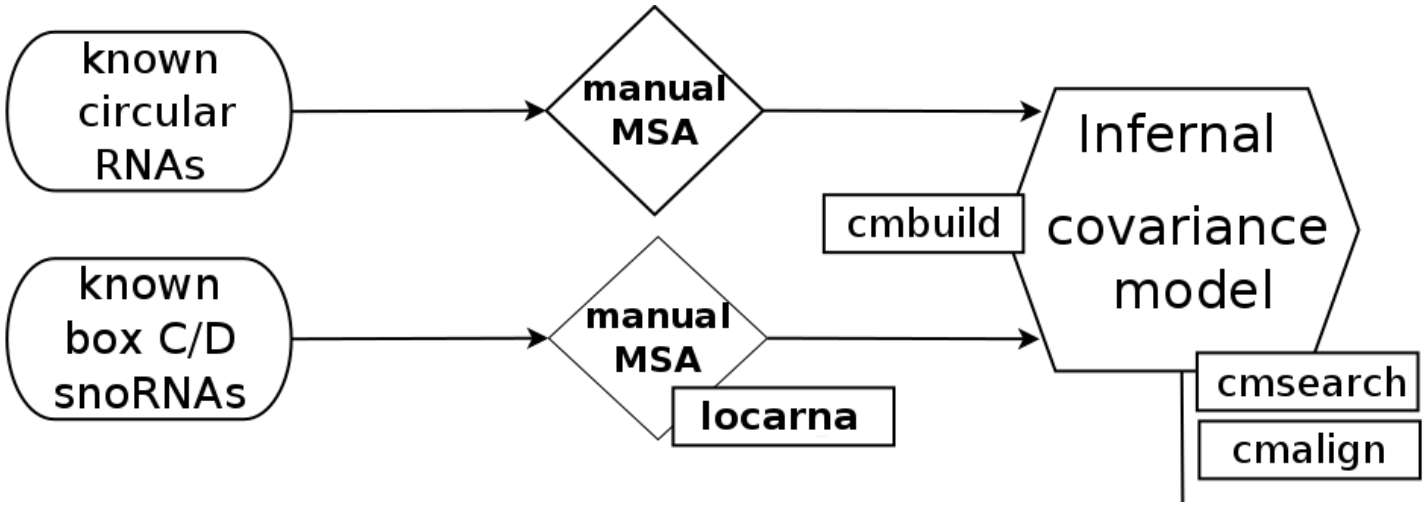


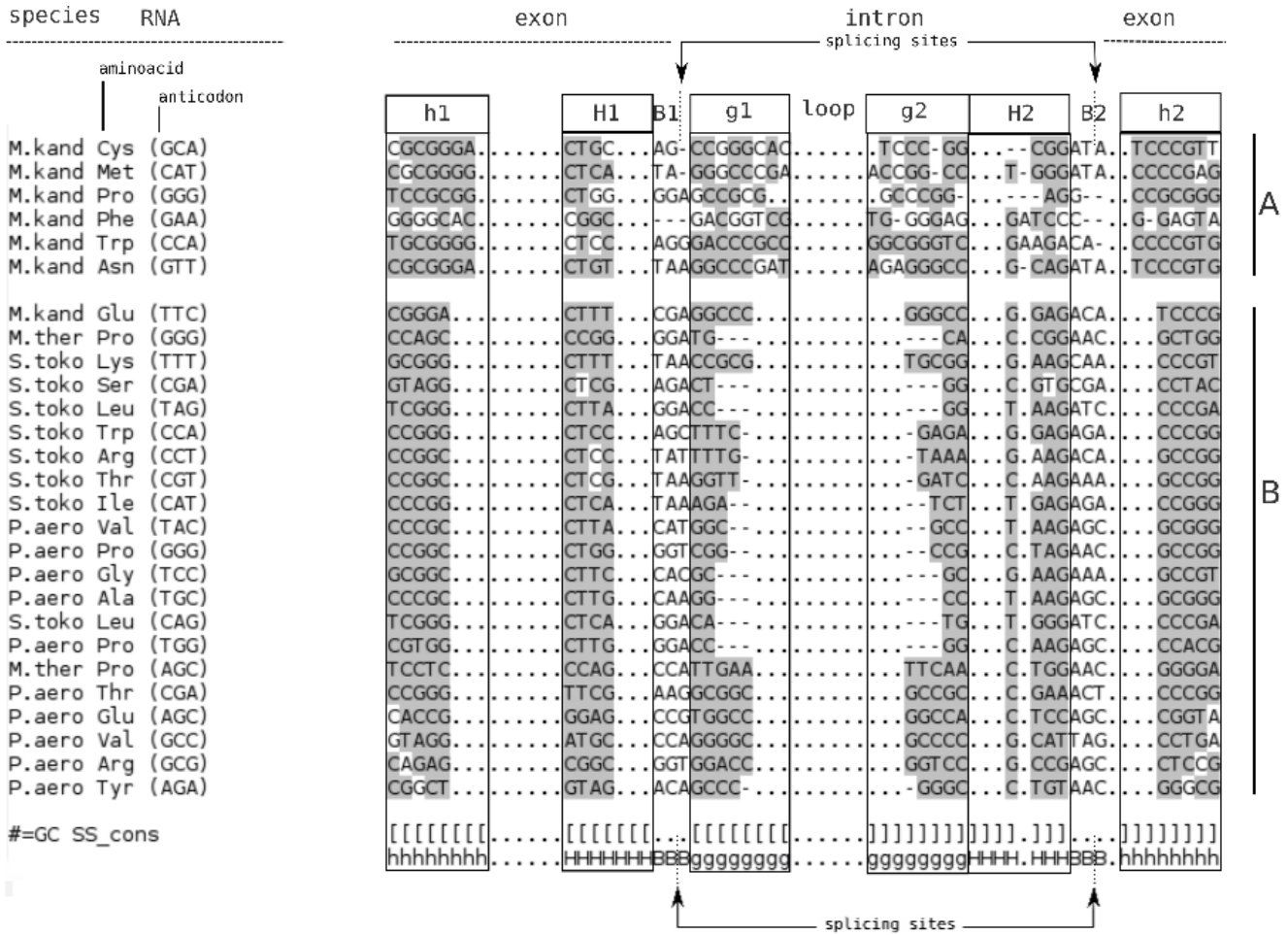

Figure: Splice sites in tRNAs [Yoshihisa, 2014]

# Candidate cRNAs

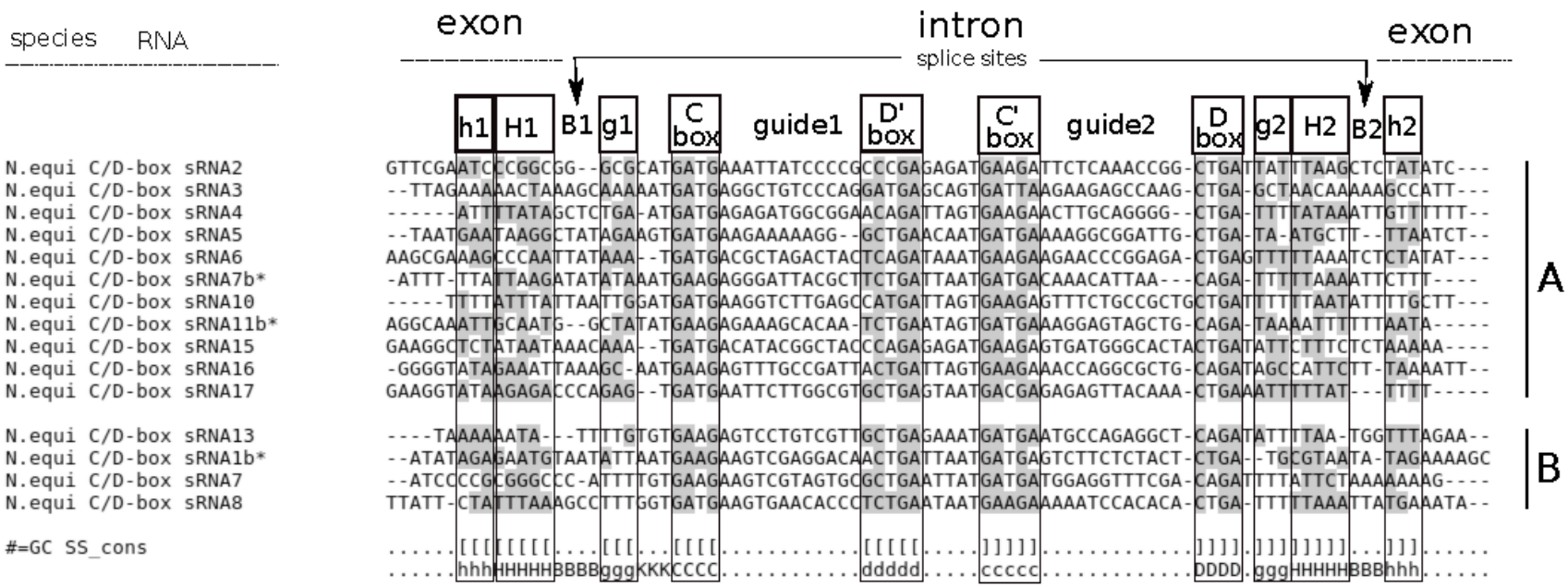



known
circular
RNAs
manual
MSA
known
box C/D
snoRNAs
manual
MSA
locarna
Infernal
cmbuild
covariance
model
cmsearch
cmalign

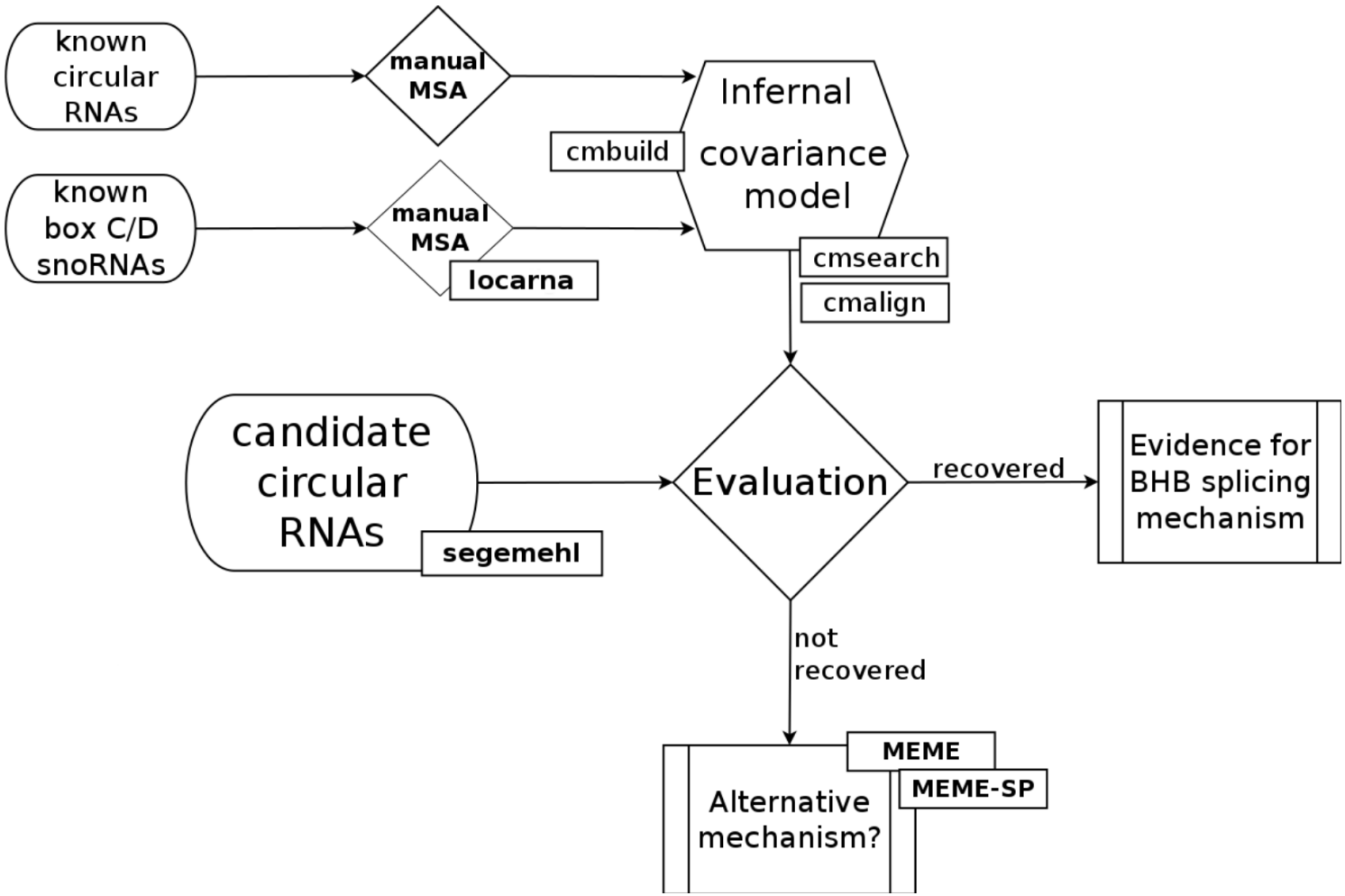
# MSA


species
RNA
aminoacid
anticodon
M.kand Cys (GCA)
M.kand Met (CAT)
M.kand Pro (GGG)
M.kand Phe (GAA)
M.kand Trp (CCA)
M.kand Asn (GTT)
M.kand Glu (TTC)
M.ther Pro (GGG)
S.toko Lys (TTT)
S.toko Ser (CGA)
S.toko Leu (TAG)
S.toko Trp (CCA)
S.toko Arg (CCT)
S.toko Thr (CGT)
S.toko Ile (CAT)
P.aero Val (TAC)
P.aero Pro (GGG)
P.aero Gly (TCC)
P.aero Ala (TGC)
S.toko Leu (CAG)
P.aero Pro (TGG)
M.ther Pro (AGC)
P.aero Thr (CGA)
P.aero Glu (AGC)
P.aero Val (GCC)
P.aero Arg (GCG)
P.aero Tyr (AGA)
#=GC SS_cons
exon
intron
exon
splicing sites
h1
H1
B1
g1
loop
g2
H2
B2
h2
A
B
splicing sites

# MSA


species RNA
exon
intron
splice sites
exon
h1 H1 B1 g1 C box guide1 D' box C' box guide2 D box g2 H2 B2 h2
N.equi C/D-box sRNA2
N.equi C/D-box sRNA3
N.equi C/D-box sRNA4
N.equi C/D-box sRNA5
N.equi C/D-box sRNA6
N.equi C/D-box sRNA7b*
N.equi C/D-box sRNA10
N.equi C/D-box sRNA11b*
N.equi C/D-box sRNA15
N.equi C/D-box sRNA16
N.equi C/D-box sRNA17
N.equi C/D-box sRNA13
N.equi C/D-box sRNA1b*
N.equi C/D-box sRNA7
N.equi C/D-box sRNA8
#=GC SS_cons
A
B

# Workflow


known circular RNAs
manual MSA
known box C/D snoRNAs
manual MSA
locarna
cmbuild
Infernal covariance model
cmsearch
cmalign
candidate circular RNAs
segemehl
Evaluation
recovered
Evidence for BHB splicing mechanism
not recovered
MEME
MEME-SP
Alternative mechanism?

# Results

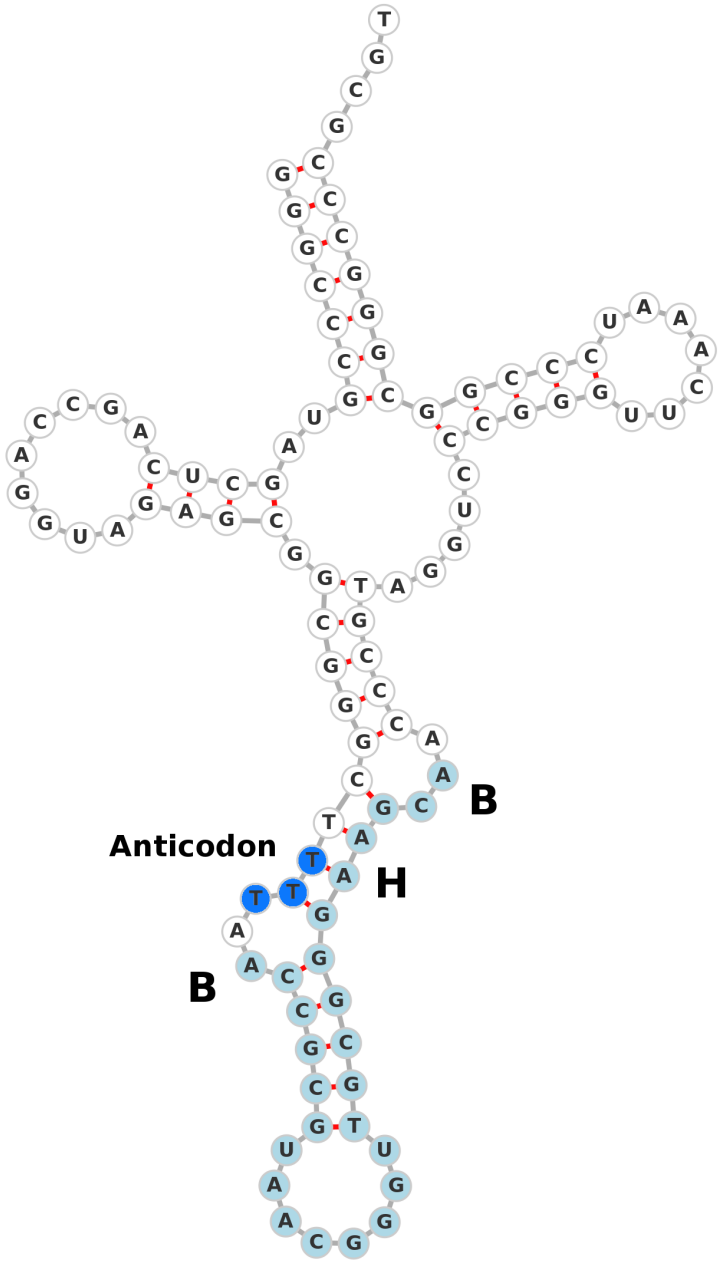
| Species | C/D | BHB | | no BHB | |
|---|---|---|---|---|---|
| *M. kandleri* | 126 | 112 | (80) | 14 | (8) |
| *N. equitans* | 11 | 9 | (6) | 2 | (1) |
| *S. solfataricus* | 20 | 7 | (3) | 13 | (5) |
| *S. acidocaldarius* | 24 | 22 | (16) | 2 | (2) |

# Results

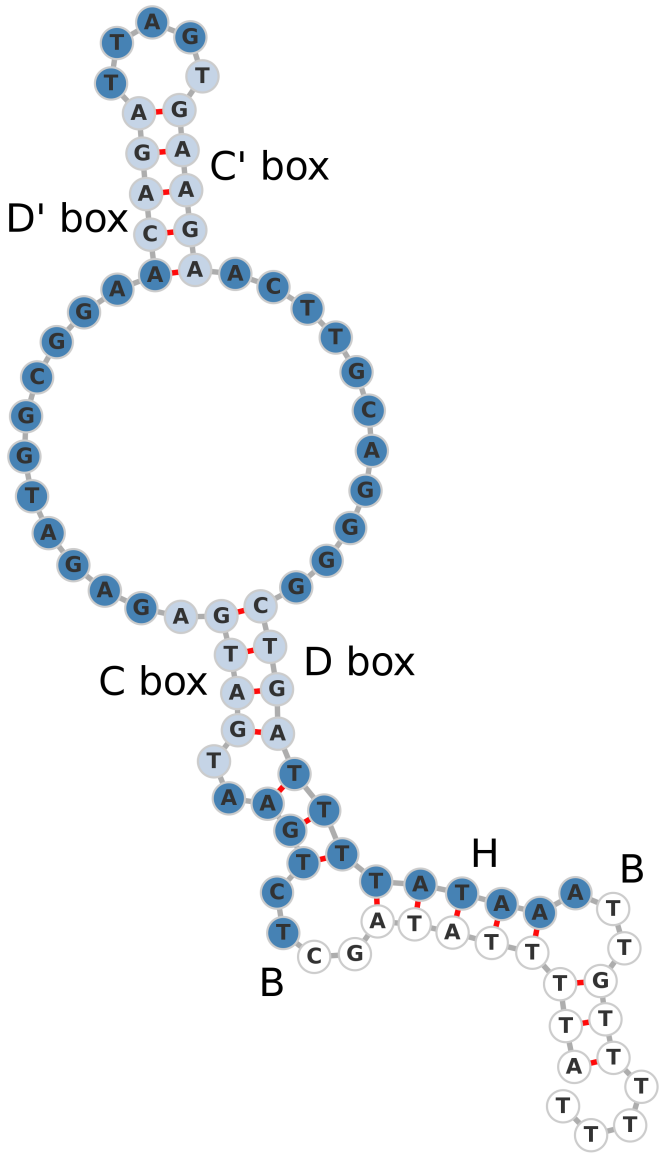
| Species | C/D | BHB | | no BHB | |
|---|---|---|---|---|---|
| *M. kandleri* | 126 | 112 | (80) | 14 | (8) |
| *N. equitans* | 11 | 9 | (6) | 2 | (1) |
| *S. solfataricus* | 20 | 7 | (3) | 13 | (5) |
| *S. acidocaldarius* | 24 | 22 | (16) | 2 | (2) |

| Species | splicesites | MSA1 | MSA2 |
|---|---|---|---|
| *M. kandleri* | 213 | 25 | 30 |
| *N. equitans* | 12 | 1 | 3 |
| *I. hospitalis* | 23 | 3 | 7 |
| *S. acidocaldarius* | 185 | 12 | 18 |
| *S. solfataricus* | 26 | 0 | 8 |

# tRNA



# C/D box

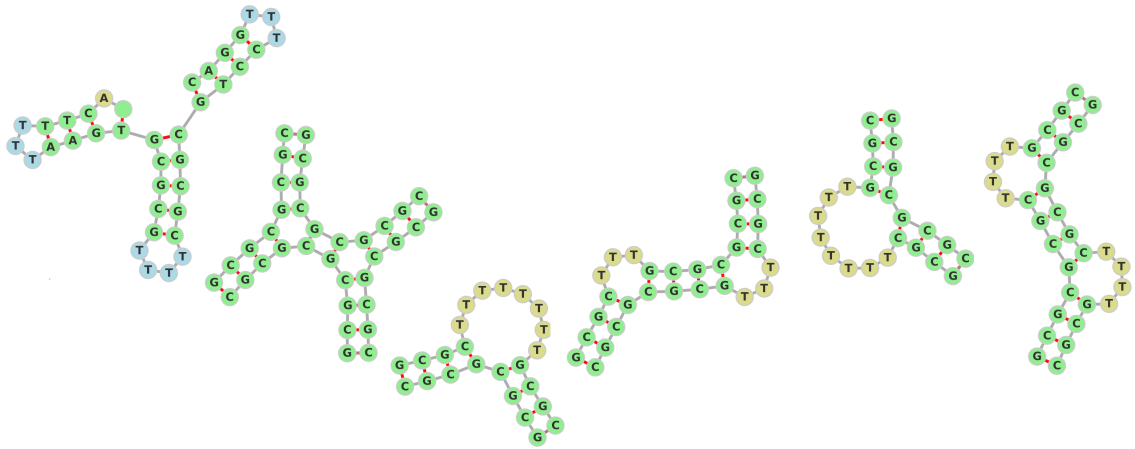


## Results

- combining forces and reducing search space helps
- most of known box C/D sRNAs flanked with BHB elements
- new candidates identified

## Outlook

- still circularized introns without BHB elements
- just known RNA classes can be tested: no generalized model



Thanks to:

- Peter Stadler
- Christian Höner zu Siederdissen
- Ivo Hofacker
- Fabian Amman
- Sebastian Will
- Peter Kerpedjiev
- ... and the TBI table soccer team

# References

## References

Brochier-Armanet, C., Forterre, P., & Gribaldo, S. (2011).

Phylogeny and evolution of the Archaea: one hundred genomes later.
*Current opinion in microbiology*, 14(3), 274–281.

Höner zu Siederdissen, C., Berkemer, S., Amman, F., Wintsche, A., Will, S., Prohaska, S. J., & Stadler, P. F. (2014).

Comparative Detection of Processed Small RNAs in Archaea.
*accepted: IWBBio, Granada.*

Marck, C. & Grosjean, H. (2003).

Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications.
*RNA*, 9.

Williams, T. A., Foster, P. G., Cox, C. J., & Embley, T. M. (2013).

An archaeal origin of eukaryotes supports only two primary domains of life.
*Nature*, 504(7479), 231–236.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990).

Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.
*PNAS*, 87(12).

Yoshihisa, T. (2014).

Handling tRNA introns, archaeal way and eukaryotic way.
*Front Genet.*, 5, 213.