'Bioinformatics of organ regeneration' is the collaborative project from @amrojasmendoza's and @fcasfer's lab. They have an open joint PhD position to work on organ regeneration, exploring a new model system. If you are enthusiastic about this subject send them an email!

Offering a 4Y PhD position to work on organ regeneration
a.rojas.m@csic.es
fcaster@upo.es

A **novel** MMR pathway in prokaryotes

Jesús Blázquez, CNB-CSIC (Madrid, ES)

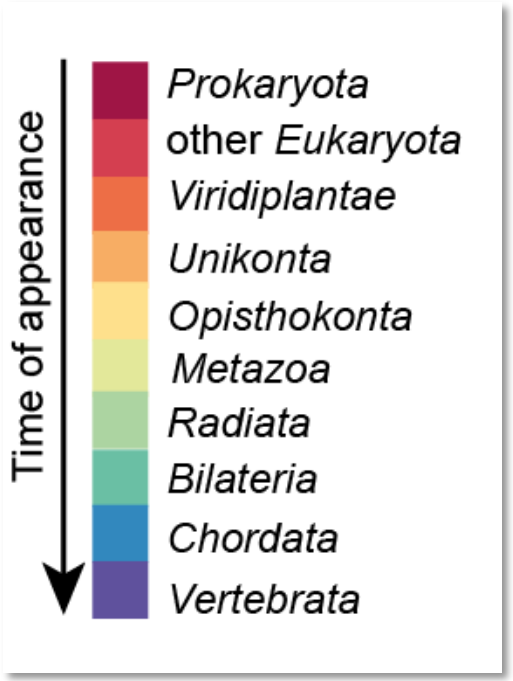
Ana M. Rojas
Jesús Blázquez

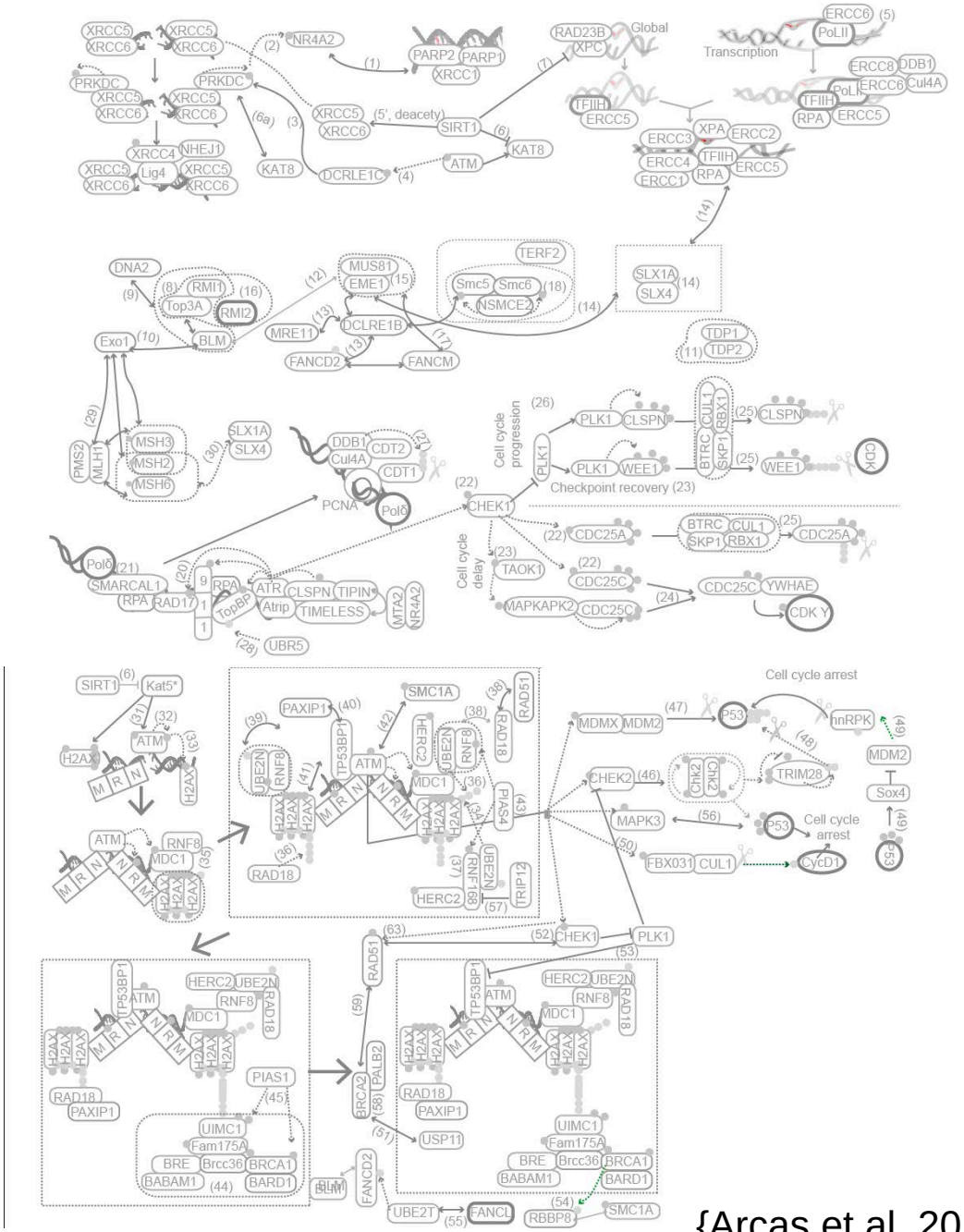
(Castañeda-García et al., Nat. Comm., 2017)

Jesus' interest: to identify novel **mutation avoidance** genes in bacteria

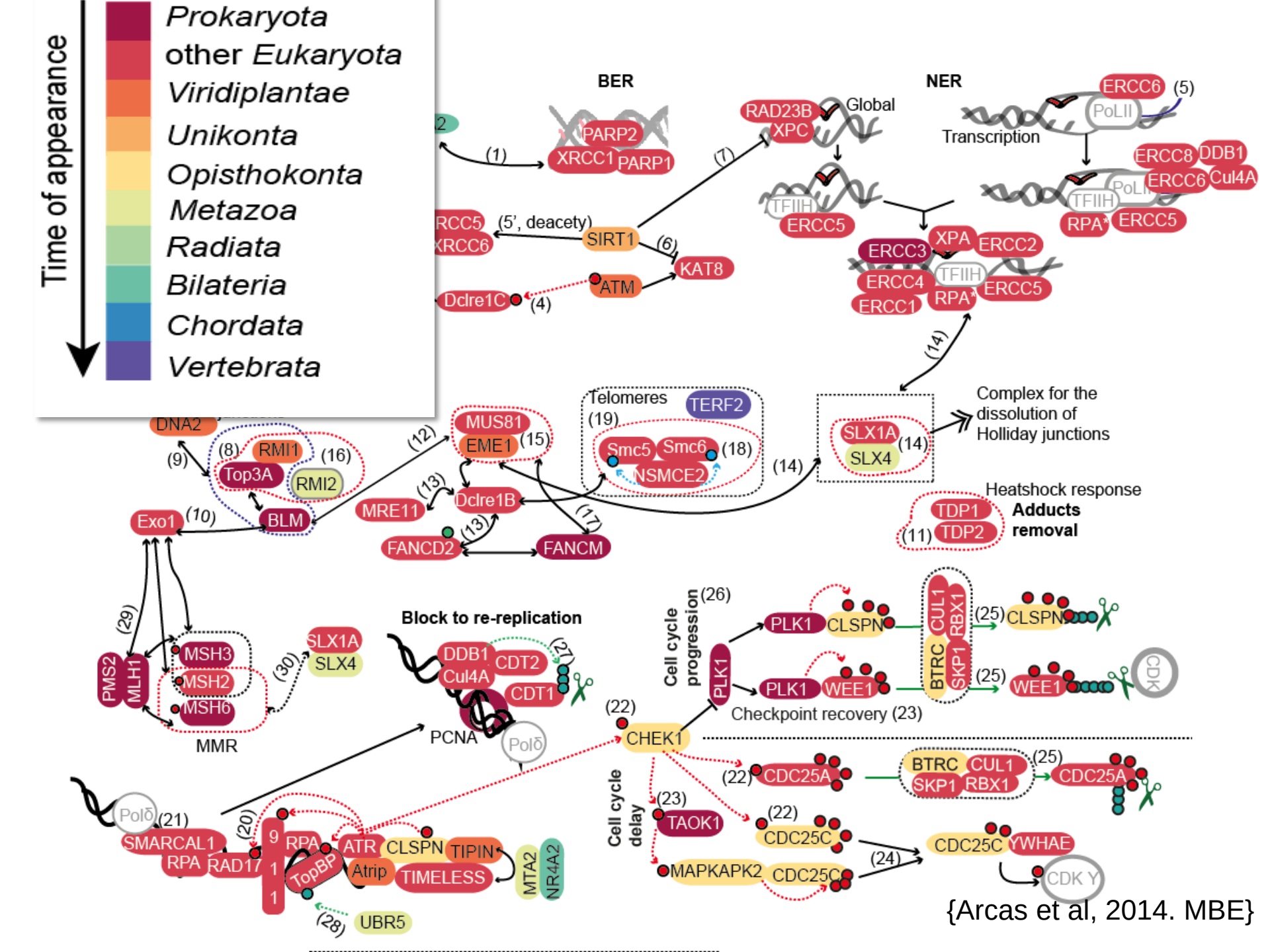
Ana' interest: **evolutionary analyses and emergence** of DNA repair and response proteins

Conservation





{Arcas et al, 2014. MBE}



{Arcas et al, 2014. MBE}

MMR repairs **mismatches** in dsDNA

(Friedberg et al, 2006).

Loss of this activity:

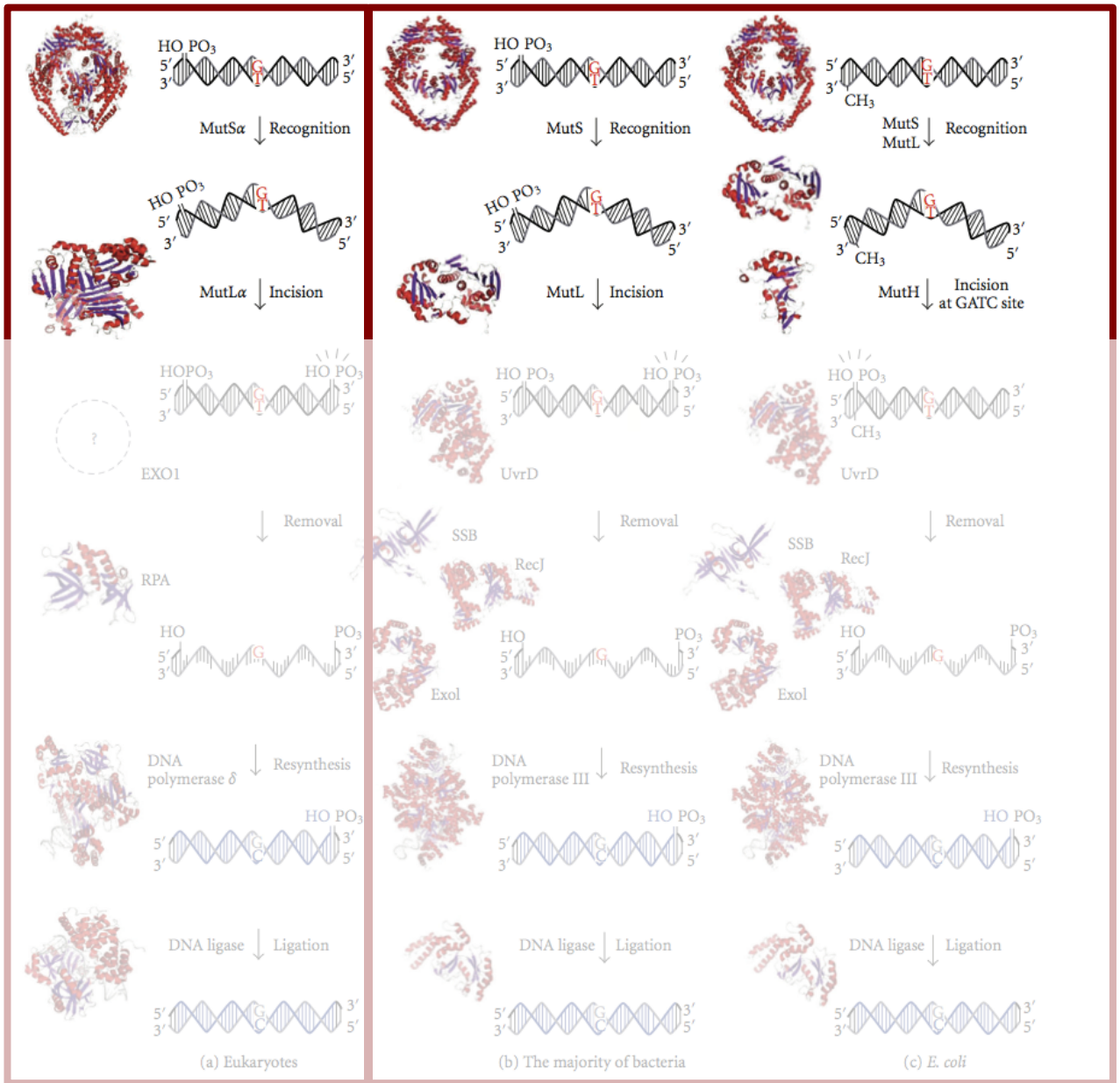
① **Hypermutability.**
② **Increased recombination** in homeologous DNA sequences

Hypermutable bacterial pathogens pose a **serious** risk in many **clinical infections**

(Gross et al, 1991;LeClerc et al., 1996; Matic et al, 1997; Oliver et al, 2000; Picard et al, 2001).

MutS/L are the principal **MMR** proteins in nature

(Fukui, K. J Nuc Acids, 2010)



MutS/L are the principal **MMR** proteins in nature

(Fukui, K. J Nuc Acids, 2010)

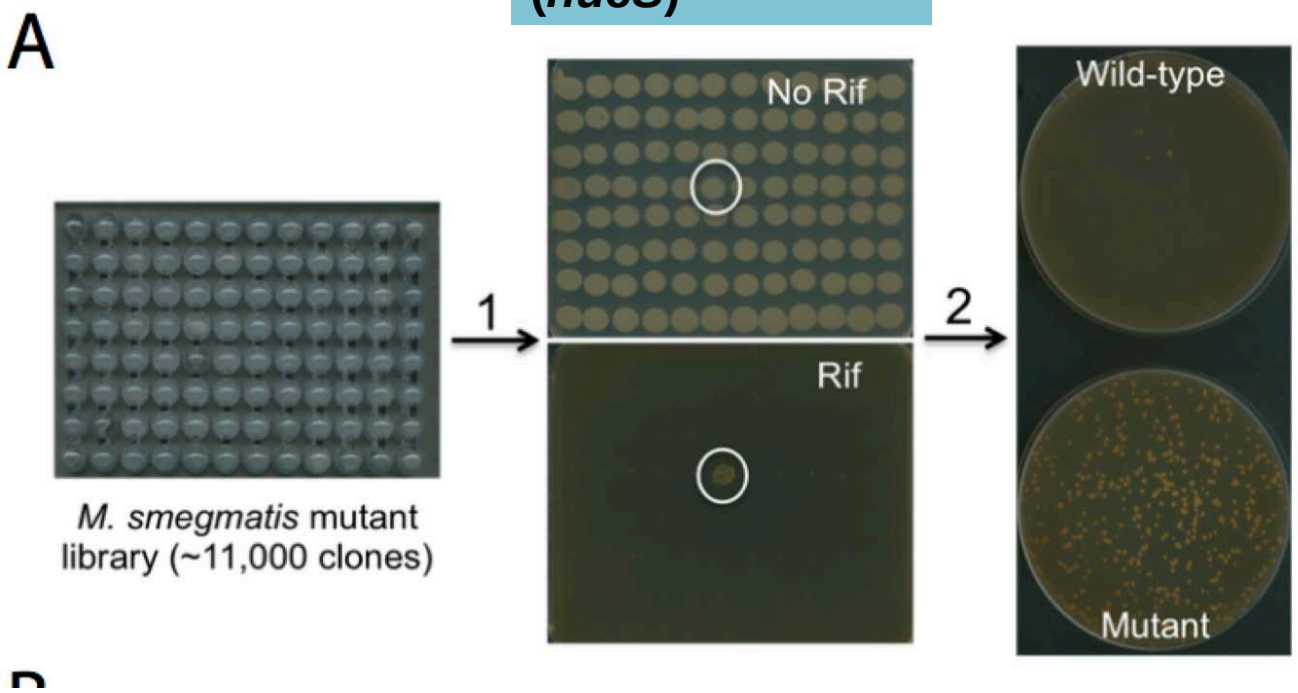
① MutS is **missing** in Actinobacteria (Mizrahi & Andersen. Mol. Microbiol. 1998; Sachadyn, Mut Res 2010; Banasik, M. & Sachadyn, Mut Res, 2014)

① And these exhibit **similar mutation** rates than MutS-containing organisms (Springer et al., Mol Microbiol, 2004; Ford et al. Nat Gen, 2011; Kucukyildirim et al. G3 2016).

Genetic screening

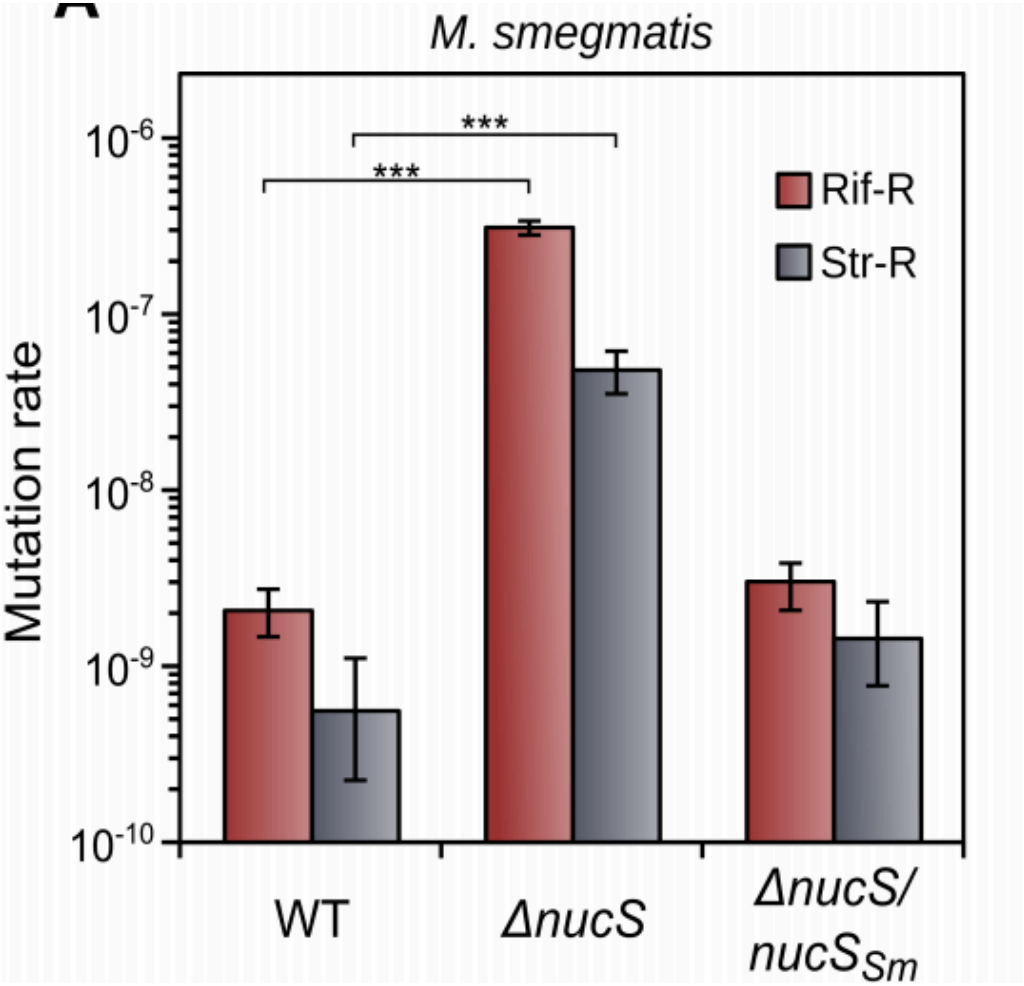
*M. smegmatis mc2 155 library of **~11,000 independent transposon** insertion mutants was generated and screened for spontaneous mutations that confer rifampicin resistance (Rif-R), used as a hypermutator hallmark (Fig. 1A).*

MSMEG_4923 (nucS)



*Using *M. smegmatis* as a surrogate model

MMR inactivation Hallmarks

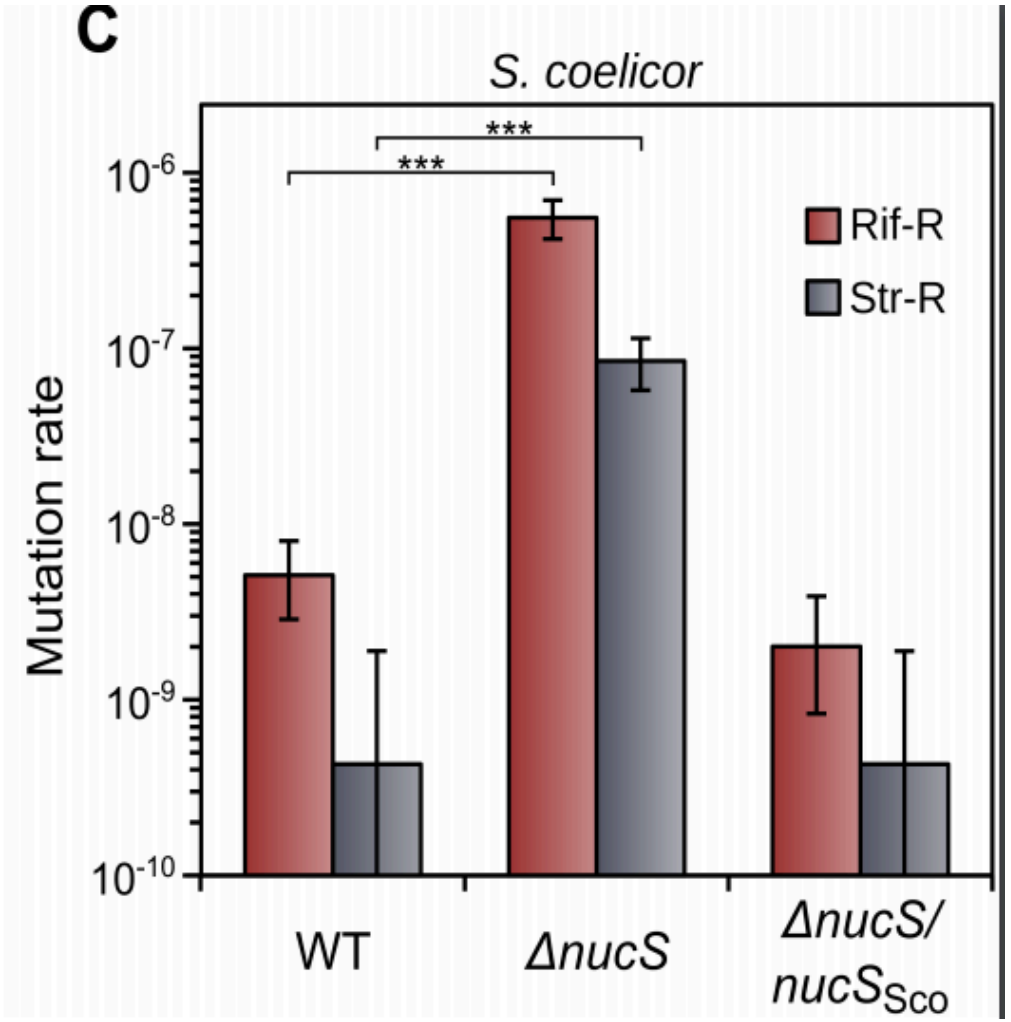


150X increased ripamficin
86X increased streptomycin

MutS - *E.coli* 102X
MutL – *E.coli* 103X

NucS is **essential** to maintain low levels of spontaneous mutation

MMR inactivation Hallmarks



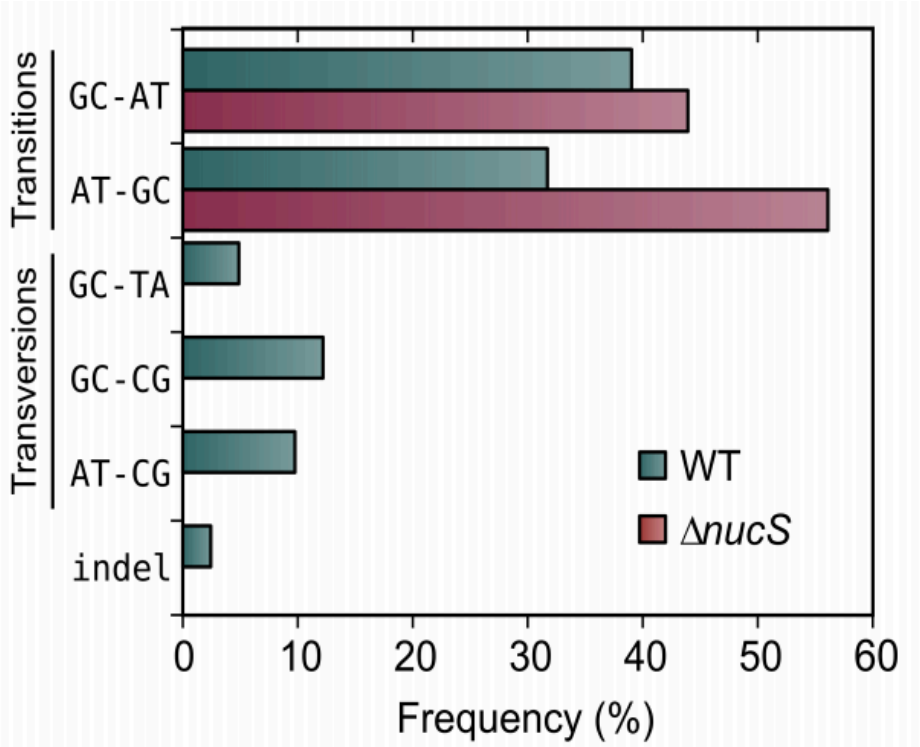
108X Rif-R
197X Srt-R

NucS is **essential** to maintain low levels of spontaneous mutation… **ALSO IN OTHER ACTINOMYCETES**

MMR inactivation Hallmarks

Δ*nucS* mutations are transitions

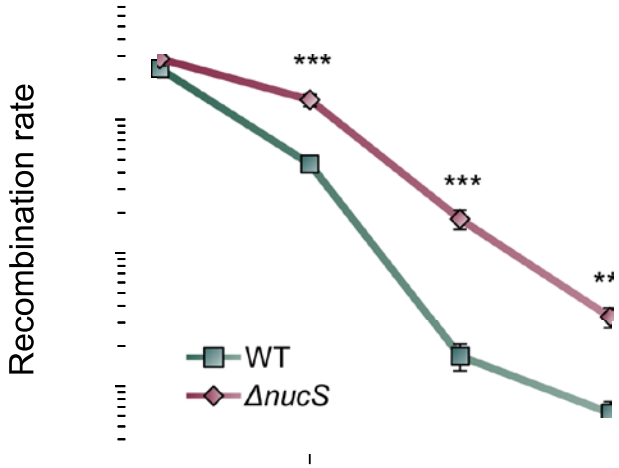
A:T->G:C
G:C->A:T



NucS mutational profile is **biased towards transitions**

MMR inactivation Hallmarks

Is NucS involved in reduction of recombination between non-identical(homeologous) DNA sequences, but not between 100% identical? (as described for the canonical MMR-null mutants in other bacterial species)



NucS **inhibits homeologous** but not homologous recombination

P value $<10^{-4}$ in all cases

NucS is a bona fide MMR protein

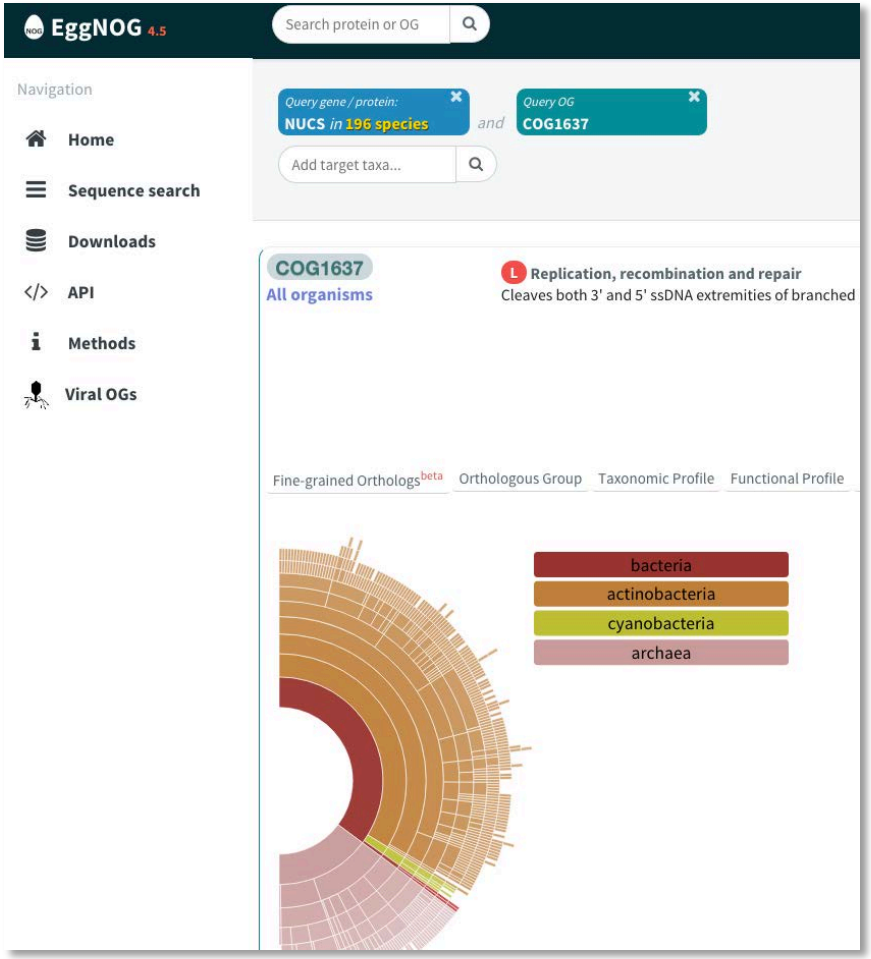
Some arising questions

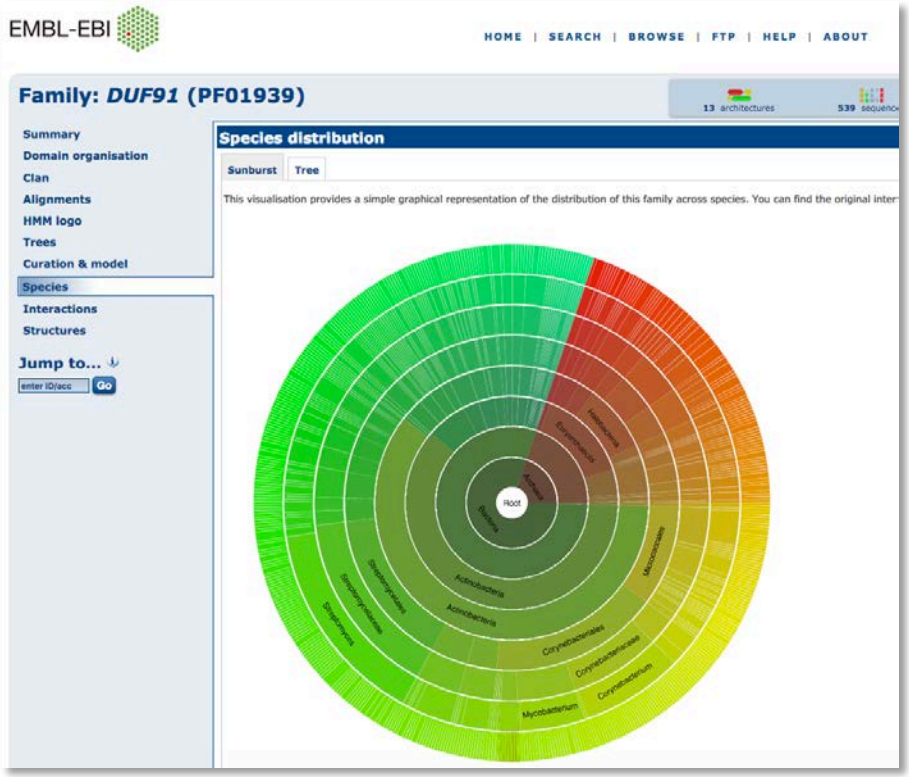
① **Where it NucS?**

① Where does it come from?

Estimating **NucS taxonomic distribution**

Screen for orthologues in public repositories

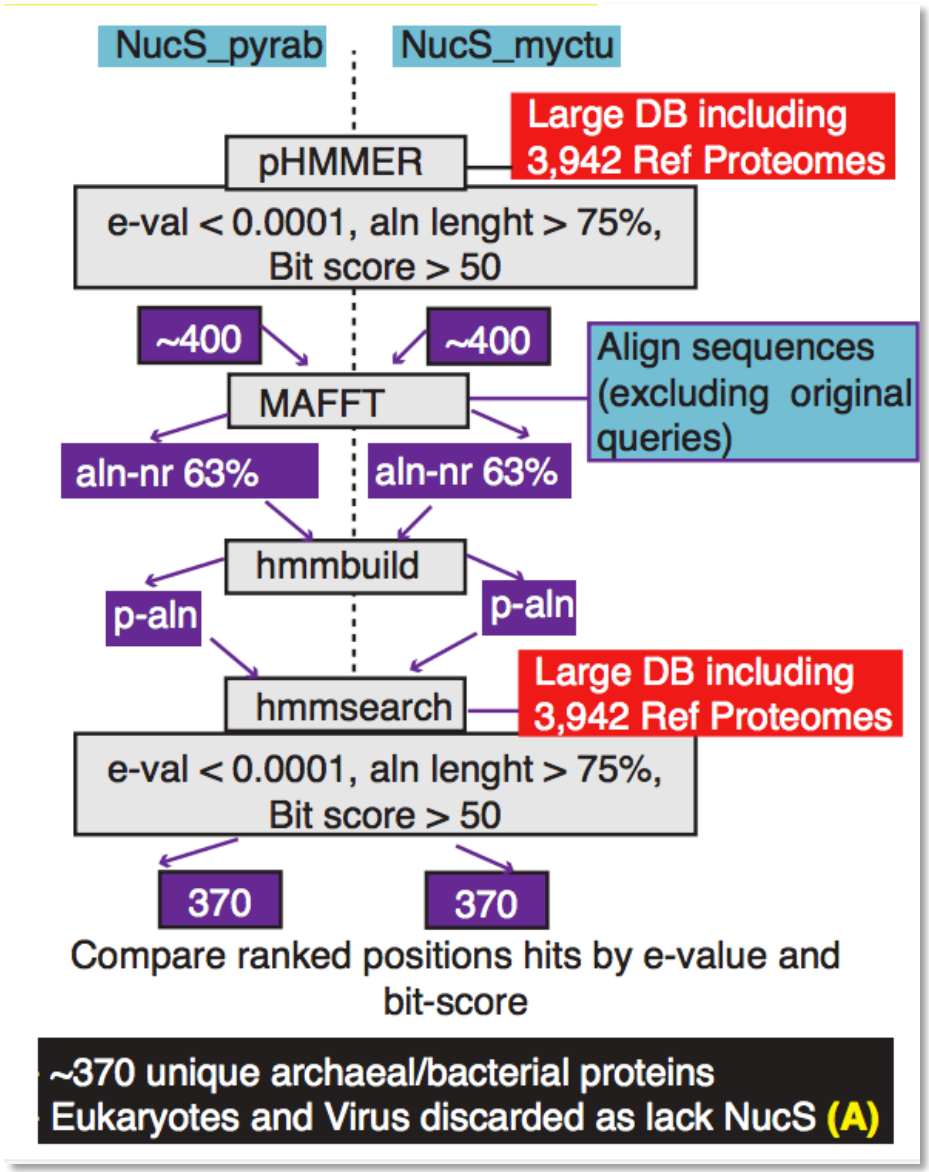




Only found in Bacteria and Archaea (**~500 seqs**)

Estimating **NucS taxonomic distribution**

BUT… when I run refined searches

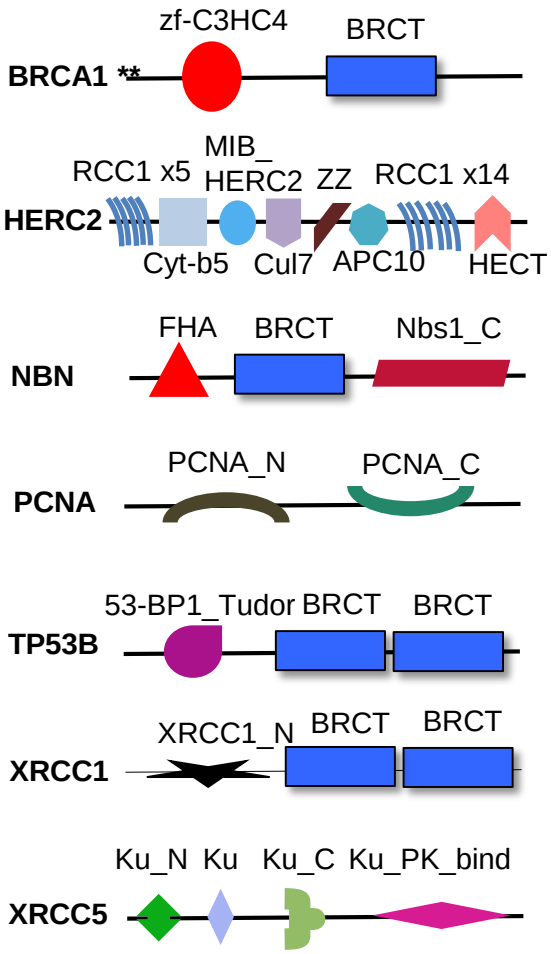


FULL NucS Only found in Bacteria and Archaea (~370 seqs)

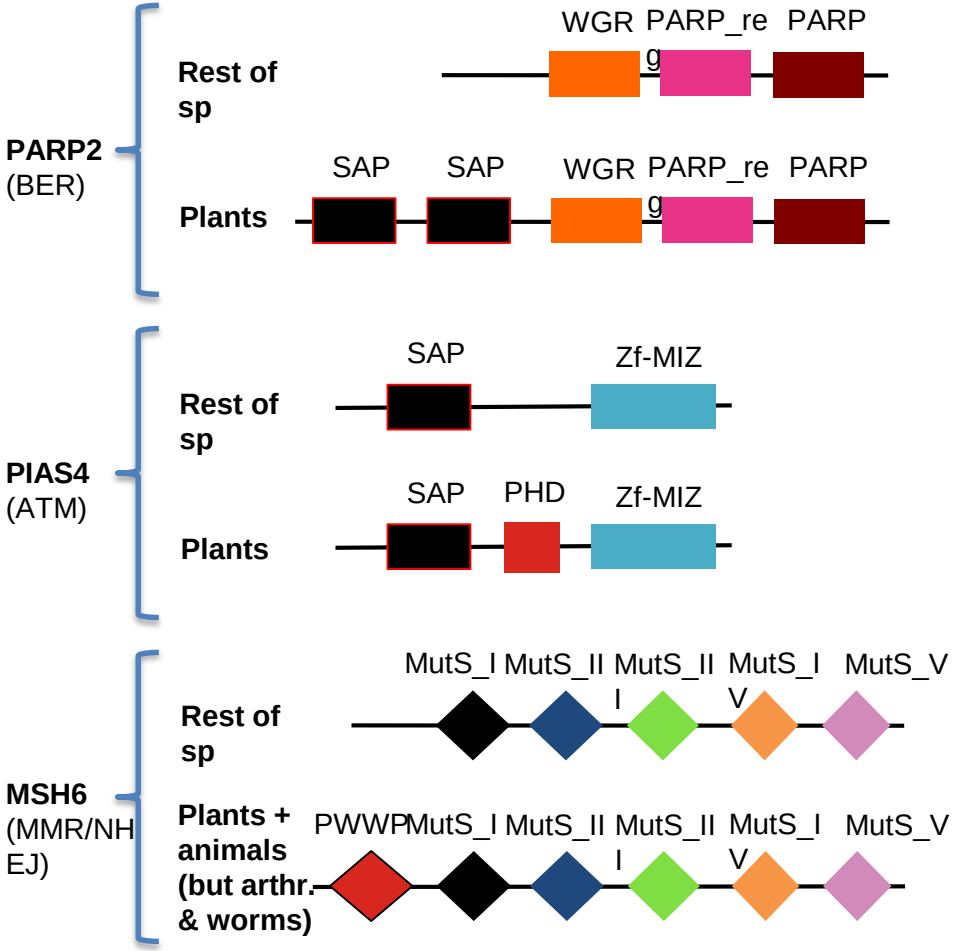
There is a GAP of about a 200 sequences size!

Check conservation of domain architecture

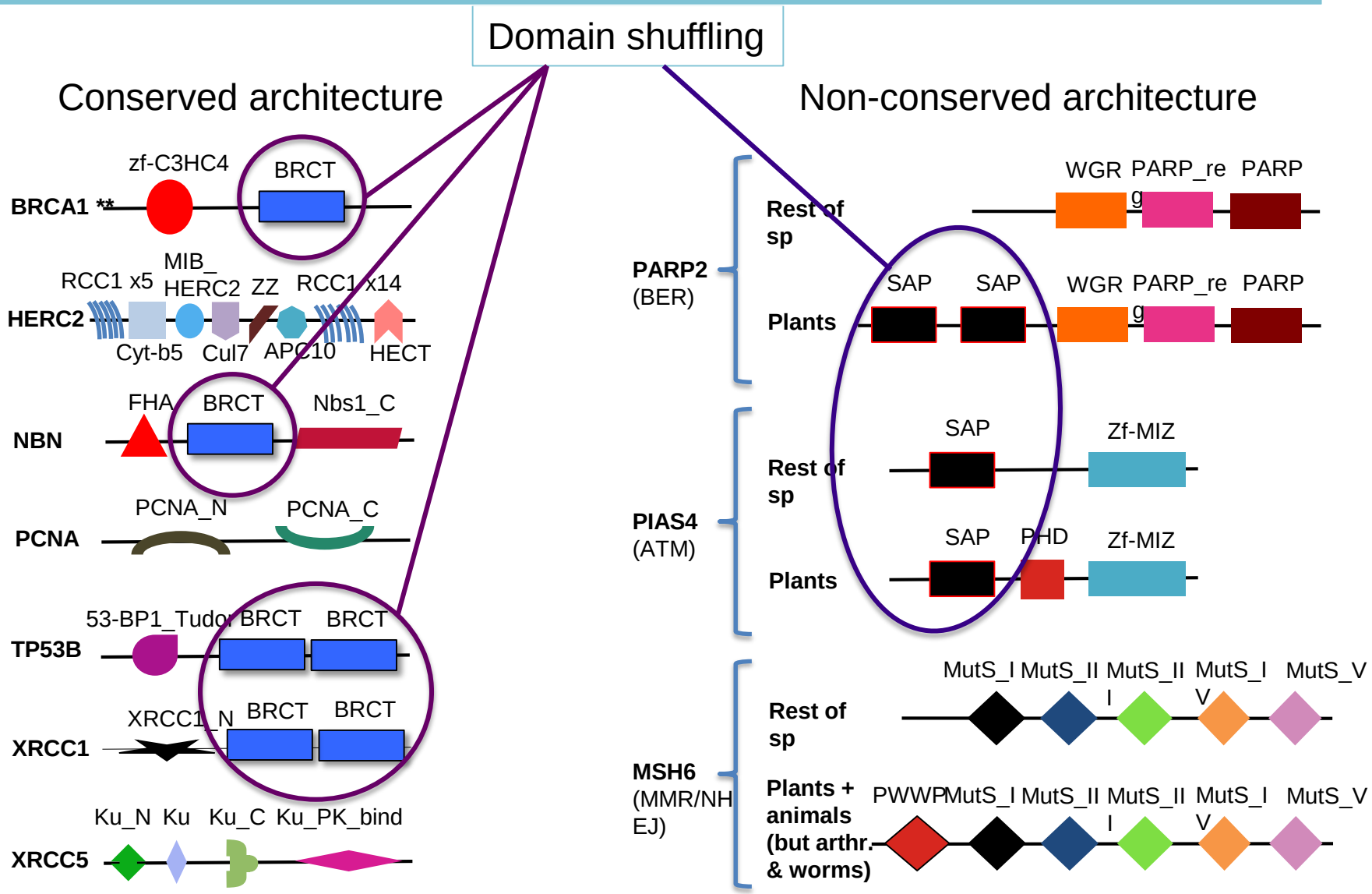
Conserved architecture



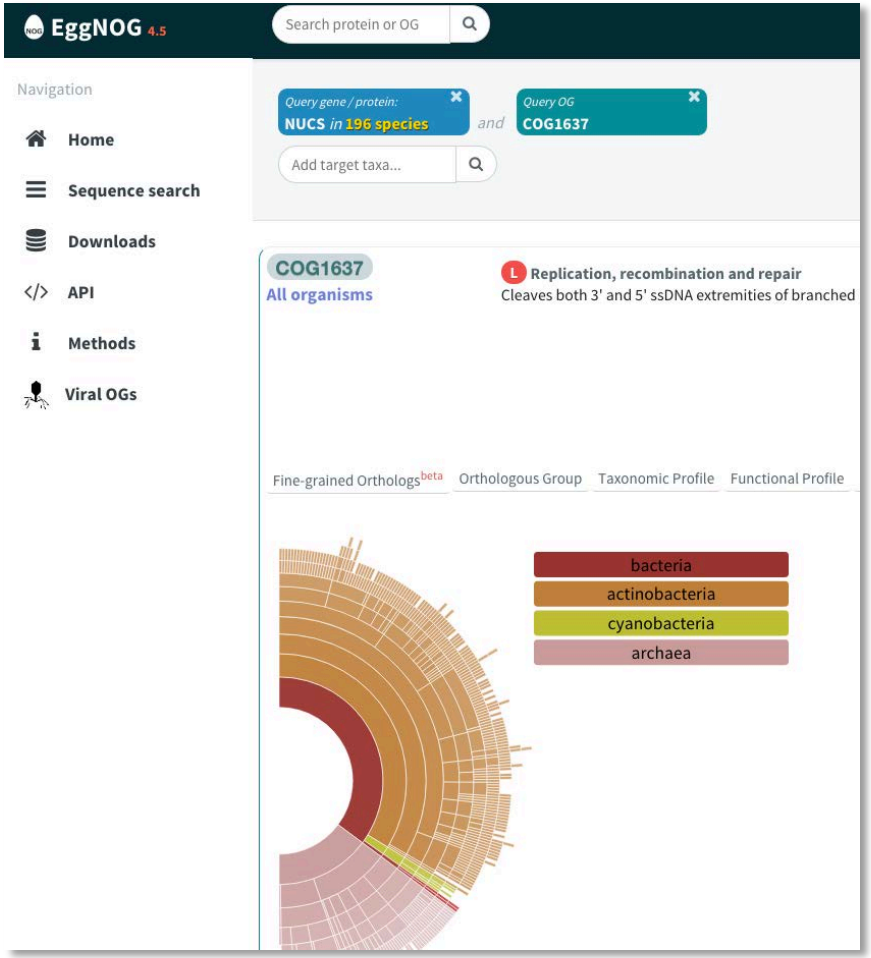
Non-conserved architecture

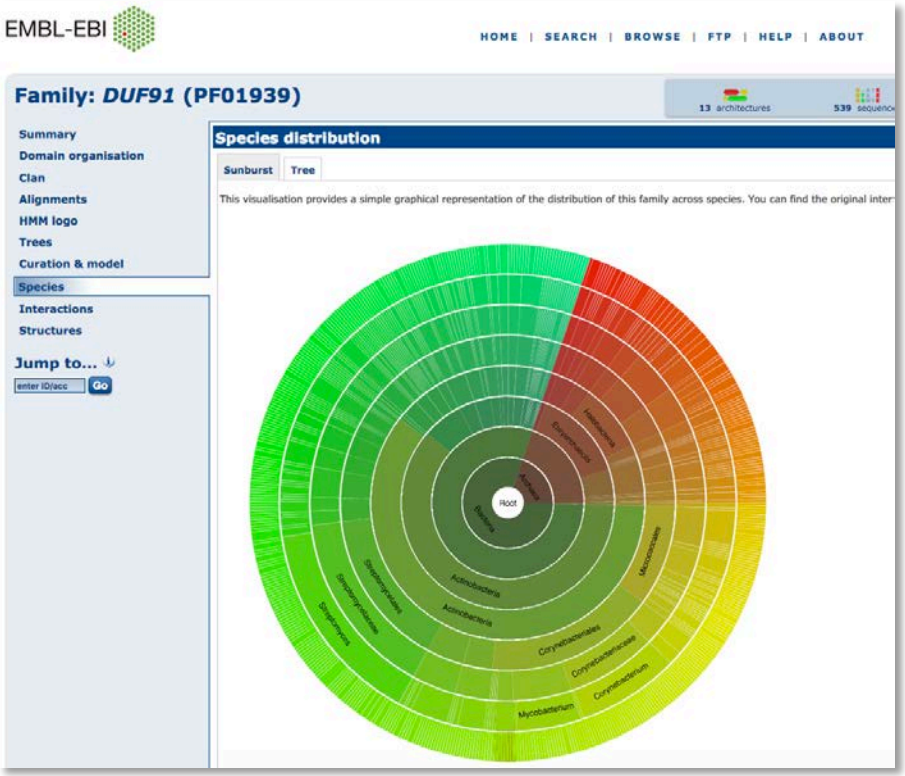


Detect potential domain shuffling= cause of misleading orthology assignations



Partial hits at significant e-values (suggest a multidomain protein)





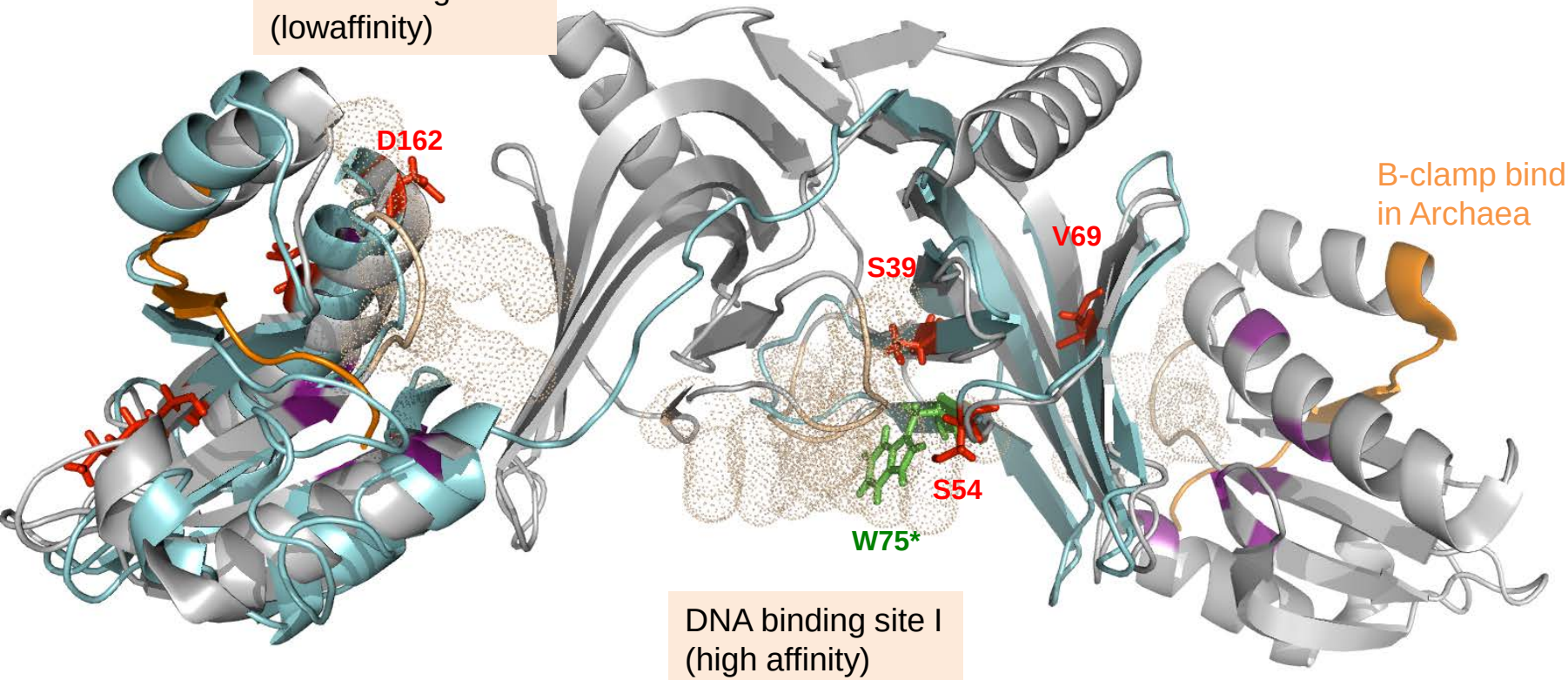
~500 seqs all hits vs. ~250 full hits

NucS is a two-domain protein

Grey: Archaea NucS
Cyan: Bacterial
Purple: catalytic sites in archea
Red: SNPs

N-terminal domain

DNA binding site II
(lowaffinity)



B-clamp binding
in Archaea

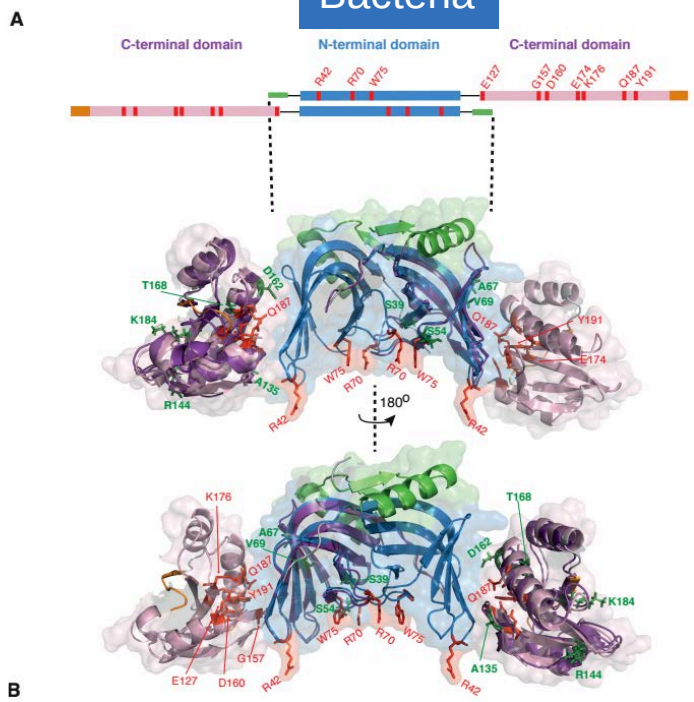
DNA binding site I
(high affinity)

C-terminal domain

***Its disruption totally abolish binding of ssDNA in the archaea**

NucS is a two-domain protein

Archaea
Bacteria



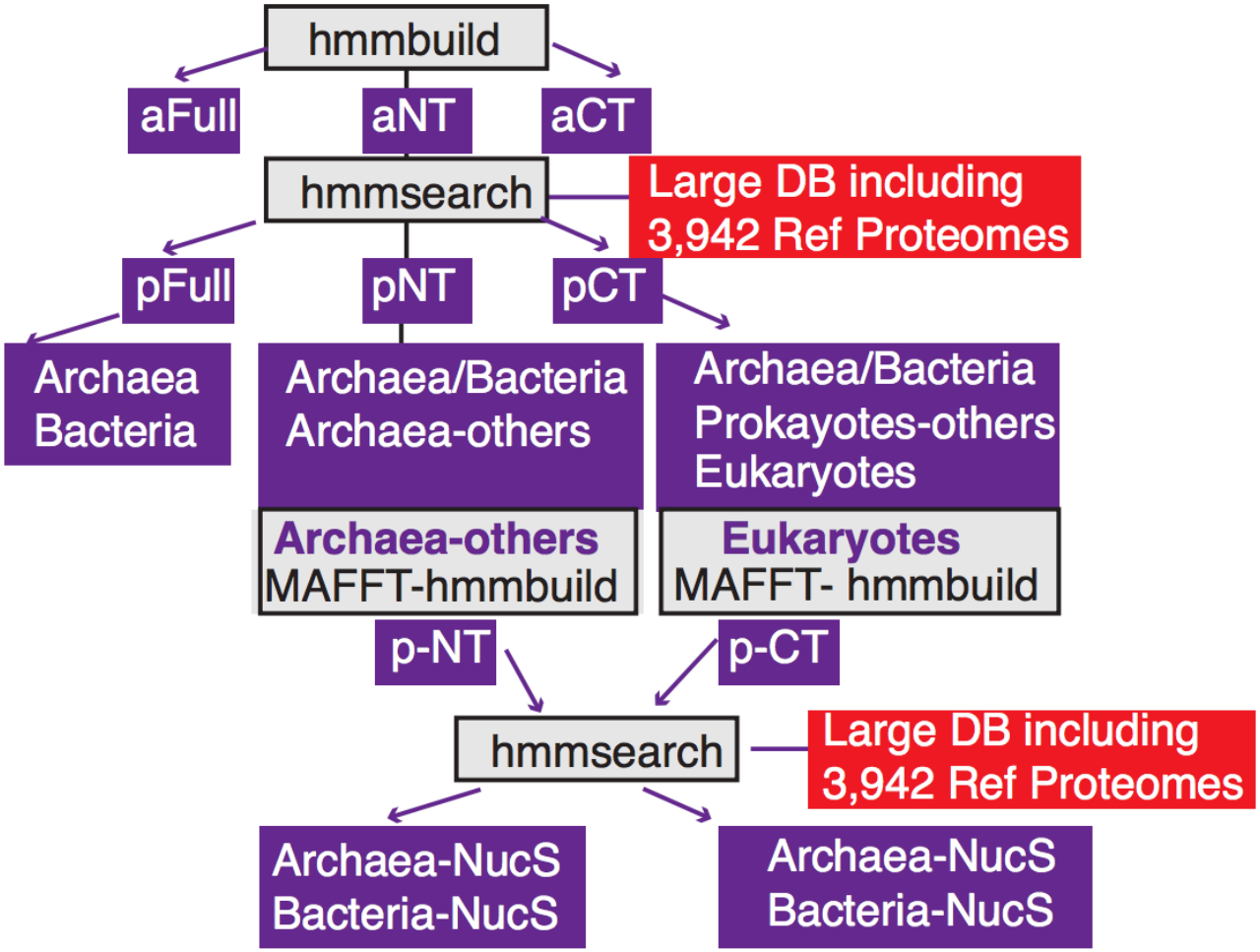

```
2VLD_Paby      25 HGGVVTIFARCKVHYEGRAKSELGEGDRIIIIKPDGSFLIHQN-KKREPVNWQPPGSKVT
NucS_Mycsme     1 M---RLVIAQCTVDYVGRLTAHLPSARRLLLFKADGSVSVHADDRAYKPLNWMSPPCWVT
NucS_Myctu      1 MSRVRLVIAQCTVDYIGRLTAHLPSARRLLLFKADGSVSVHADDRAYKPLNWMSPPCWLT
                    * * * * **    *    *    * *** R *       * ** I*    *

2VLD_Paby      84 FK--E--NSmISIRRRPYERLEVEIIEPYSLVVFLAEDYEELaltgSEAEmANLIFENPR
NucS_Mycsme    58 EQDTETGVALWVVENKTGEQLRITVEDIEHDSHHELGVDPGLVKDGVEAHLQALLAEHVE
NucS_Myctu     61 EESGGQ-APVWVVENKAGEQLRITIEGIEHDSSHELGVDPGLVKDGVEAHLQALLAEHIQ
                     S A         * *                   *   * **    *  *

2VLD_Paby     140 VIEEGFKPIYREKPIRHGIVDVMGVDKDGNIVVLELKRRKADLHAVSQLKRYVDSLKEEY
NucS_Mycsme   118 LLGAGYTLVRREYPTPIGPVDLLCRDELGRSVAVEIKR-RGEIDGVEQLTRYLELLNRDS
NucS_Myctu    120 LLGEGYTLVRREYMTAIGPVDLLCRDERGGSVAVEIKR-RGEIDGVEQLTRYLELLNRDS
                    *     **   S * **   S*  *  *  * **     H * **A**   *

2VLD_Paby     200 -GENVRGILVAPSLTEGAKKLLEKEGLEFRKLEPP 233
NucS_Mycsme   177 LLAPVAGVFAAQQIKPQARTLATDRGIRCVTLDYD 211
NucS_Myctu    179 VLAPVKGVFAAQQIKPQARILATDRGIRCLTLDYD 213
                   *E*   *      *  *    *     *
```

Archaea
Bacteria
Eukaryota*
(3 early metazoans)

M. smegmatis modeled on P abyssi structure.

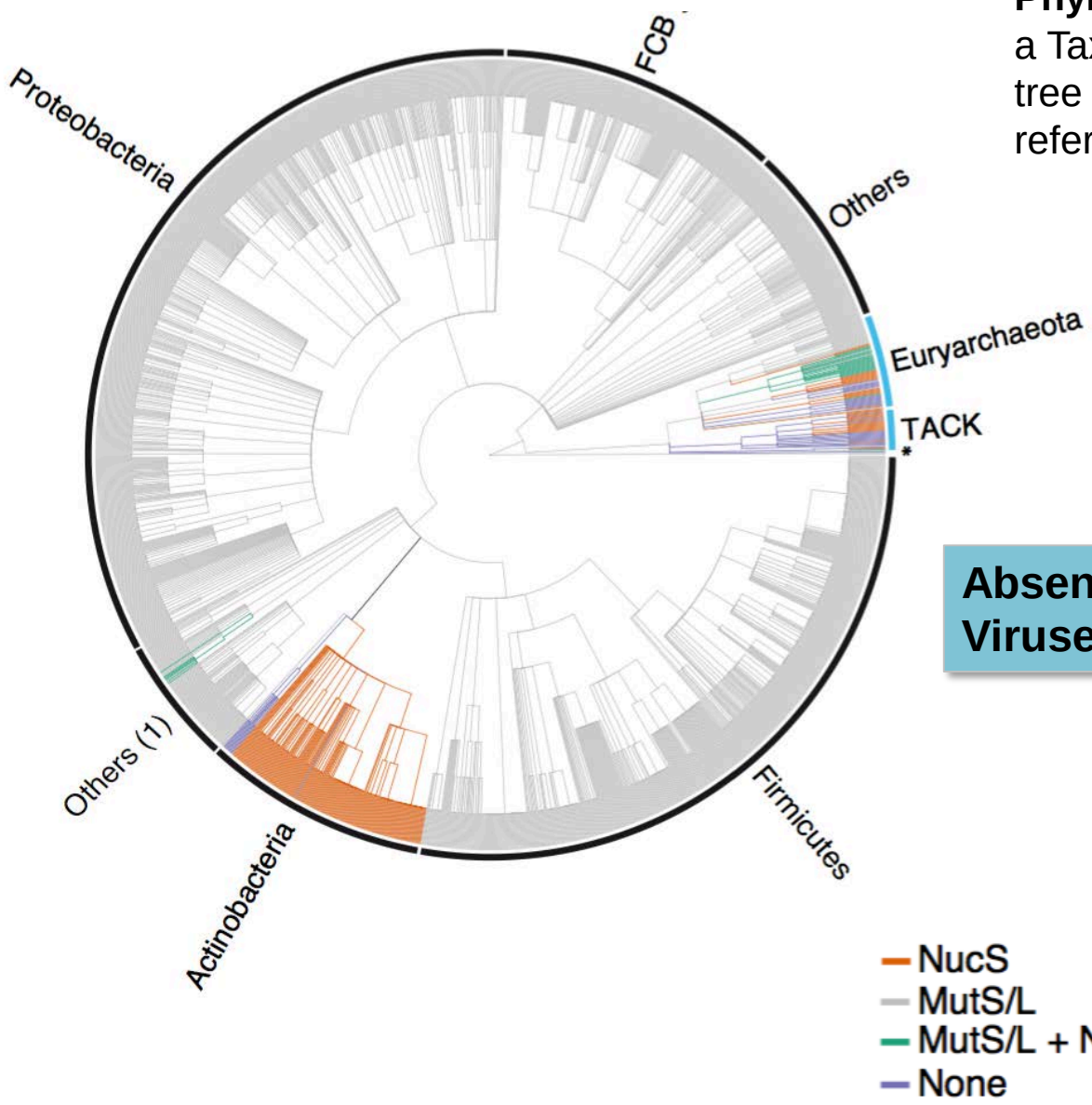
Domain Analyses I

Structure-based definition using NucS_pyrab: full, N-terminal and C-terminal.



NucS is built on two distinct domains **(B)** Supp. Fig 3

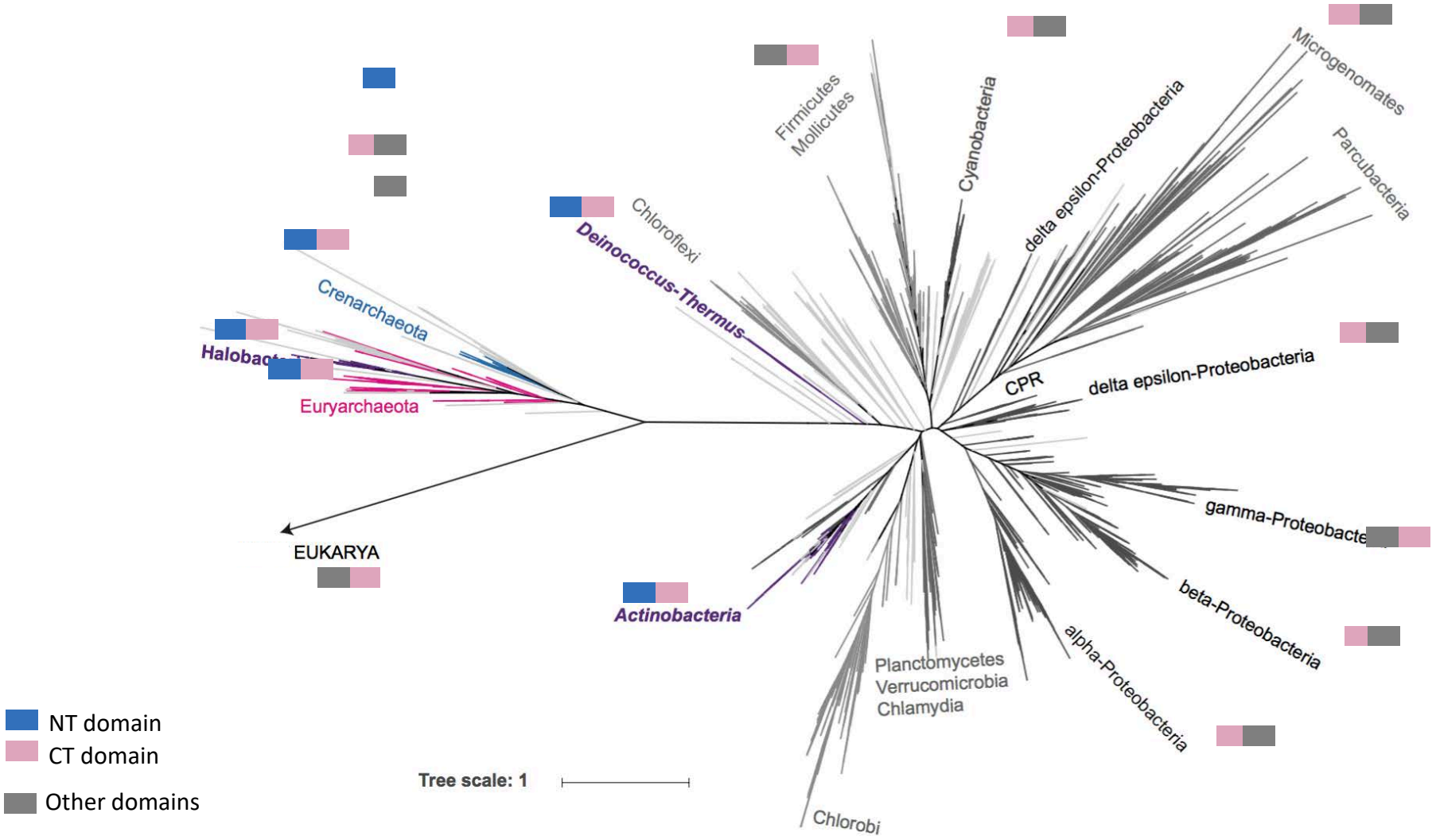
Full NucS is restricted to prokaryotes



Phylogenetic profiling on a Taxonomy NCBI based tree of ~2709 prokaryotic reference proteomes

Absent in Eukaryotes and Viruses

NucS's domains **exhibit** different distributions

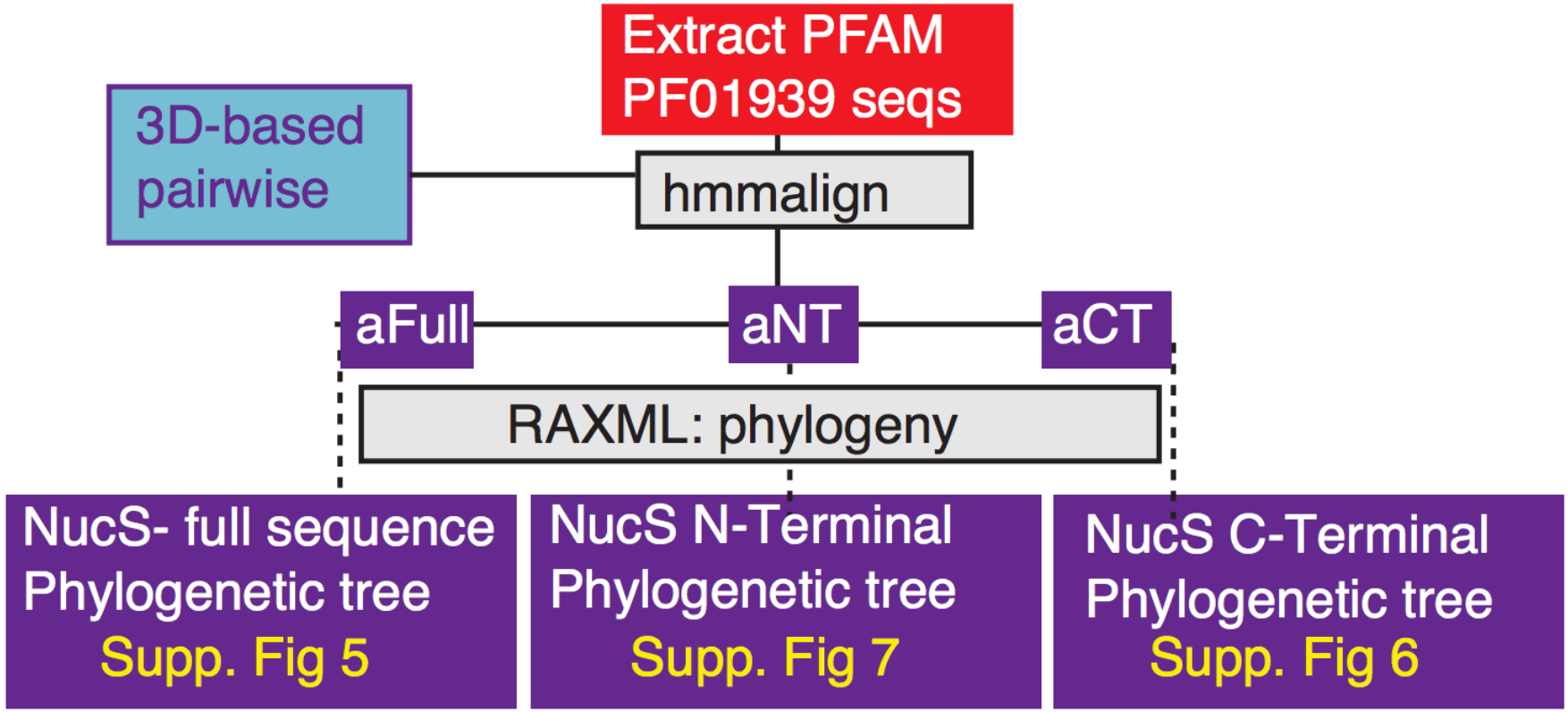


Some arising questions

① Where it NucS?

① **Where does it come from?**

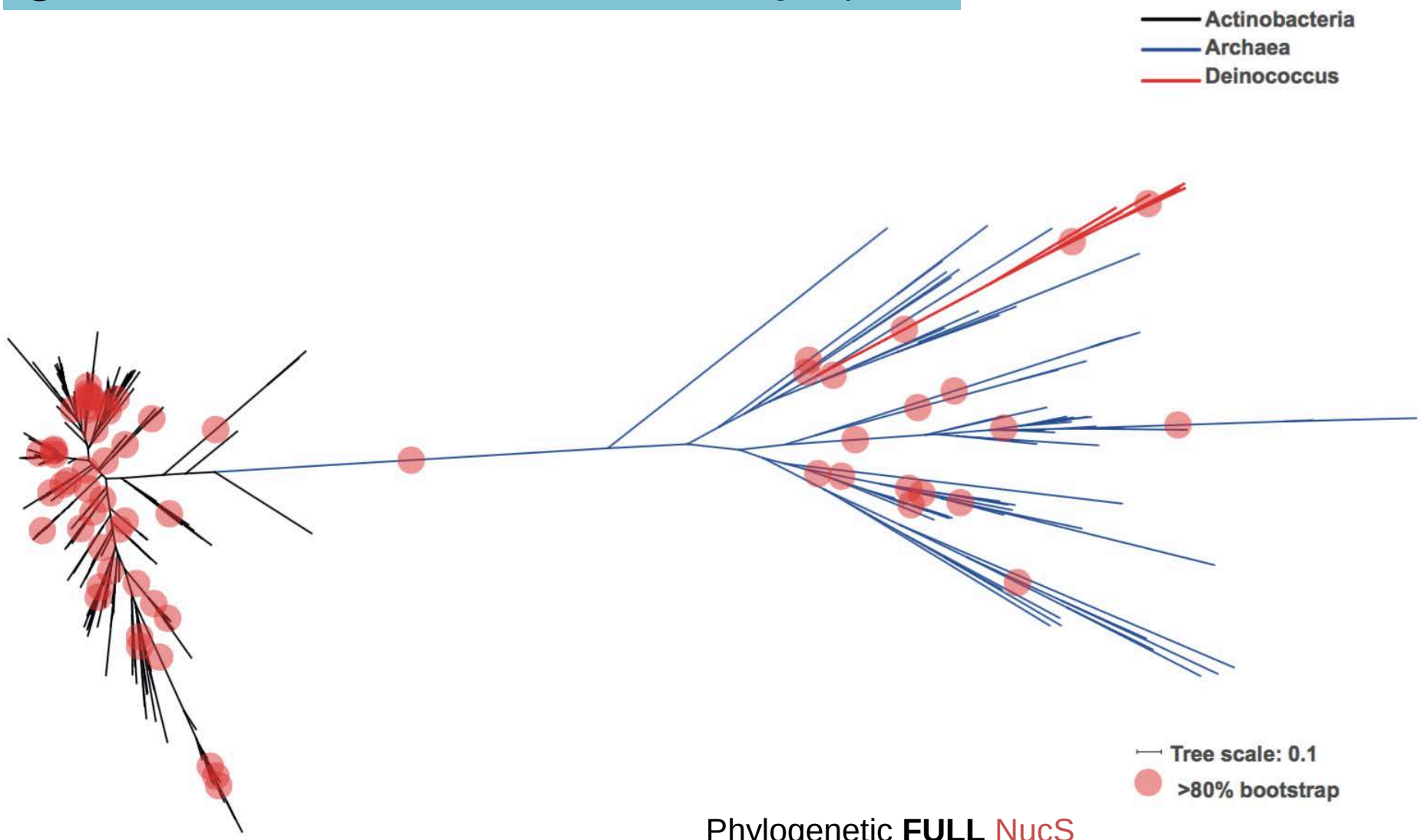
Domain Analyses II



NucS in some bacteria have been transferred from Archaea **(C)**

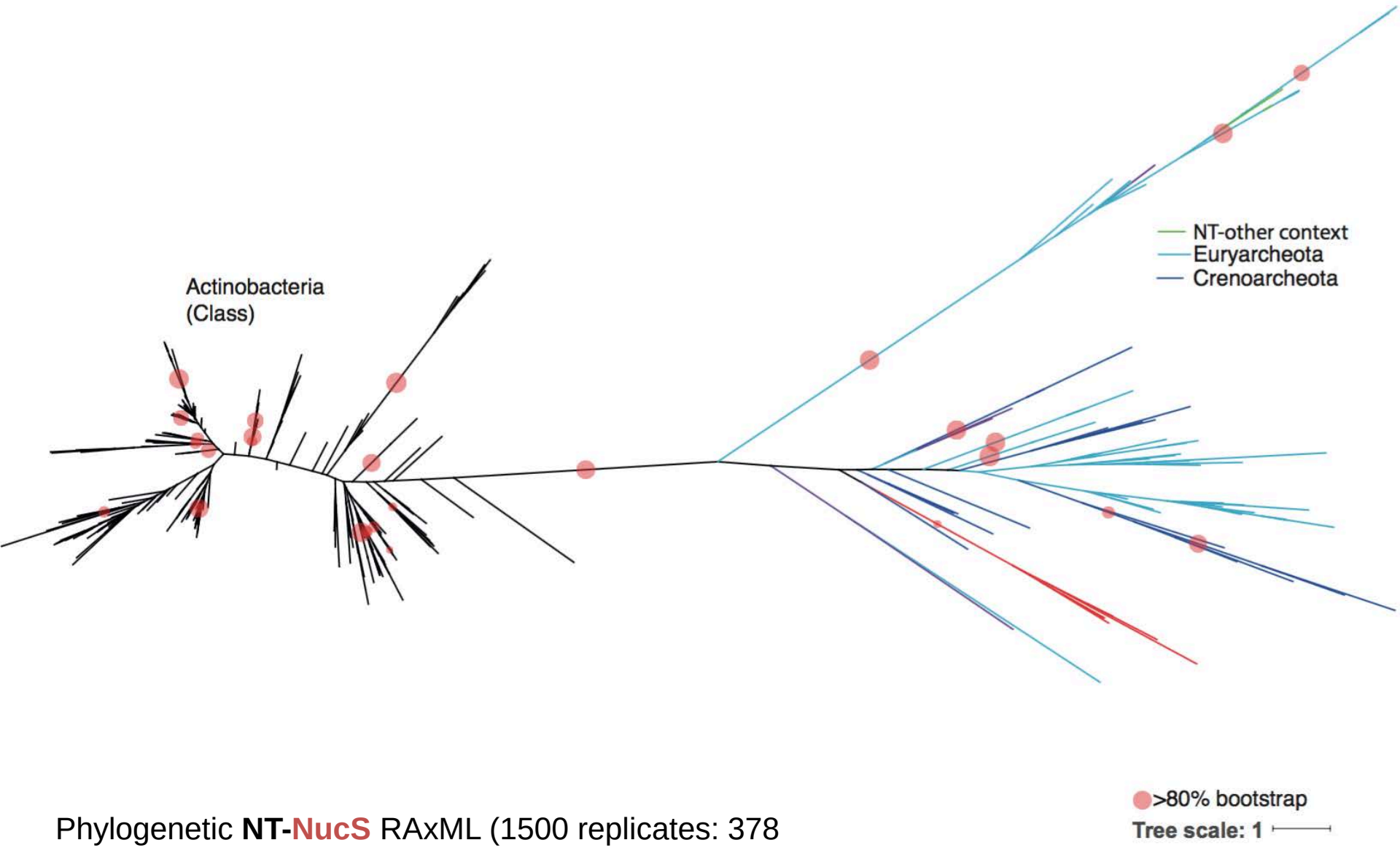
By Phylogenetic analyses

① HGT from Archaea to some Deinococcus group



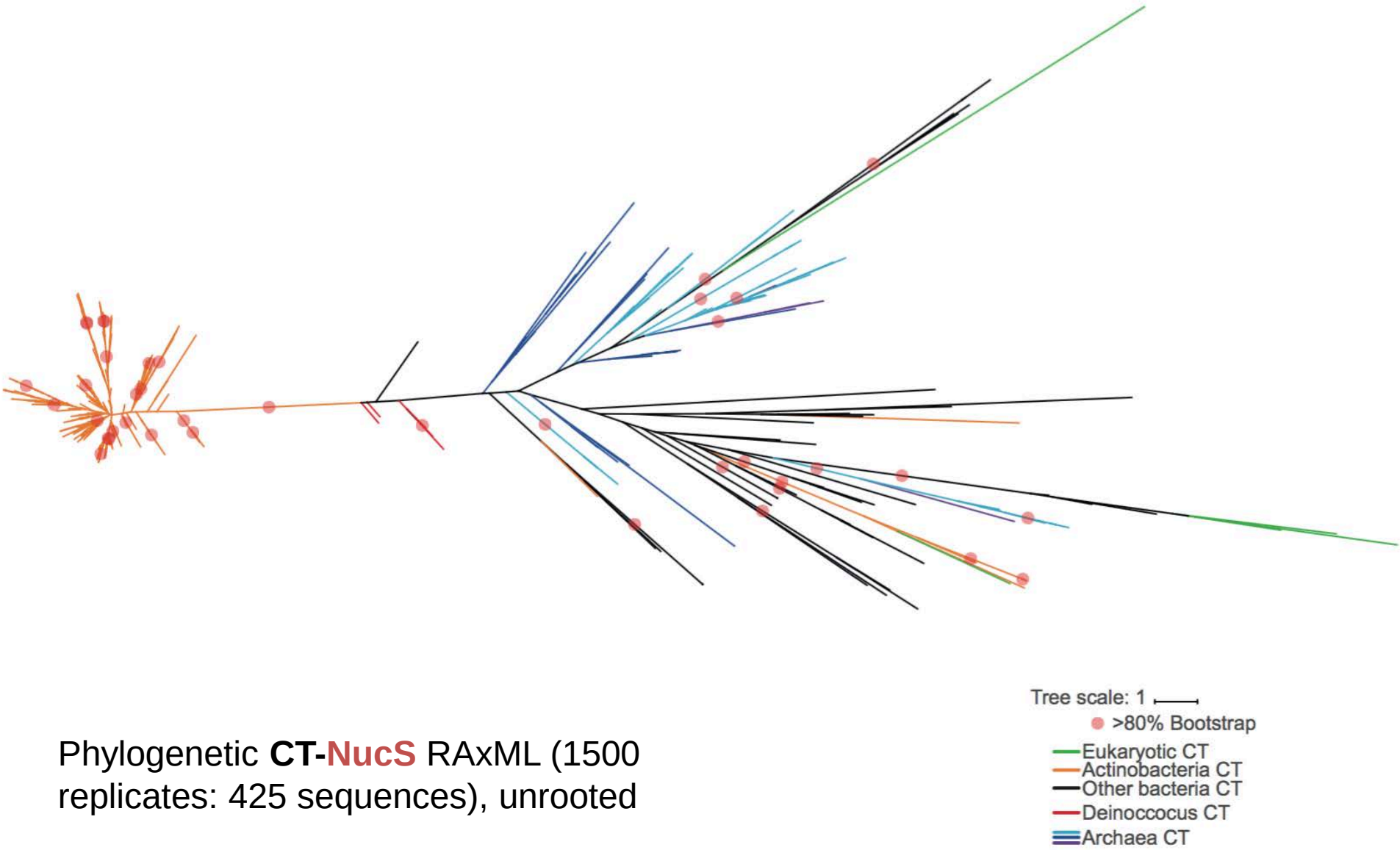
Phylogenetic **FULL** NucS
RAxML (1500 replicates:
378 sequences), unrooted

Where does N-terminal of NucS come from?



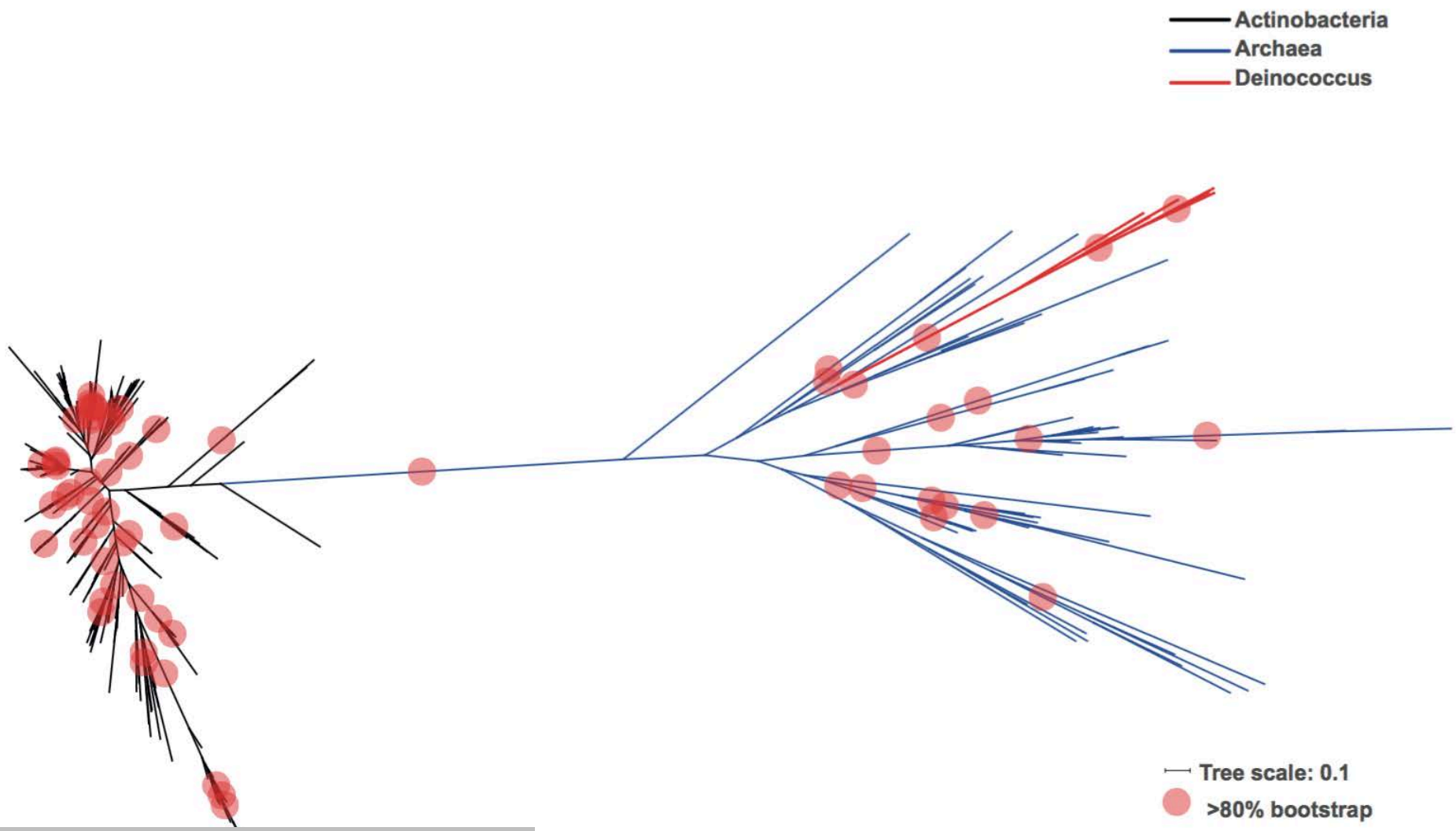
Phylogenetic **NT-NucS** RAxML (1500 replicates: 378 sequences), unrooted

Where does C-terminal of NucS come from?



Phylogenetic **CT-NucS** RAxML (1500 replicates: 425 sequences), unrooted

Full NucS has been horizontally transferred



HGT from Archaea to some Deinococcus group

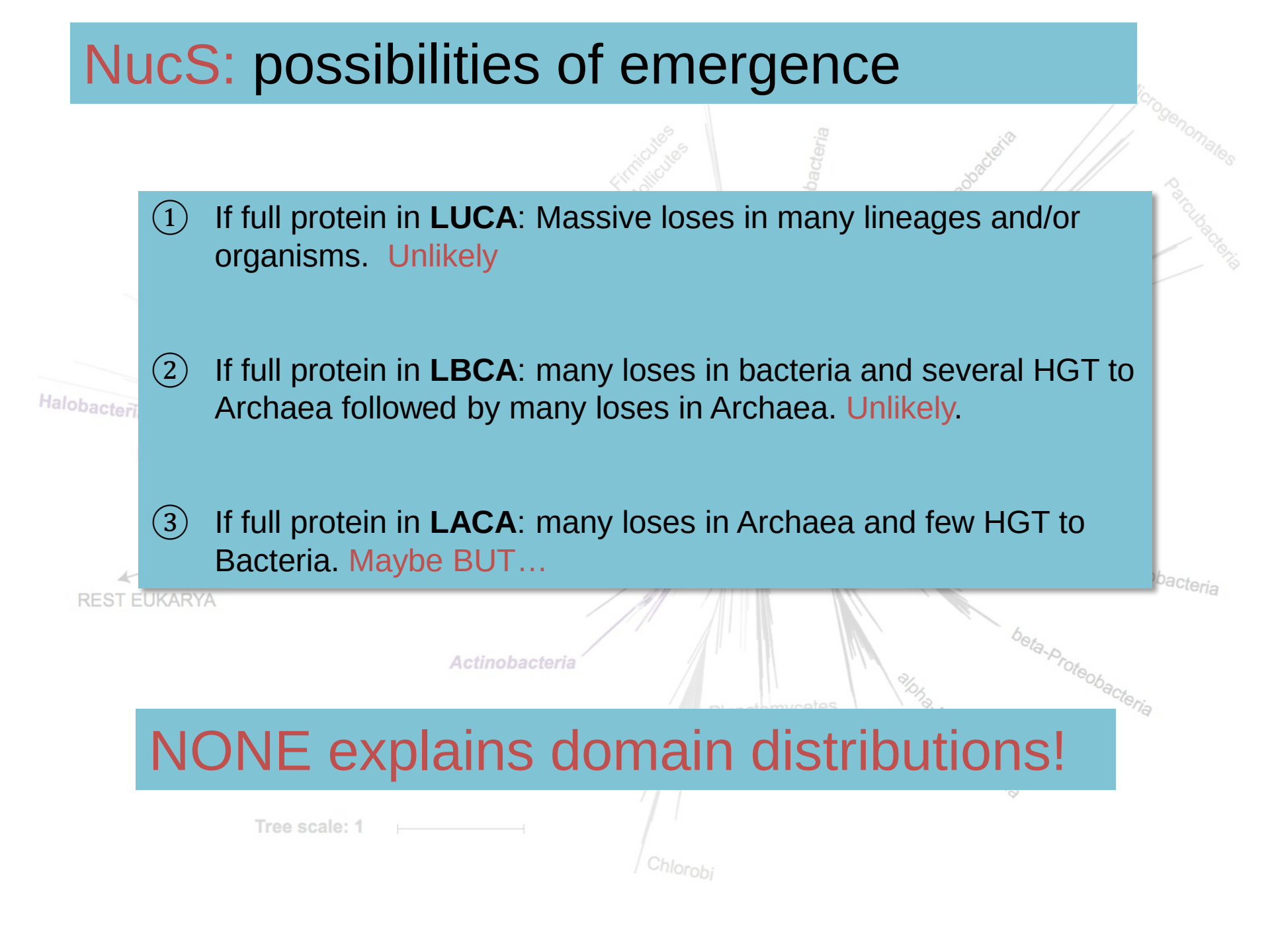
Phylogenetic **FULL** NucS RAxML (1500 replicates: 378 sequences), unrooted

NucS: possibilities of emergence

① If full protein in **LUCA**: Massive loses in many lineages and/or organisms. Unlikely

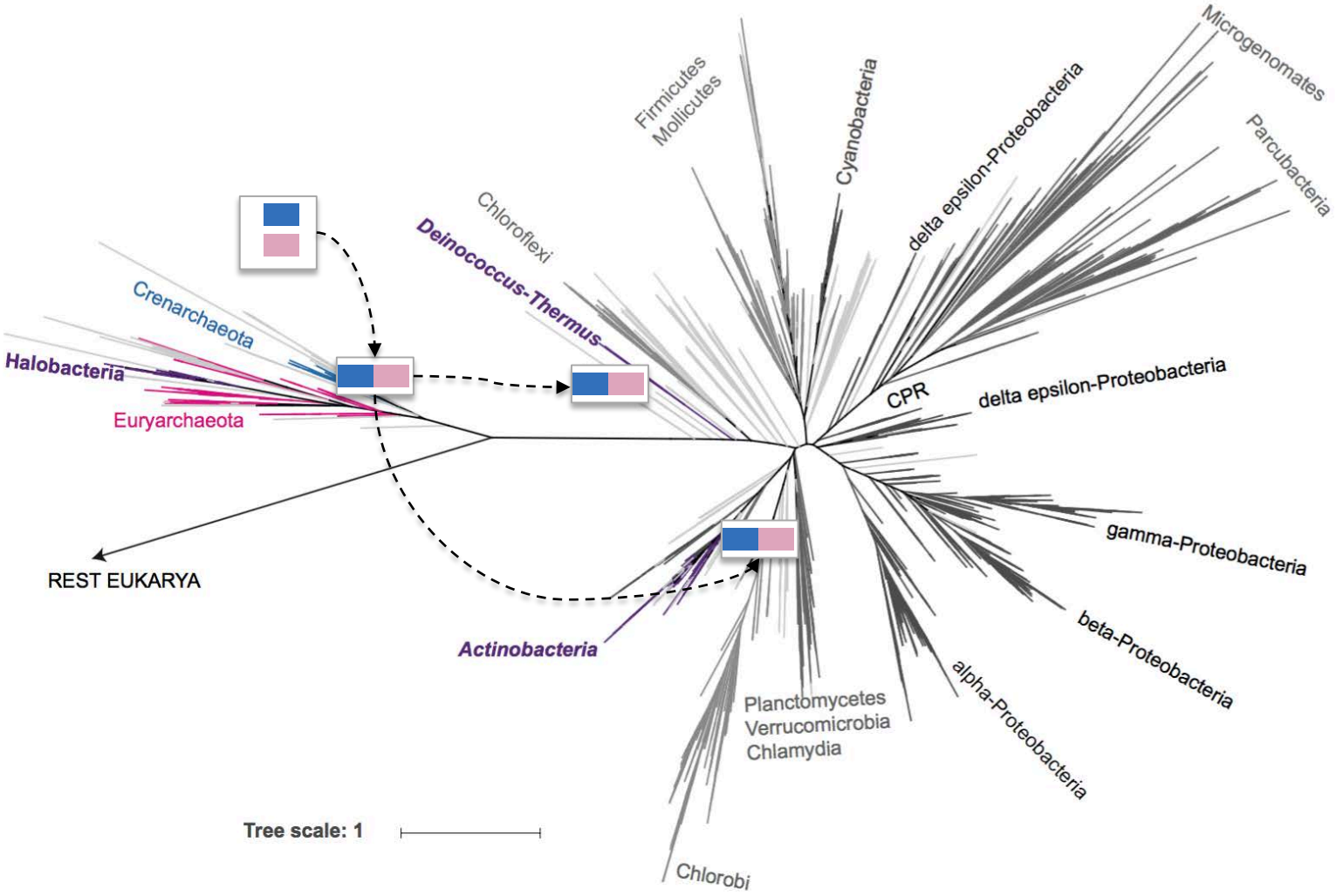
② If full protein in **LBCA**: many loses in bacteria and several HGT to Archaea followed by many loses in Archaea. Unlikely.

③ If full protein in **LACA**: many loses in Archaea and few HGT to Bacteria. Maybe BUT…

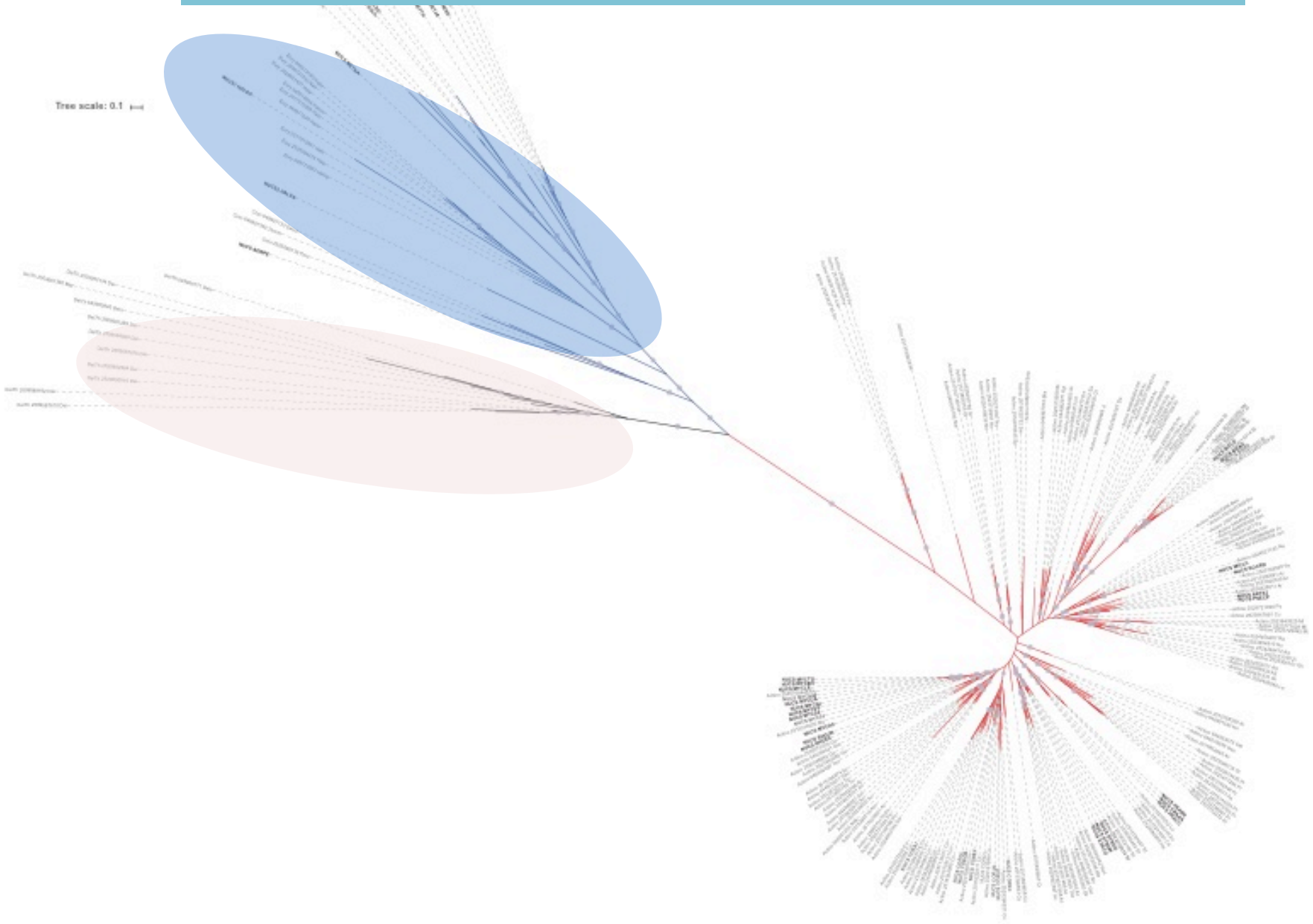


NONE explains domain distributions!

NucS is likely of archaeal origin



Is NucS distribution supported in GEBA?



Implications

① NucS could server as a **therapeutic target** to fight MDR strains

① Still more MMR pathways to unravel…

Focus on Actinobacteria. . .

Reported that *Mycobacterium tuberculosis* **does NOT** have hypermutable phenotypes

- Acquires Ab resistance exclusively trough chromosomal mutations (Muller et al, 2013; Ford et al, 2013)
- Presents variability in mutation rates among strains (Ford et al, 2013)
- Lacks MutS/L

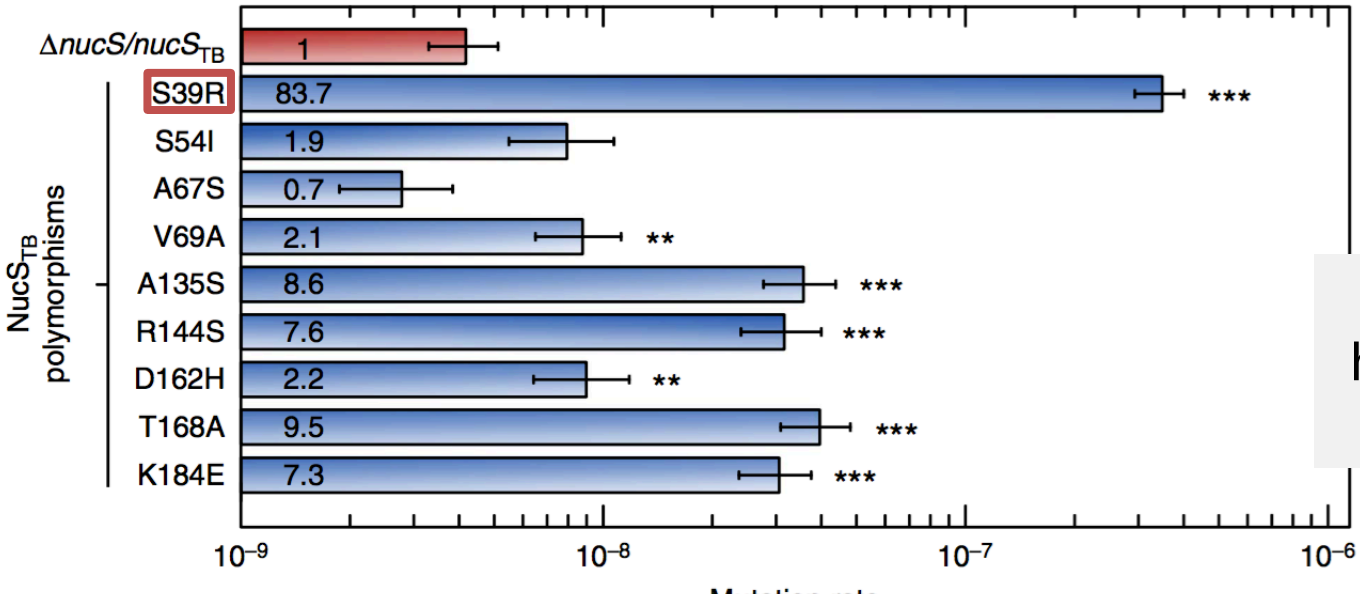
But it has MDR profiles

GenomeID/ name	Resistance profile	Lineage	Origin
CDC1551	Susceptible	4	North America
TKK_02_0079	MDR	4	South Africa
MTB_N1057	Susceptible	4	South Asia
KT-0040	Susceptible	2	S. Korea (Broad Inst)
ERR036236	Susceptible	1	Unknown
BTB 04-388	MDR	3	Sweden (Broad Inst)
BTB 07-246	MDR	4	Sweden (Broad Inst)
TKK_03_0044	Susceptible	4	South Africa
HN2738	Unknown	Unknown	Unknown (Broad Inst)
MTB_X632	MDR	4	Central America

1,600 clinical *M. tuberculosis* strains*, 9 SNPs

*Ensembl bacteria

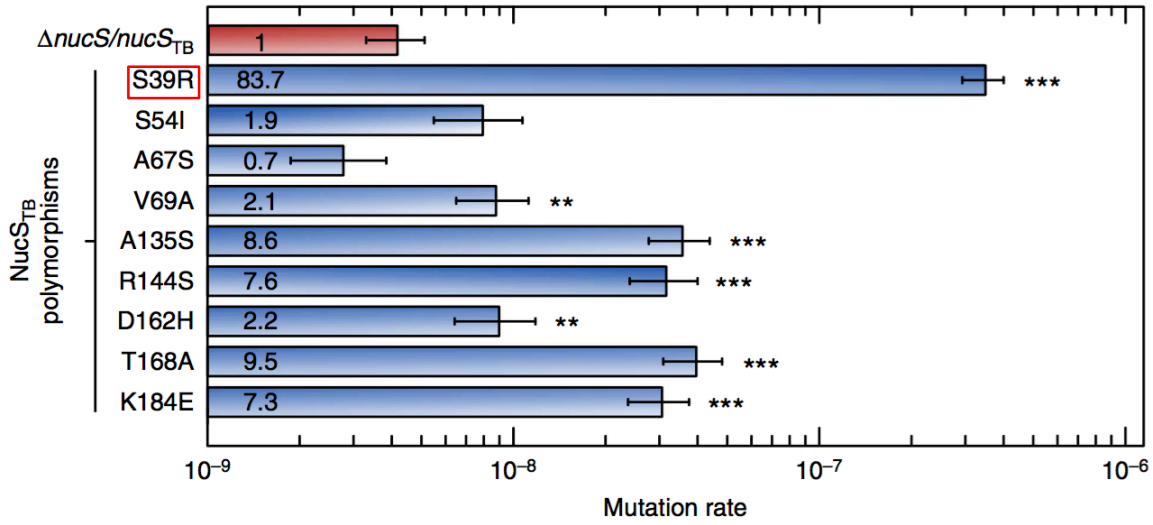
And associated Polymorphysms

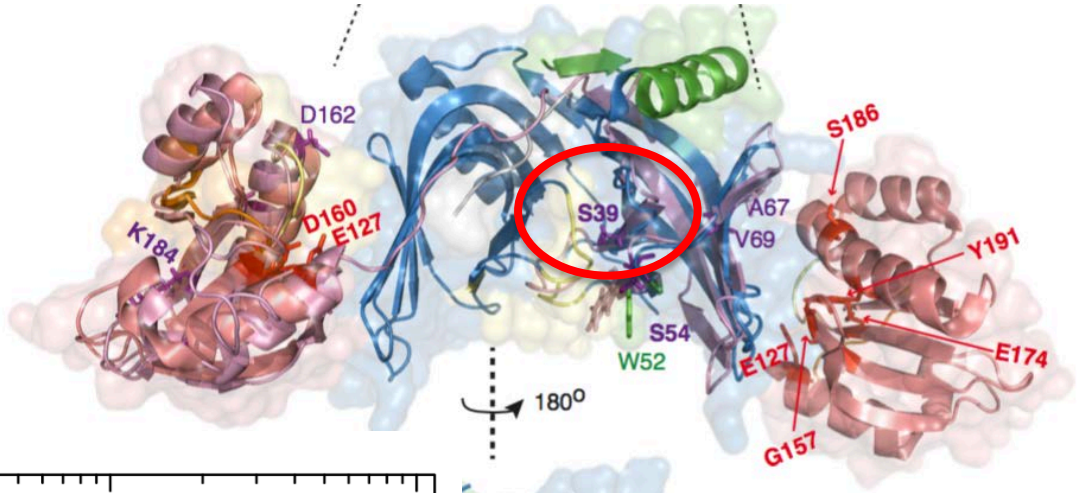


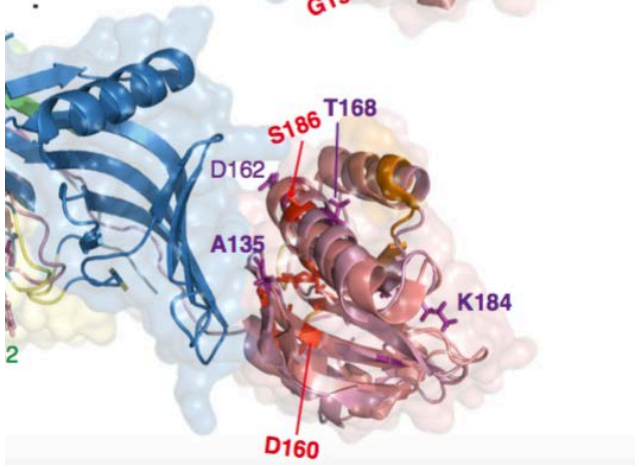
There are hypermutable strains of M. tuberculosis

```
                                                                                KKREPVNWQPPGSKVT
NucS_Mycsme     1 M---RLVIAQCTVDYVGRLTAHLPSARRLLLFKADGSVSVHADDRAYKPLNWMSPPCWVT
NucS_Myctu      1 MSRVRLVIAQCTVDYIGRLTAHLPSARRLLLFKADGSVSVHADDRAYKPLNWMSPPCWLT
                            * * * * **     *     *     * *** R *         * ** I*     *

2VLD_Paby      84 FK--E--NSmISIRRRPYERLEVEIIEPYSLVVFLAEDYEELaltgSEAEmANLIFENPR
NucS_Mycsme    58 EQDTETGVALWVVENKTGEQLRITVEDIEHDSHHELGVDPGLVKDGVEAHLQALLAEHVE
NucS_Myctu     61 EESGGQ-APVWVVENKAGEQLRITIEGIEHDSSHELGVDPGLVKDGVEAHLQALLAEHIQ
                     S A         * *                         *   * **    *  *

2VLD_Paby     140 VIEEGFKPIYREKPIRHGIVDVMGVDKDGNIVVLELKRRKADLHAVSQLKRYVDSLKEEY
NucS_Mycsme   118 LLGAGYTLVRREYPTPIGPVDLLCRDELGRSVAVEIKR-RGEIDGVEQLTRYLELLNRDS
NucS_Myctu    120 LLGEGYTLVRREYMTAIGPVDLLCRDERGGSVAVEIKR-RGEIDGVEQLTRYLELLNRDS
                      *     **    S * **    S*  *  *  * **     H * **A**   *

2VLD_Paby     200 -GENVRGILVAPSLTEGAKKLLEKEGLEFRKLEPP 233
NucS_Mycsme   177 LLAPVAGVFAAQQIKPQARTLATDRGIRCVTLDYD 211
NucS_Myctu    179 VLAPVKGVFAAQQIKPQARILATDRGIRCLTLDYD 213
                   *E*   *       *  *    *     *
```

NucS polymorphisms are associated to MDR





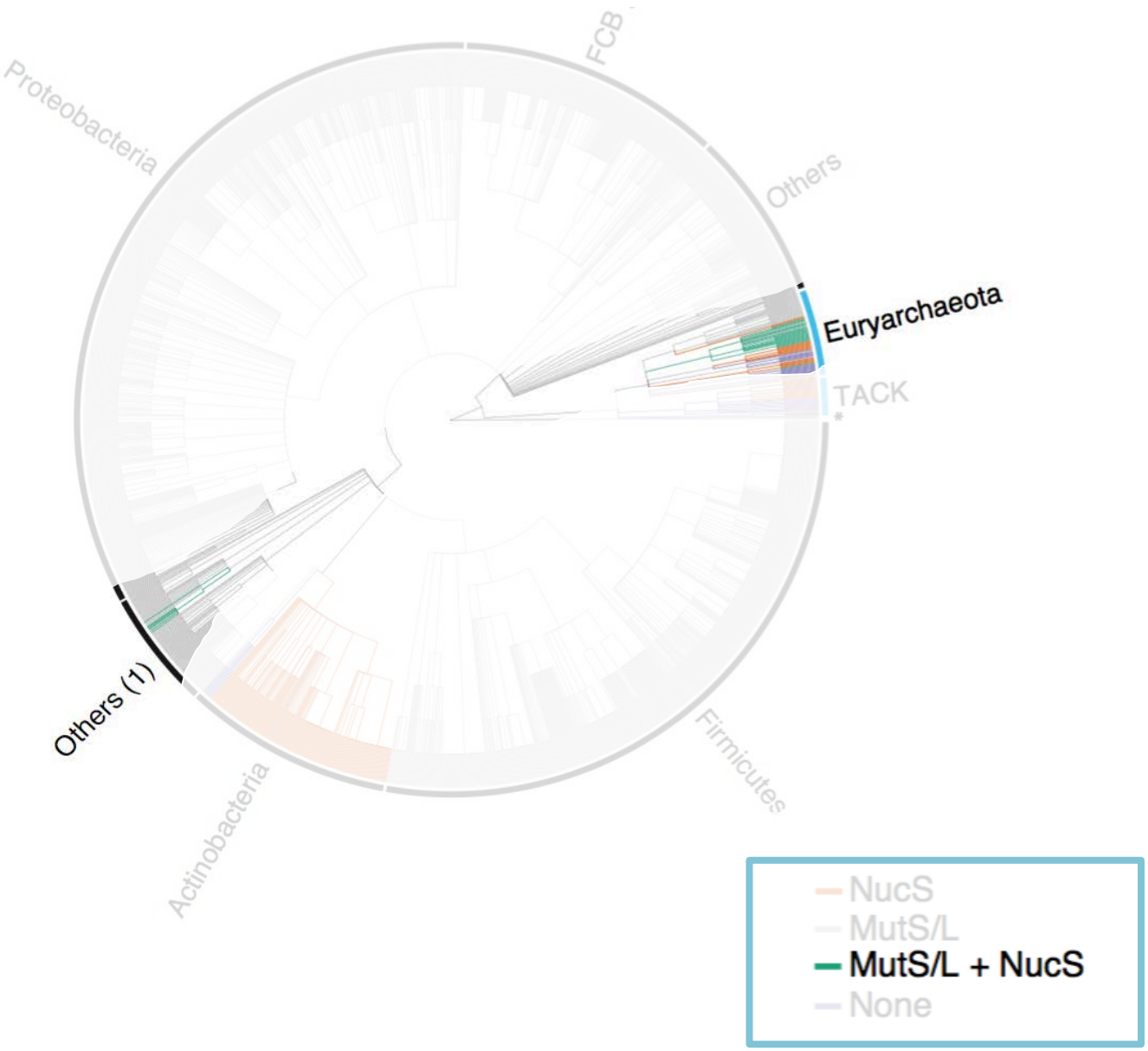


Implications

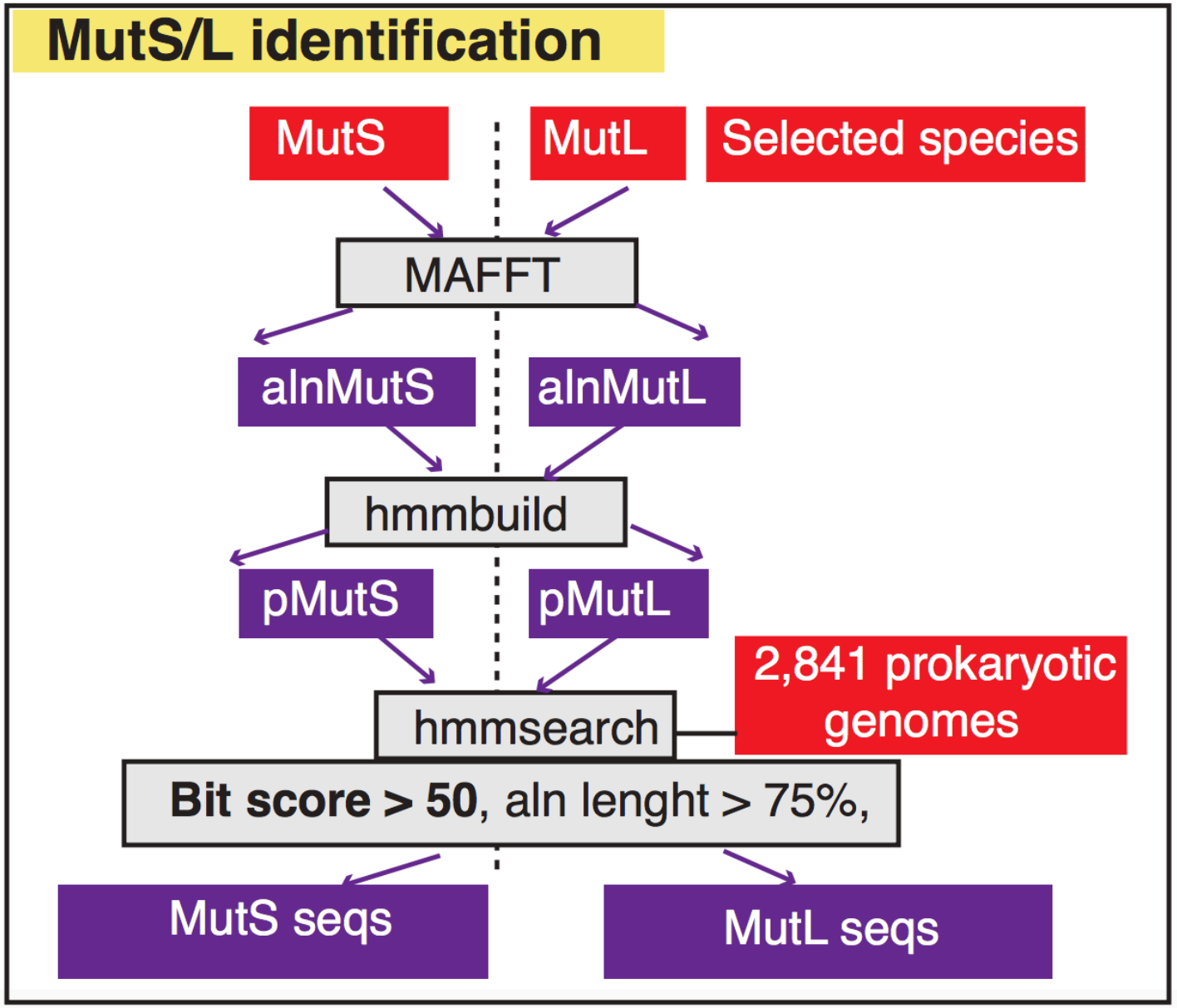
① NucS could server as a therapeutic target to fight MDR strains

① Still **more MMR pathways to unravel**…

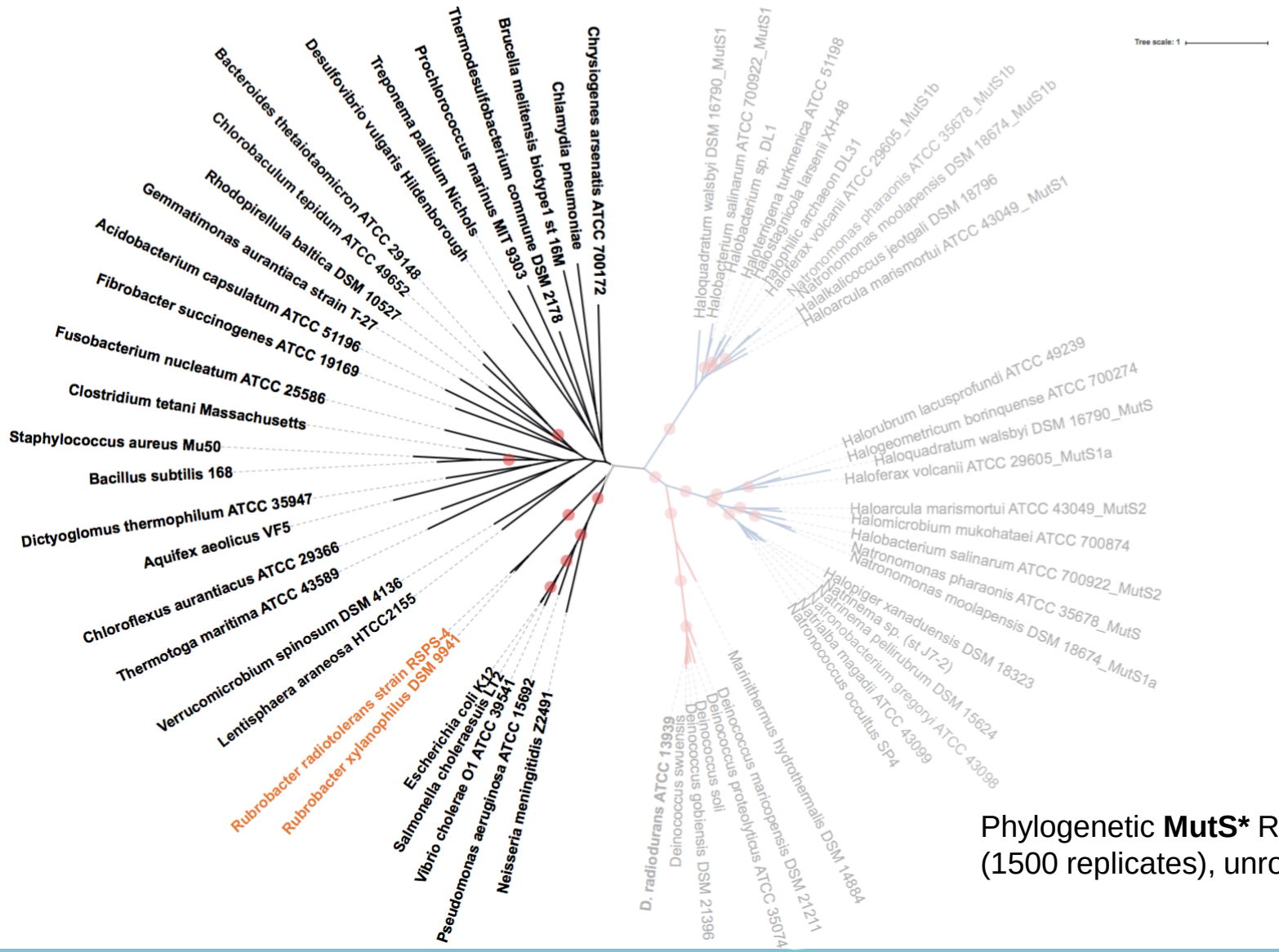
Pathway replacement?



Pathway replacement?



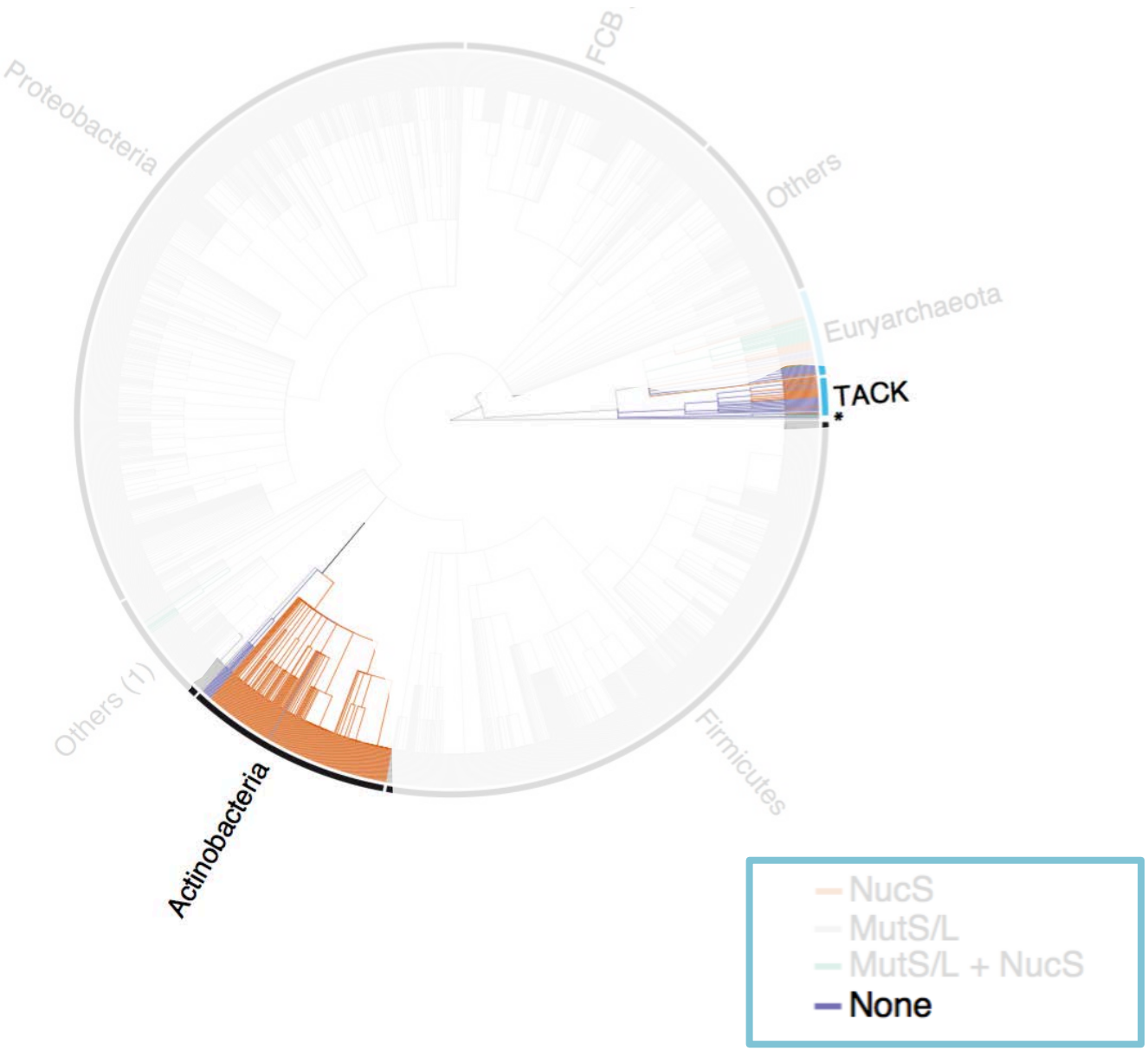
MutS in MutS/L-NucS species **is** archaeal



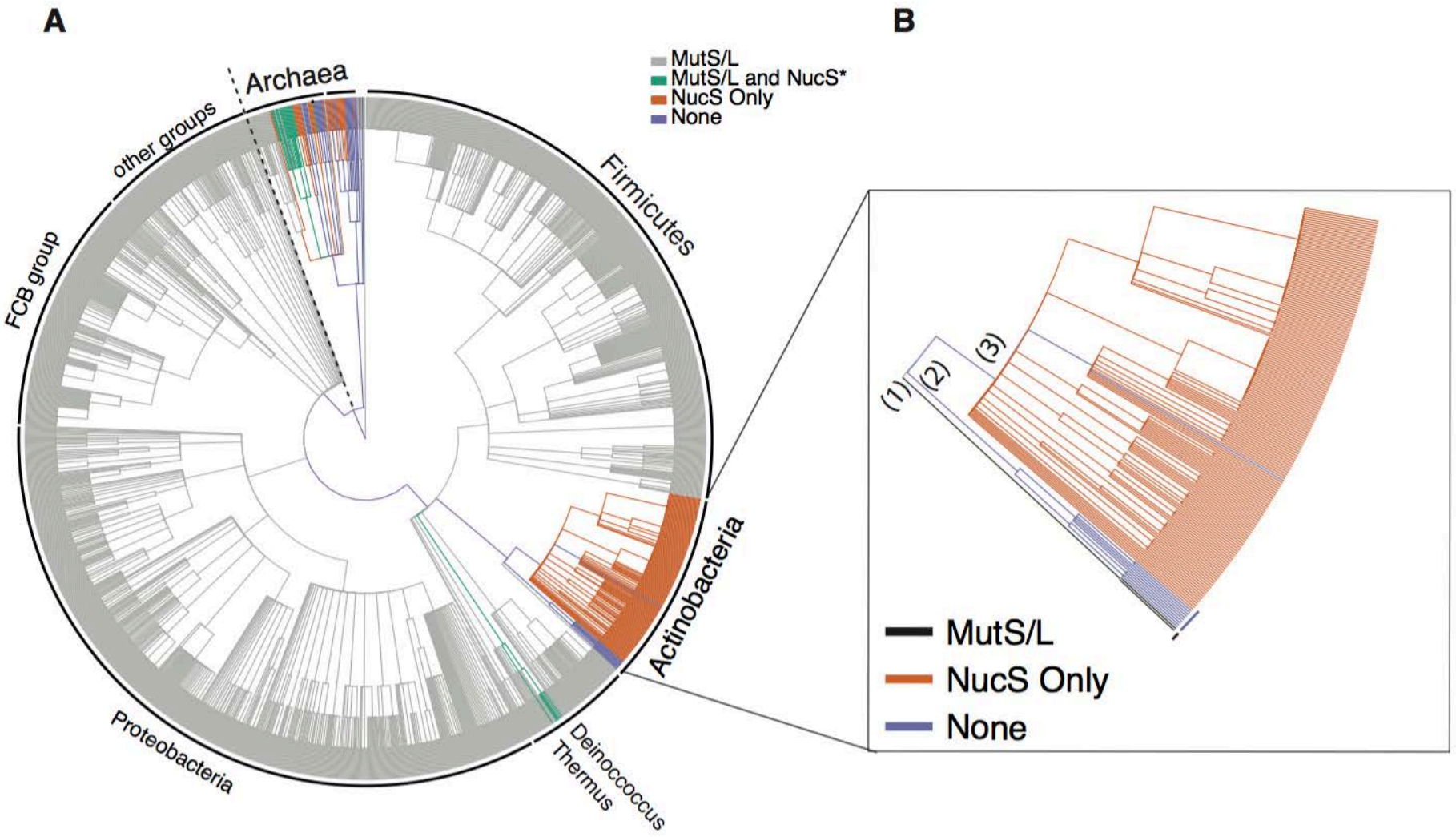
Phylogenetic **MutS*** RAxML (1500 replicates), unrooted

Topology testing: AU approximation by Shimodaira in 2002 (CONSEL). ML trees constrained for strict monophylia [((bacteria1,..bacteriaN),(archaea1,..,archaeaN))]. Monophylia in Bacteria is discarded for MutS (p-val is less than 0.005), supporting our initial observations.

Novel pathways/proteins?



Who is in charge?



We need MORE data from unexplored regions of the pToL


Proteobacteria
FCB
MutS/L
Thank you!
nature COMMUNICATIONS
Altmetric: 302
Views: 912
More detail »
Article | OPEN
A non-canonical mismatch repair pathway in prokaryotes
A. Castañeda-García, A. I. Prieto, J. Rodríguez-Beltrán, N. Alonso, D. Cantillon, C. Costas, L. Pérez-Lago, E. D. Zegeye, M. Herranz, P. Plociński, T. Tonjum, D. García de Viedma, M. Paget, S. J. Waddell, A. M. Rojas, A. J. Doherty & J. Blázquez

NucS: model of emergence

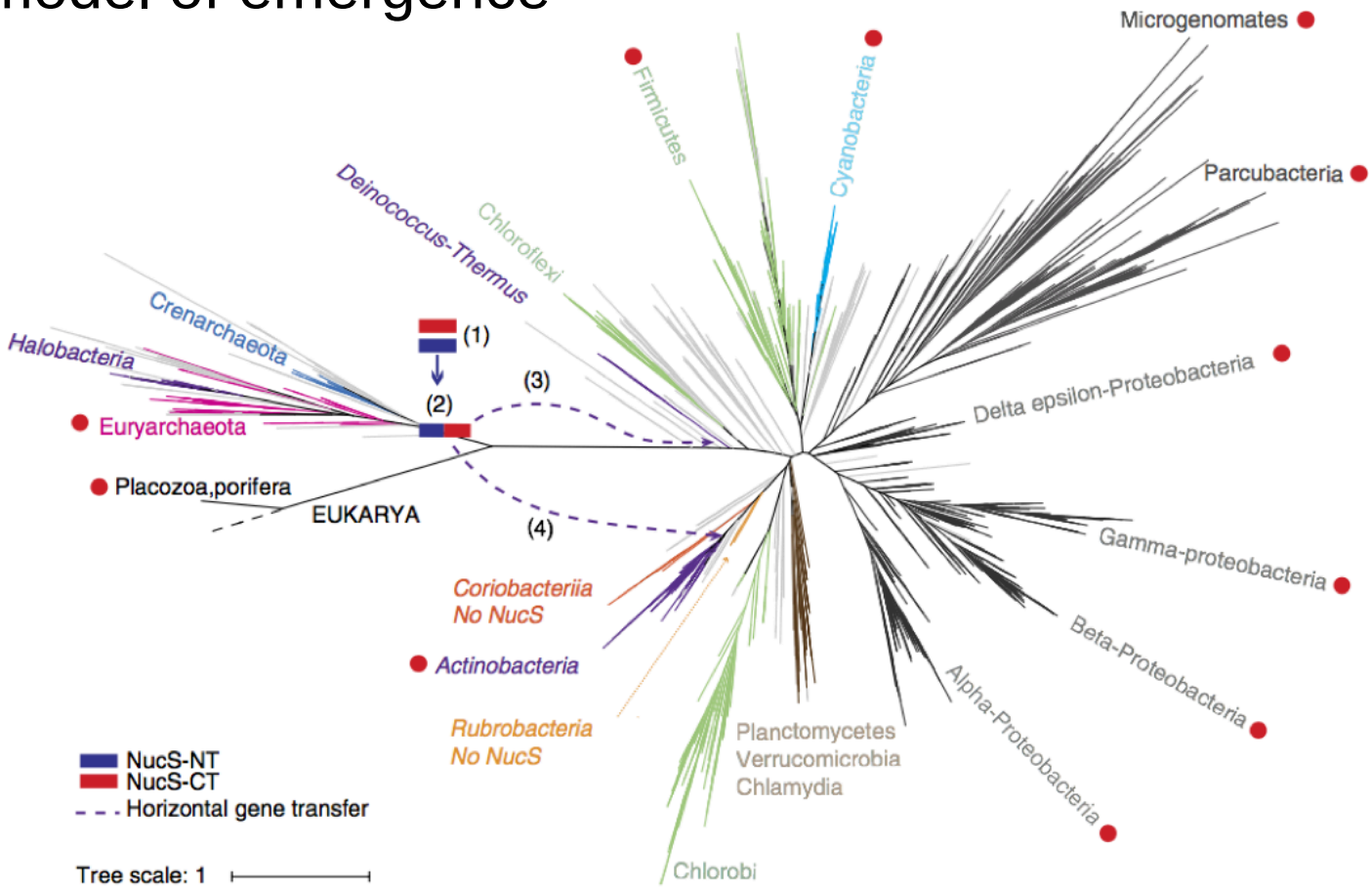
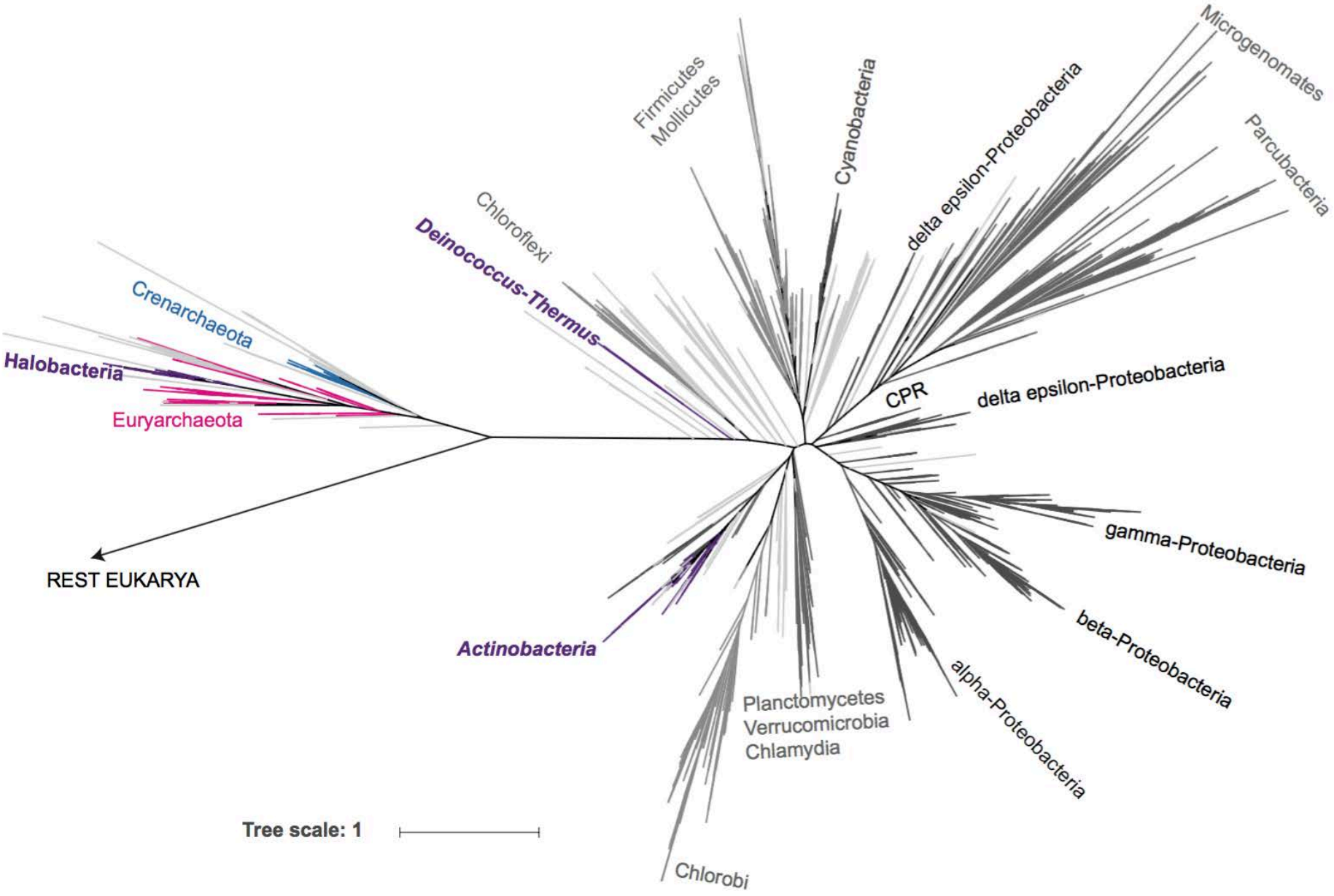


Figure 6 | A model for NucS protein emergence and evolution. The unrooted Tree of Life (available and based on ref. 50) was used to depict the proposed evolutionary history of NucS according to our data. The groups relevant to our model are highlighted. Coloured squares depict the NucS-NT (blue) and NucS-CT (red) terminal regions. This model proposes that NucS has an archaeal origin and emerged as a combination of two independent protein domains with complex evolutionary history. Numbers indicate the steps of the model: Both N-terminal and C-terminal regions likely emerged in the archaeal lineage (1). The CT region was transferred via HGT to very few Eukaryotes and to some Bacteria (main groups with any species having the NucS-CT region are labelled with red circles), where the CT domain combined with other regions outside the context of NucS. In the archaeal lineage, NT and CT regions fused to produce the full NucS (2). NucS expanded in many archaeal groups but was also lost in some others. The full NucS protein was transferred to Bacteria by at least two independent HGT events, one to some Deinoccocus-Thermus species (3) and another to Actinobacteria (4).

Phylogenetic tree ToL (3,083 organisms in total)



Hug, L. A. *et al.* A new view of the tree of life. *Nat Microbiol*. 2016
Concatenated set of 16 ribosomal protein sequences from each organism.

Some arising questions

① Where it NucS?

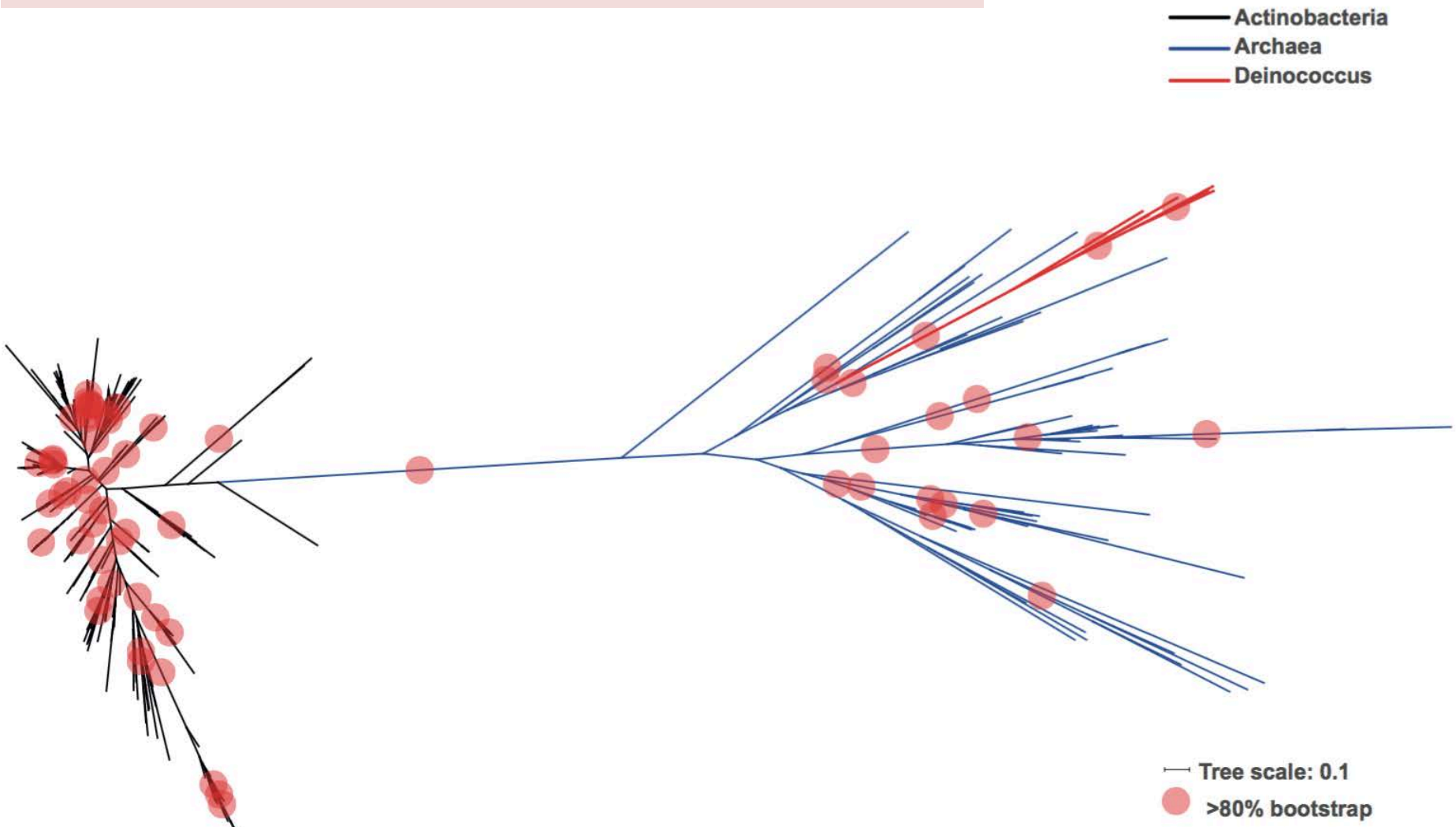
① Where does it come from?

② Are there polymorphisms exploitable for clinical purposes?

③ Are MutS/L also horizontally transferred?

By Phylogenetic analyses

① HGT from Archaea to some Deinococcus group

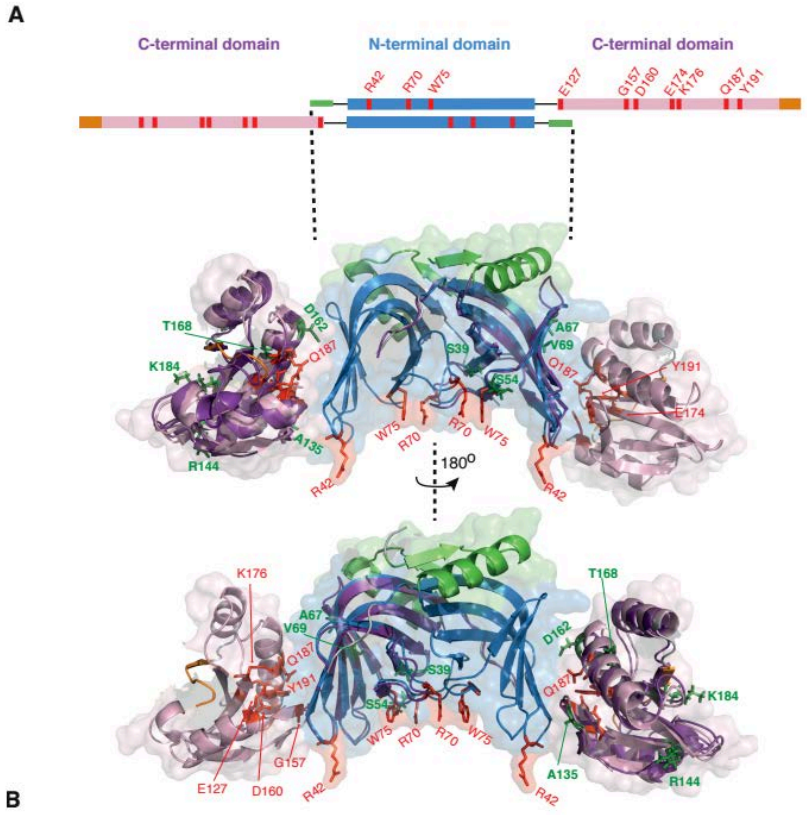


Phylogenetic **FULL** NucS
RAxML (1500 replicates:
378 sequences), unrooted

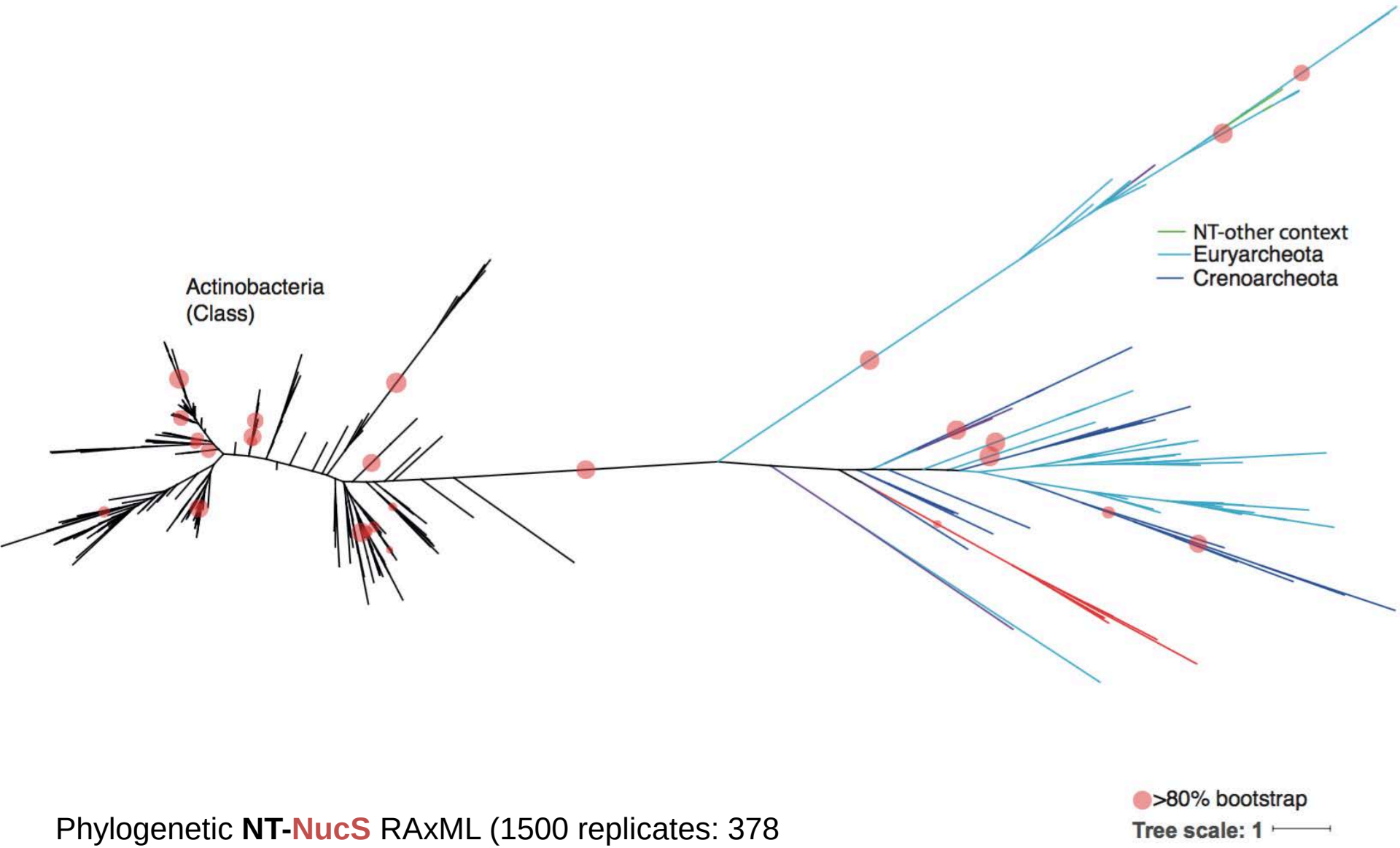
Where does NucS come from?

Archea
Bacteria

Archea
Bacteria
Eukaryota* (3 early metazoans)

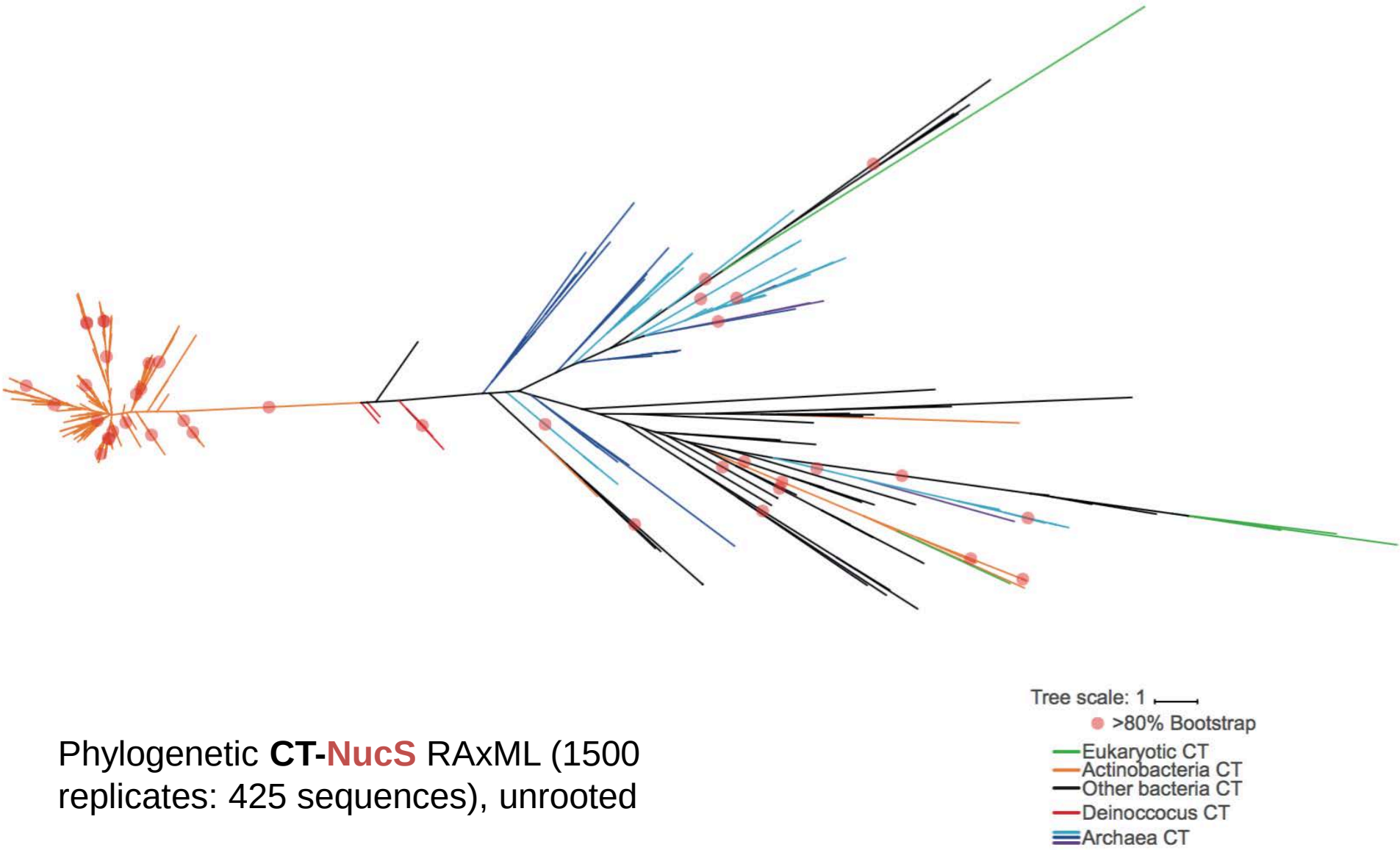


Where does N-terminal of NucS come from?



Phylogenetic **NT-NucS** RAxML (1500 replicates: 378 sequences), unrooted

Where does C-terminal of NucS come from?



Phylogenetic **CT-NucS** RAxML (1500 replicates: 425 sequences), unrooted

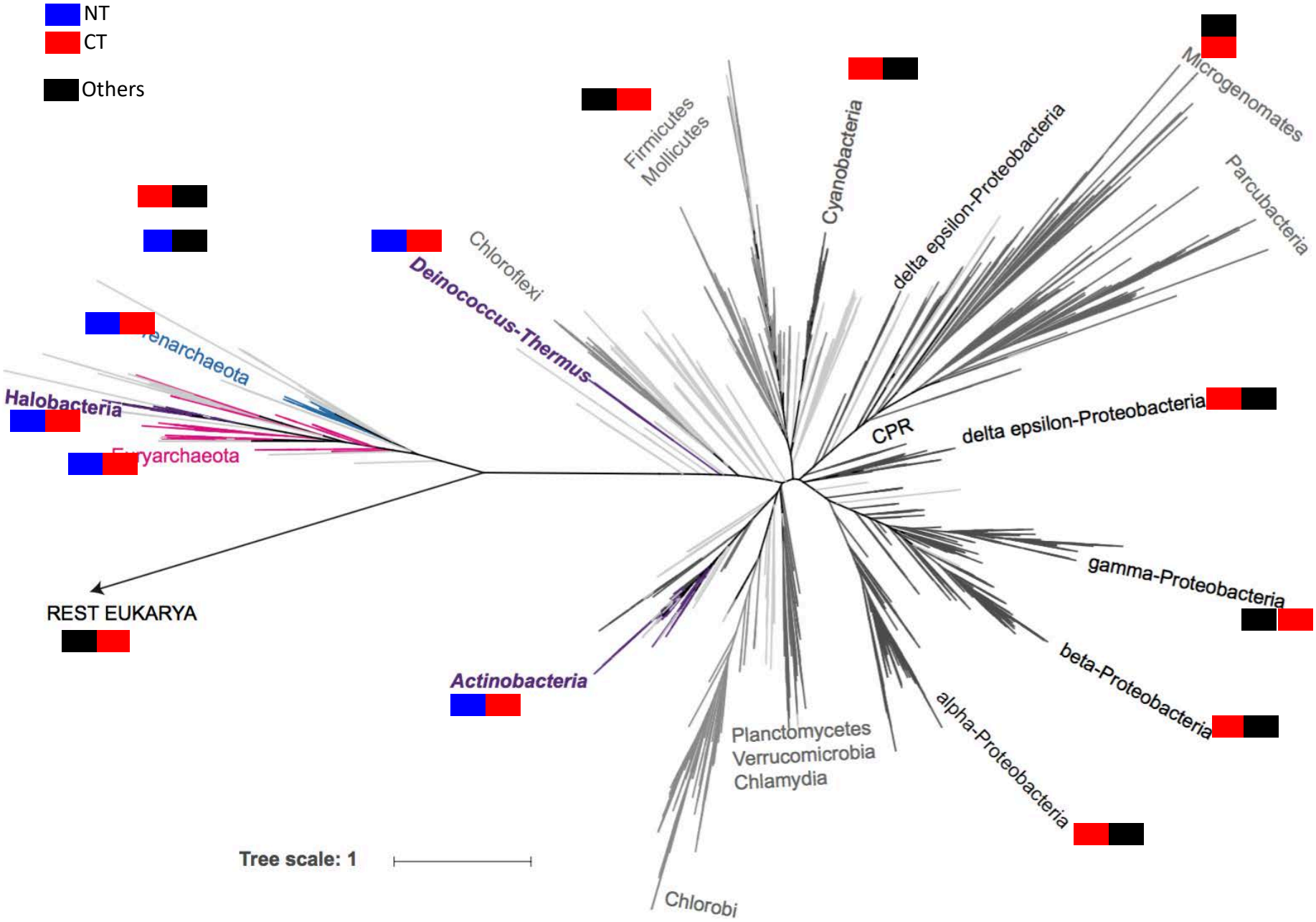
NucS: possibilities of emergence

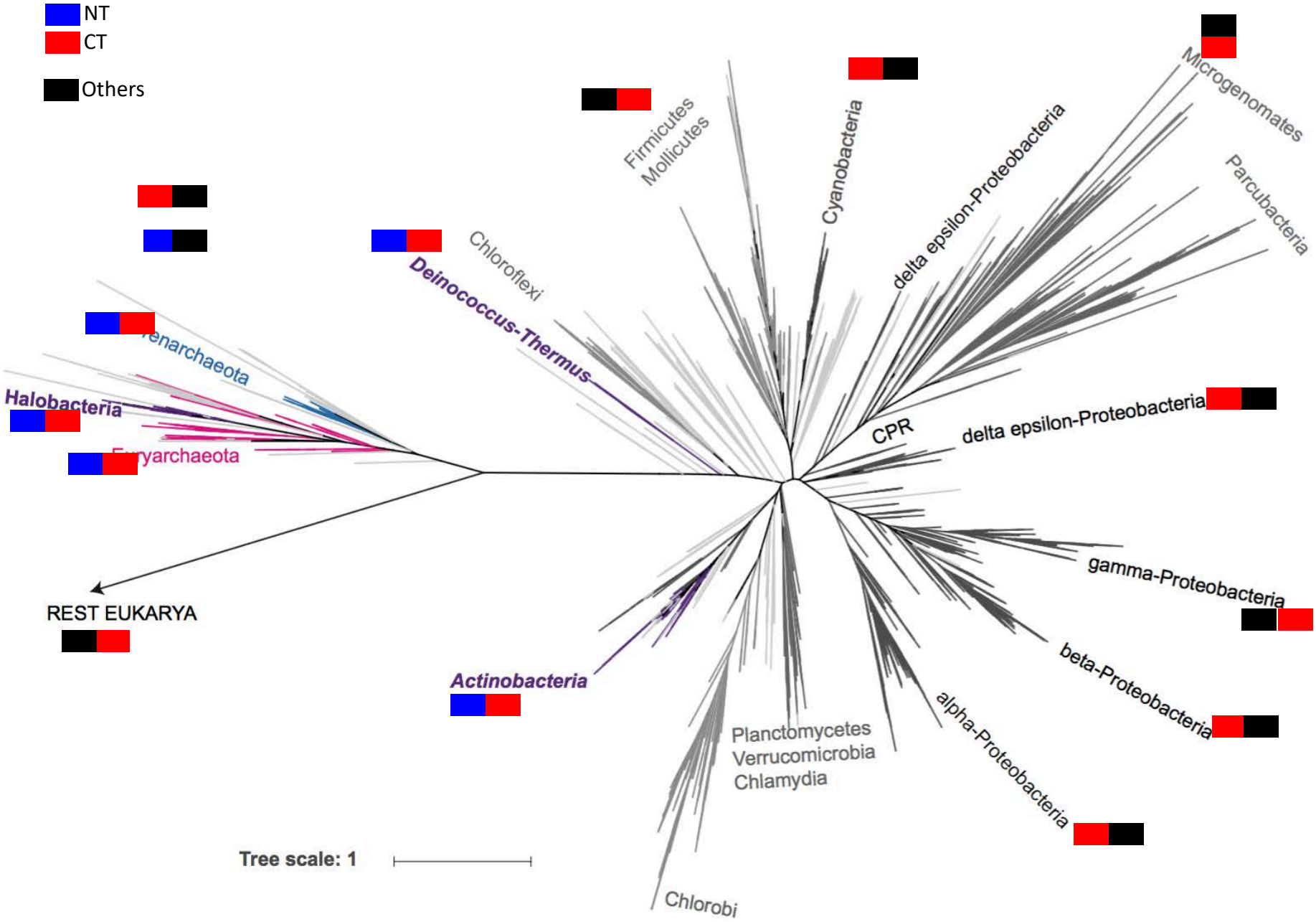
① If full protein in **LUCA**: Massive loses in many lineages and/or organisms. Unlikely

② If full protein in **LBCA**: many loses in bacteria and several HGT to Archaea followed by many loses in Archaea. Unlikely.

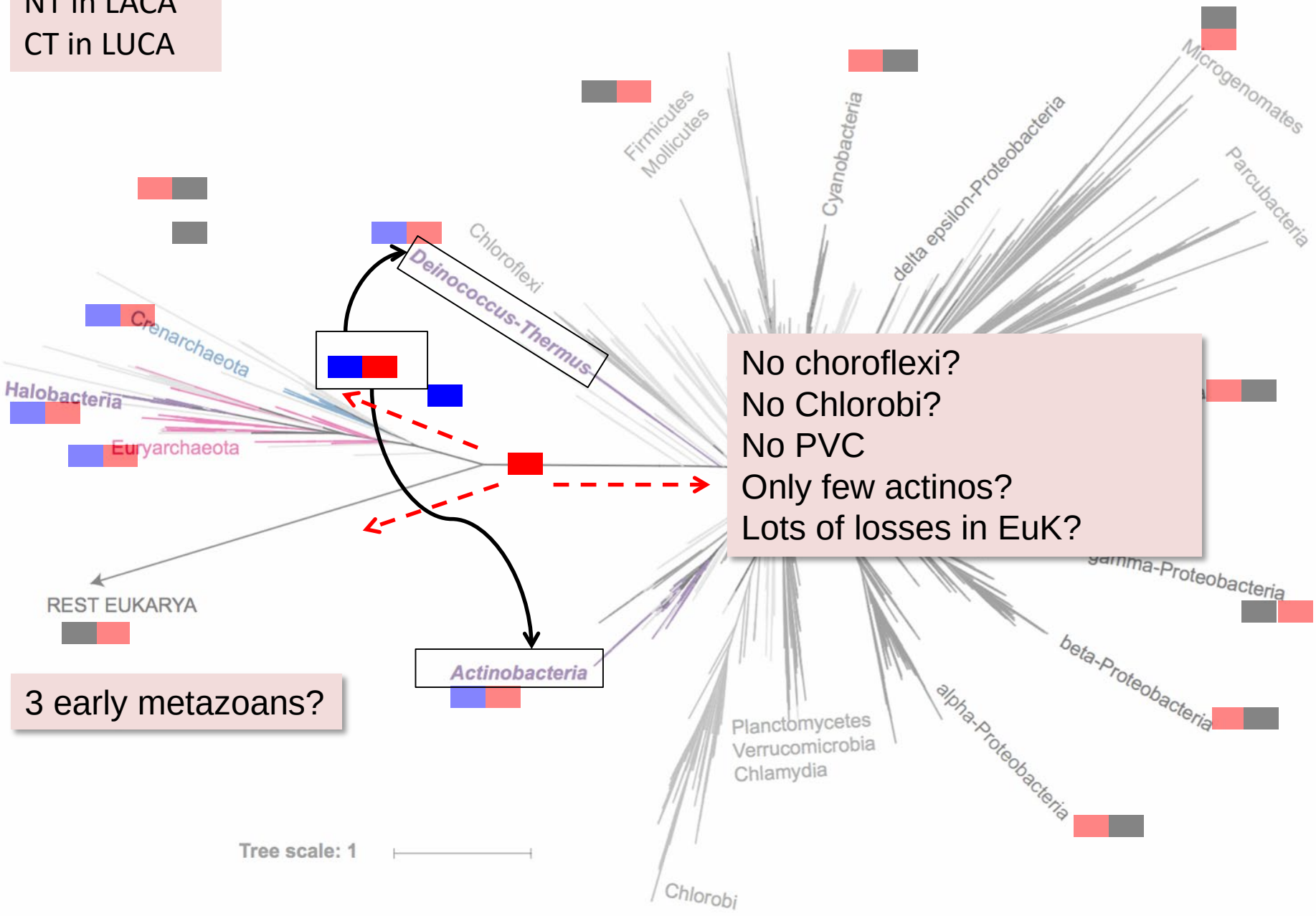
③ If full protein in **LACA**: many loses in Archaea and few HGT to Bacteria. Maybe BUT…

NONE explains domain distributions!


NT
CT
Others
Firmicutes
Mollicutes
Cyanobacteria
Microgenomates
delta epsilon-Proteobacteria
Parcubacteria
Chloroflexi
Deinococcus-Thermus
Halobacteria
Euryarchaeota
CPR
delta epsilon-Proteobacteria
REST EUKARYA
gamma-Proteobacteria
beta-Proteobacteria
Actinobacteria
Planctomycetes
Verrucomicrobia
Chlamydia
alpha-Proteobacteria
Chlorobi
Tree scale: 1


NT
CT
Others
Firmicutes
Mollicutes
Cyanobacteria
Microgenomates
delta epsilon-Proteobacteria
Parcubacteria
Chloroflexi
Deinococcus-Thermus
Halobacteria
Euryarchaeota
CPR
delta epsilon-Proteobacteria
REST EUKARYA
gamma-Proteobacteria
beta-Proteobacteria
Actinobacteria
Planctomycetes
Verrucomicrobia
Chlamydia
alpha-Proteobacteria
Chlorobi
Tree scale: 1

NT in LACA
CT in LUCA



No choroflexi?
No Chlorobi?
No PVC
Only few actinos?
Lots of losses in EuK?

3 early metazoans?

NucS: model of emergence

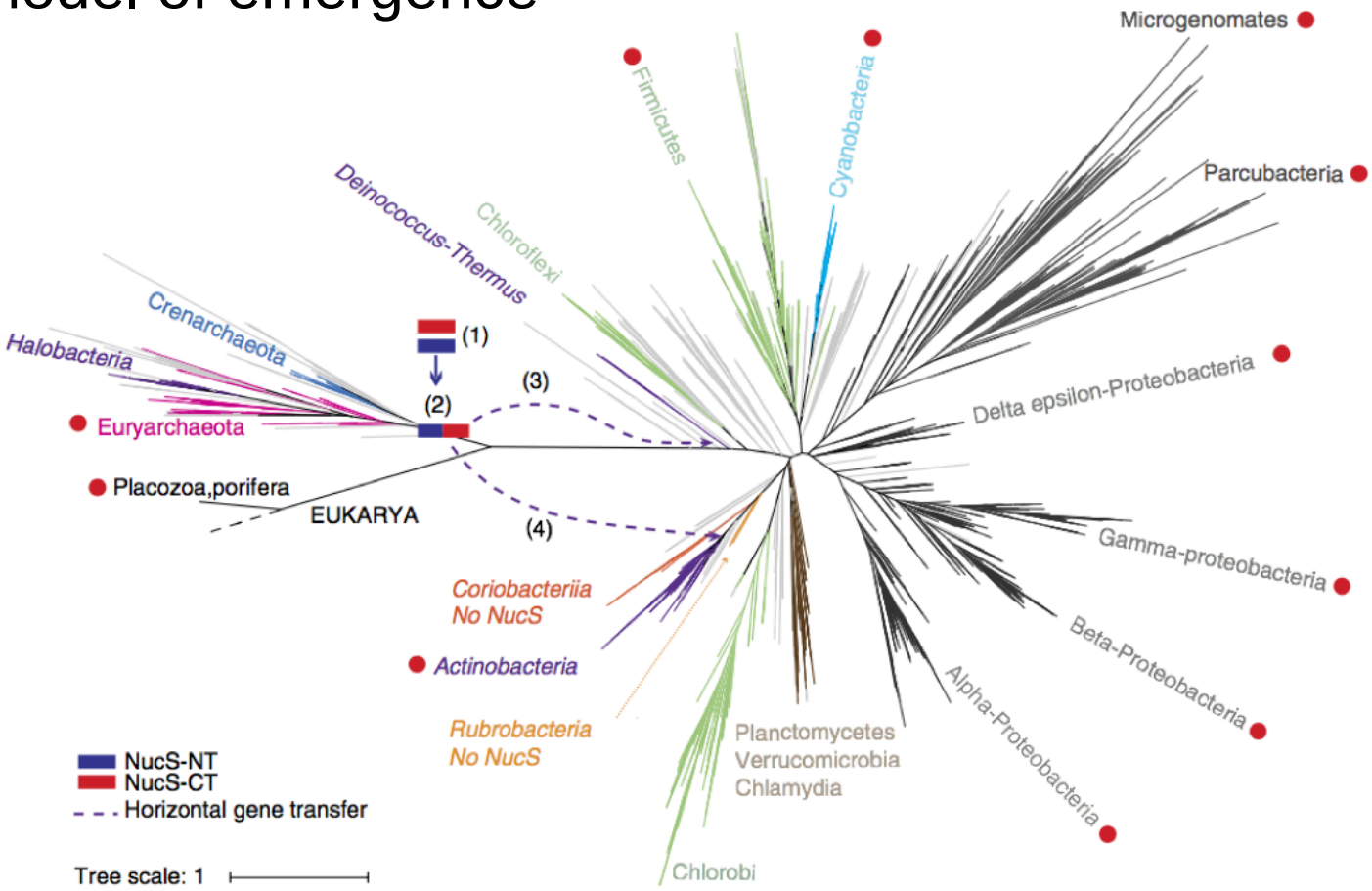
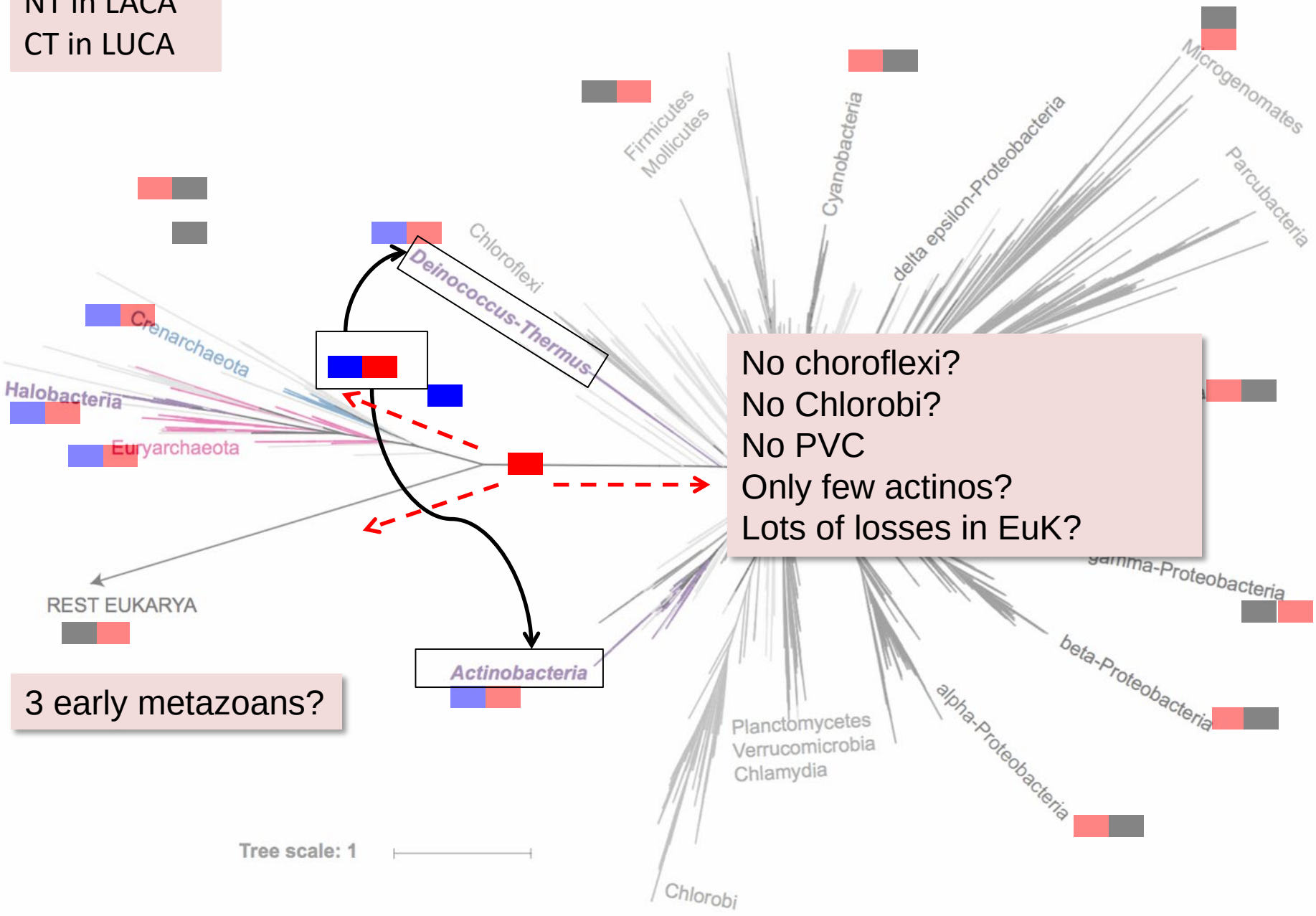


Figure 6 | A model for NucS protein emergence and evolution. The unrooted Tree of Life (available and based on ref. 50) was used to depict the proposed evolutionary history of NucS according to our data. The groups relevant to our model are highlighted. Coloured squares depict the NucS-NT (blue) and NucS-CT (red) terminal regions. This model proposes that NucS has an archaeal origin and emerged as a combination of two independent protein domains with complex evolutionary history. Numbers indicate the steps of the model: Both N-terminal and C-terminal regions likely emerged in the archaeal lineage (1). The CT region was transferred via HGT to very few Eukaryotes and to some Bacteria (main groups with any species having the NucS-CT region are labelled with red circles), where the CT domain combined with other regions outside the context of NucS. In the archaeal lineage, NT and CT regions fused to produce the full NucS (2). NucS expanded in many archaeal groups but was also lost in some others. The full NucS protein was transferred to Bacteria by at least two independent HGT events, one to some Deinoccocus-Thermus species (3) and another to Actinobacteria (4).

NT in LACA
CT in LUCA



No choroflexi?
No Chlorobi?
No PVC
Only few actinos?
Lots of losses in EuK?

3 early metazoans?

Open new questions

① Was **MutS/L/NucS** transferred at the same time to *Deinoccocus*?

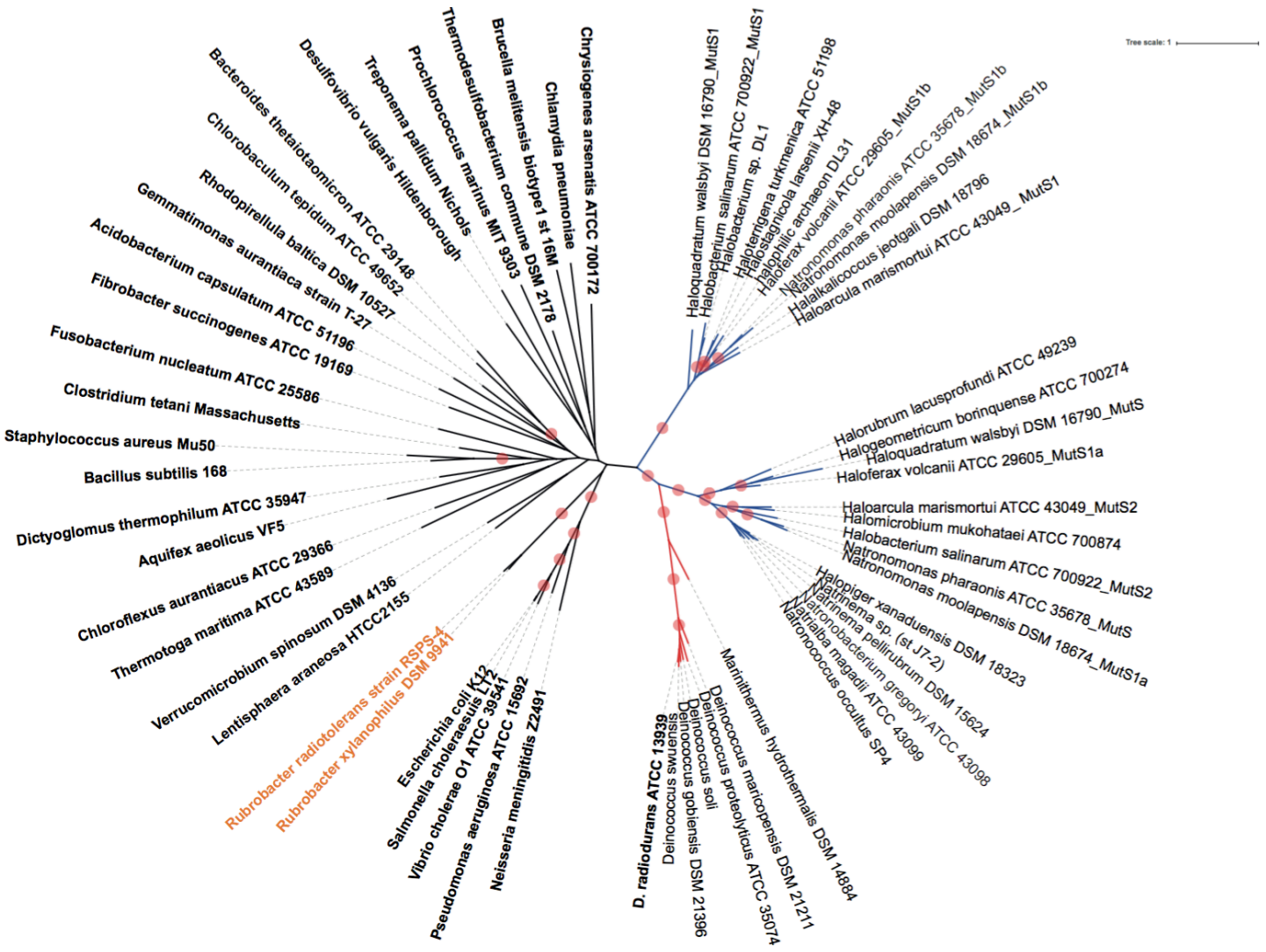
② Did **a**MutS/L take over **b**MutS/L?

③ Deinococcus never had MutS/L and got aMutS/L?*

④ What's going on with those lacking the three?

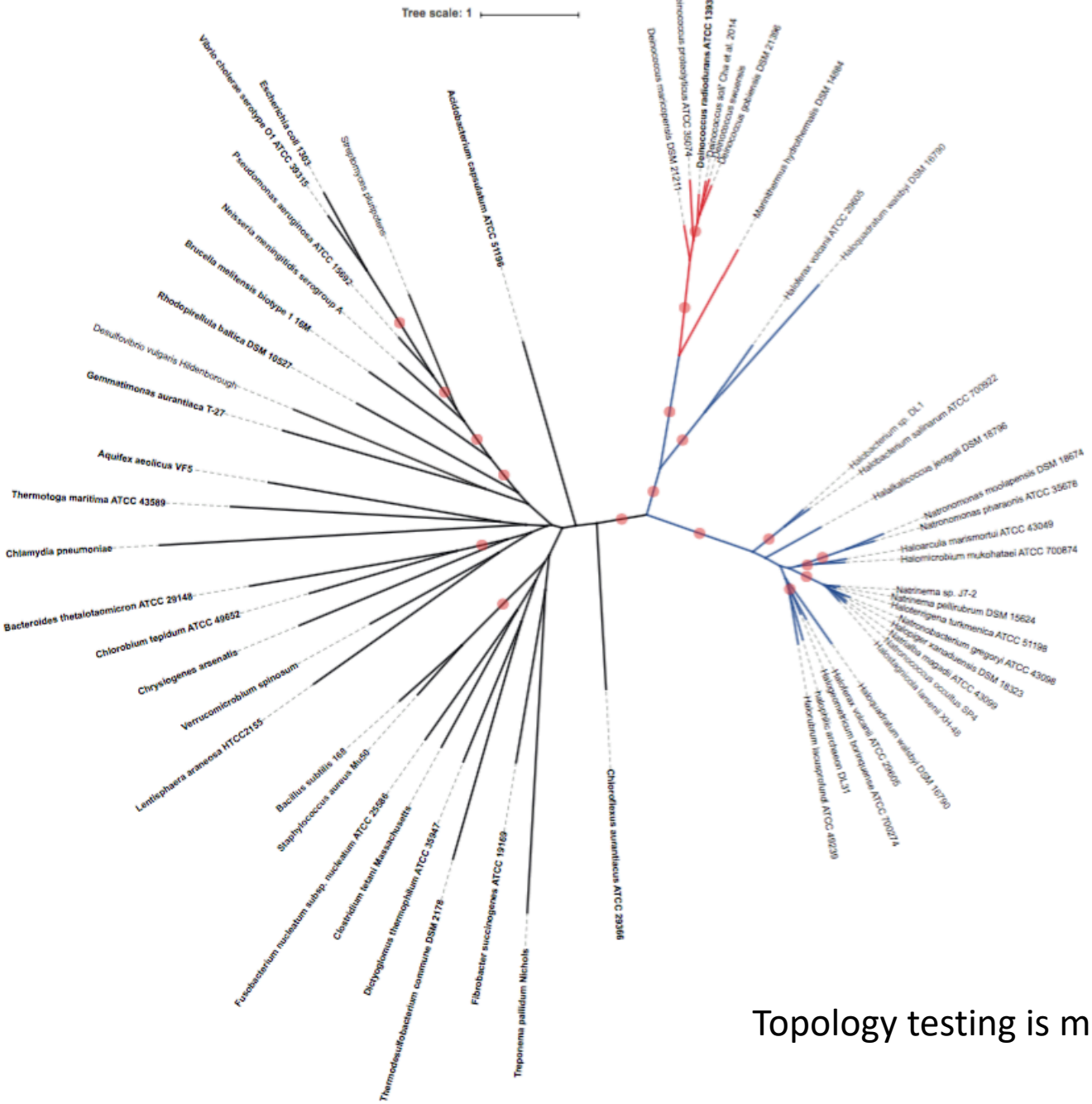
⑤ What is the origin of each independent domain?

Phylogenetic **MutS*** RAxML (1500 replicates), unrooted



Topology testing: AU approximation by Shimodaira in 2002 (CONSEL). ML trees constrained for strict monophylia [((bacteria1,..bacteriaN),(archaea1,..,archaeaN))]. Monophylia in Bacteria is discarded for MutS (p-val is less than 0.005), supporting our initial observations.

Phylogenetic **MutL** RAxML (1500 replicates), unrooted



Topology testing is marginal P=0.05

Open new questions

① Was **MutS/L/NucS** transferred at the same time to *Deinoccocus*?

② Did **a**MutS/L take over **b**MutS/L?

③ Deinococcus never had MutS/L and got aMutS/L?*

④ What's going on with those lacking the three? (Coriiobacteria)

⑤ What is the origin of each independent domain?

Open questions

nature biotechnology

OPEN

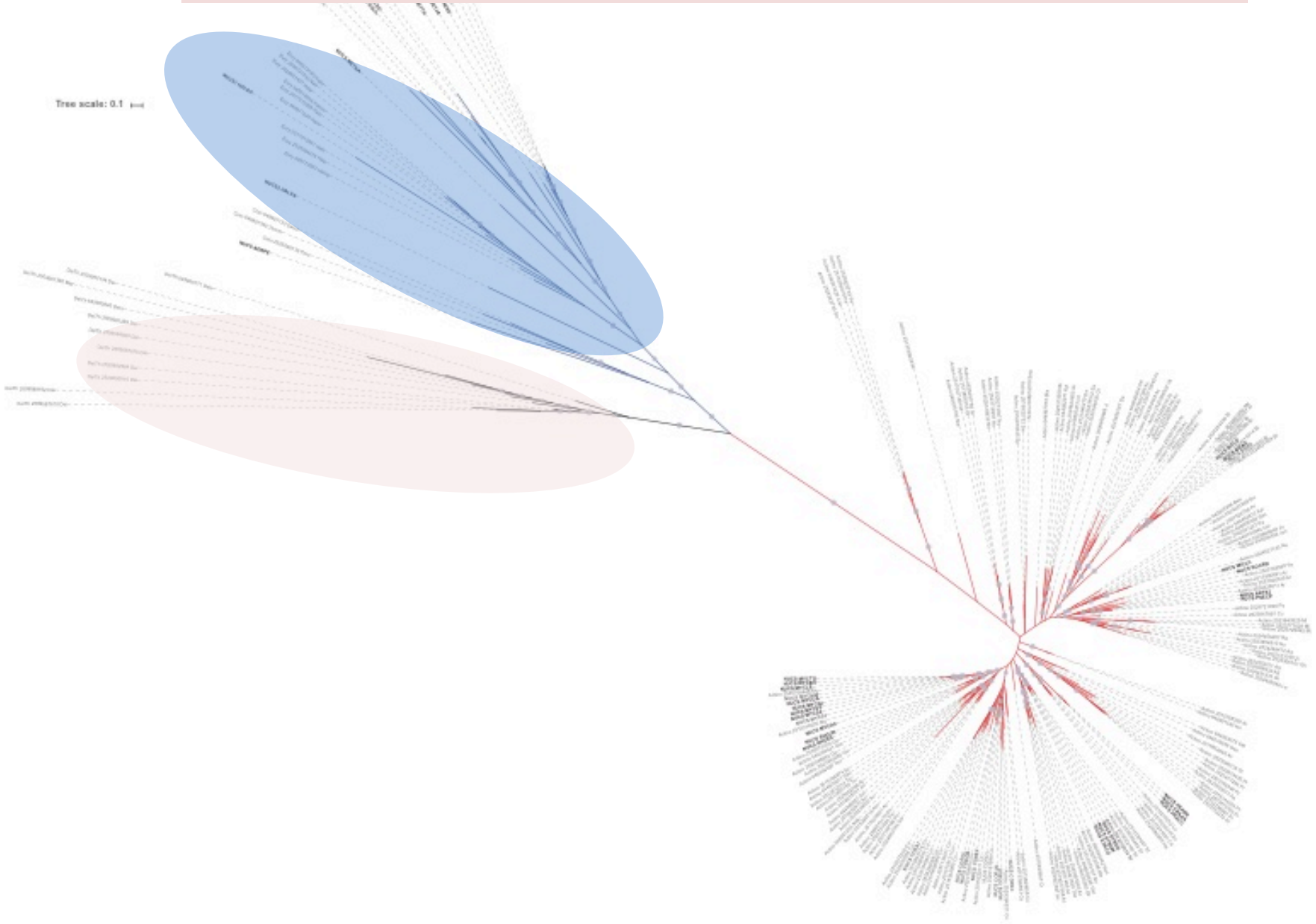
1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life

Supratim Mukherjee[1,10], Rekha Seshadri[1,10], Neha J Varghese[1], Emiley A Eloe-Fadrosh[1], Jan P Meier-Kolthoff[2], Markus Göker[2], R Cameron Coates[1,9], Michalis Hadjithomas[1], Georgios A Pavlopoulos[1], David Paez-Espino[1], Yasuo Yoshikuni[1], Axel Visel[1], William B Whitman[3], George M Garrity[4,5], Jonathan A Eisen[6], Philip Hugenholtz[7], Amrita Pati[1,9], Natalia N Ivanova[1], Tanja Woyke[1], Hans-Peter Klenk[8] & Nikos C Kyrpides[1]

We present 1,003 reference genomes that were sequenced as part of the Genomic Encyclopedia of Bacteria and Archaea (GEBA) initiative, selected to maximize sequence coverage of phylogenetic space. These genomes double the number of existing type strains and expand their overall phylogenetic diversity by 25%. Comparative analyses with previously available finished and draft genomes reveal a 10.5% increase in novel protein families as a function of phylogenetic diversity. The GEBA genomes recruit 25 million previously unassigned metagenomic proteins from 4,650 samples, improving their phylogenetic and functional interpretation. We identify numerous biosynthetic clusters and experimentally validate a divergent phenazine cluster with potential new chemical structure and antimicrobial activity. This Resource is the largest single release of reference genomes to date. Bacterial and archaeal isolate sequence space is still far from saturated, and future endeavors in this direction will continue to be a valuable resource for scientific discovery.

GEBA: Taxonomic diversity increased ~25%

Is NucS distribution supported in GEBA?



Some arising questions

① Where it NucS?

① Where does it come from?

② Are there polymorphisms exploitable for clinical purposes?

Jesús Blázquez, CNB-CSIC (Madrid, ES)

Focus on Actinobacteria. . .

Reported that *Mycobacterium tuberculosis* **does NOT** have hypermutable phenotypes

- Acquires Ab resistance exclusively trough chromosomal mutations (Muller et al, 2013; Ford et al, 2013)
- Presents variability in mutation rates among strains (Ford et al, 2013)
- Lacks MutS/L

But it has MDR profiles

GenomeID/ name	Resistance profile	Lineage	Origin
CDC1551	Susceptible	4	North America
TKK_02_0079	MDR	4	South Africa
MTB_N1057	Susceptible	4	South Asia
KT-0040	Susceptible	2	S. Korea (Broad Inst)
ERR036236	Susceptible	1	Unknown
BTB 04-388	MDR	3	Sweden (Broad Inst)
BTB 07-246	MDR	4	Sweden (Broad Inst)
TKK_03_0044	Susceptible	4	South Africa
HN2738	Unknown	Unknown	Unknown (Broad Inst)
MTB_X632	MDR	4	Central America

1,600 clinical *M. tuberculosis* strains*, 9 SNPs

With SNPs in NucS

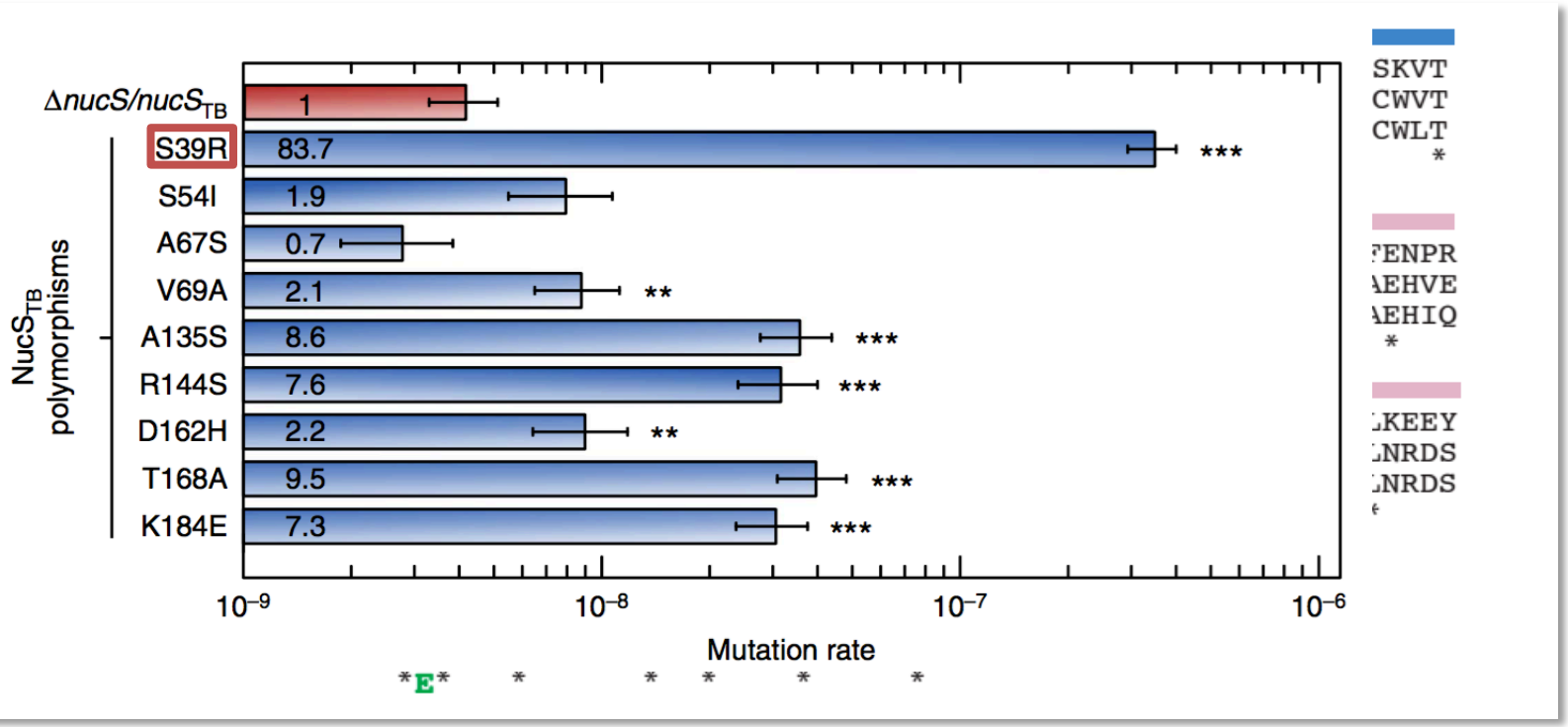
*Ensembl bacteria

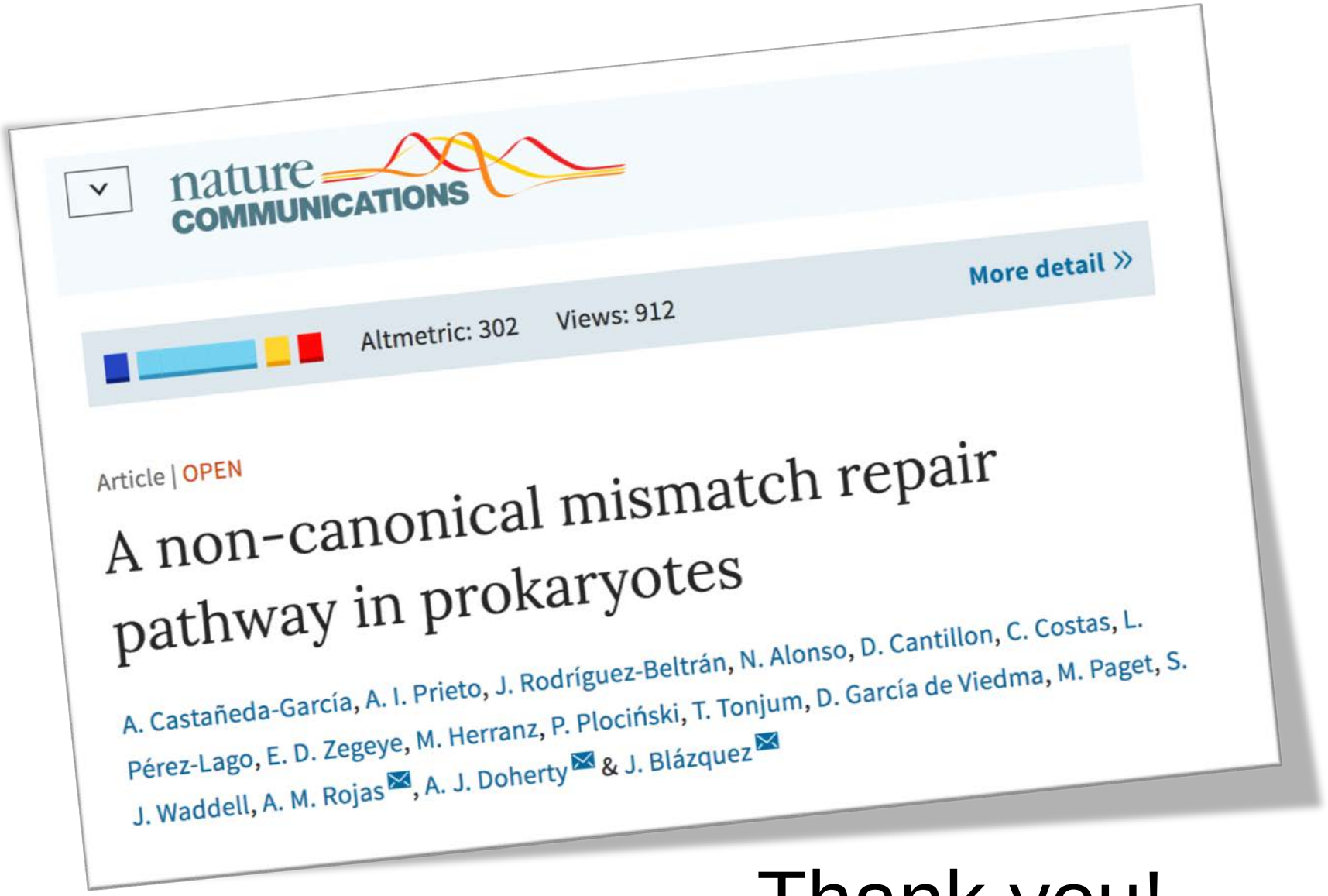
Polymorphysms

Clinical strains *M. tuberculosis*

GenomeID/ name	Polymorphism	Resistance profile	Lineage	Origin
CDC1551	WT	Susceptible	4	North America
TKK_02_0079	S39R	MDR	4	South Africa
MTB_N1057	S54I	Susceptible	4	South Asia
KT-0040	A67S	Susceptible	2	S. Korea (Broad Inst)
ERR036236	V69A	Susceptible	1	Unknown
BTB 04-388	A135S	MDR	3	Sweden (Broad Inst)
BTB 07-246	R144S	MDR	4	Sweden (Broad Inst)
TKK_03_0044	D162H	Susceptible	4	South Africa
HN2738	T168A	Unknown	Unknown	Unknown (Broad Inst)
MTB_X632	K184E	MDR	4	Central America

There are hypermutable strains of M. tuberculosis




nature COMMUNICATIONS
Altmetric: 302
Views: 912
More detail »
Article | OPEN
A non-canonical mismatch repair pathway in prokaryotes
A. Castañeda-García, A. I. Prieto, J. Rodríguez-Beltrán, N. Alonso, D. Cantillon, C. Costas, L. Pérez-Lago, E. D. Zegeye, M. Herranz, P. Plociński, T. Tonjum, D. García de Viedma, M. Paget, S. J. Waddell, A. M. Rojas, A. J. Doherty & J. Blázquez

Thank you!