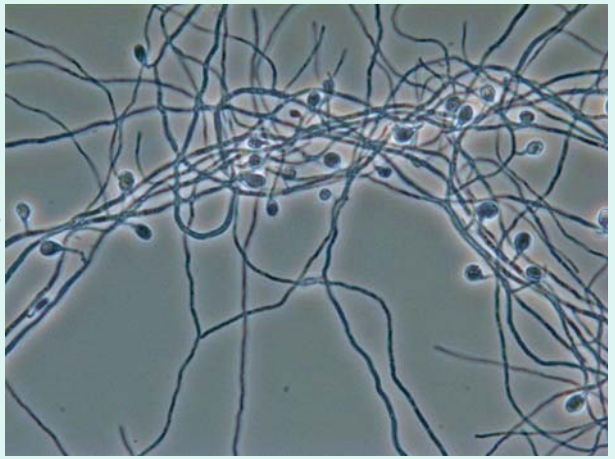
Elucidation of the *Frankia* EAN1pec Genome Sequence Shows Gene Expansion and Metabolic Versatility

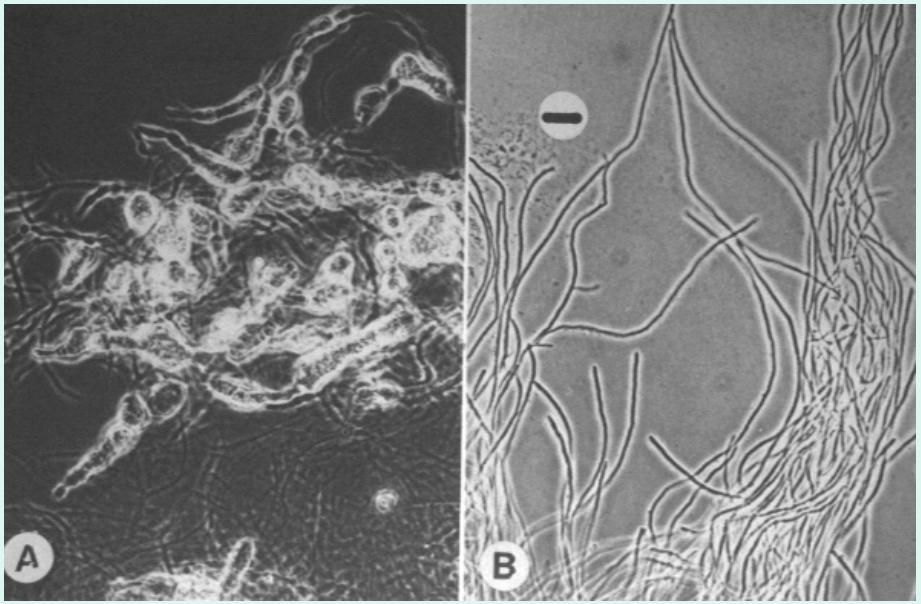
Louis S Tisa[1], James M. Niemann[1], Tania Rawnsley[1], Teal Furnholm[1], David R Benson[2], Gary B. Smejkal[4], Pascal Lapierre[2], J.Peter Gogarten[2], Ying Huang[2], Juliana Mastronunzio[2], Carla A. Bassi[2], Philippe Normand[5], M. Pilar Francino[3], and Paul Richardson[3]

[1]U. New Hampshire, Durham, NH, USA; [2]U Connecticut , Storrs, CT, USA; [3] JGI, Walnut Creek, CA, USA, [4] Pressure Biosciences, Inc, Bridgewater, MA, USA, [5] niversité Lyon, Villeurbanne, France

Frankia



- Member of the Actinomycetales
- Hyphal bacteria
  - 67-72% GC
  - generation time 24-48
- Structures
  - vesicles
  - spores

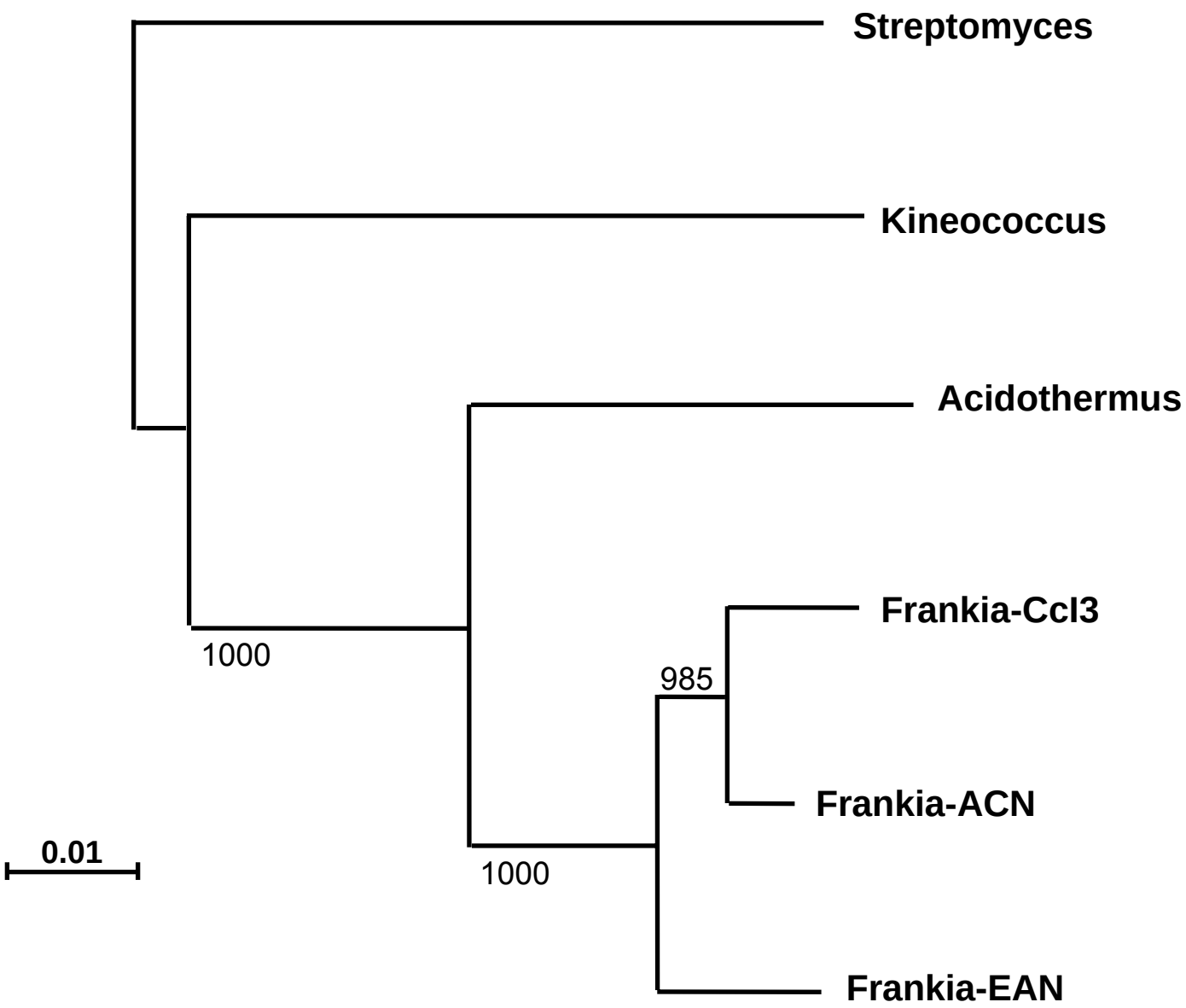


We are interested in developing Genetic and Genomic tools to elucidate the biology of actinorhizal symbiosis

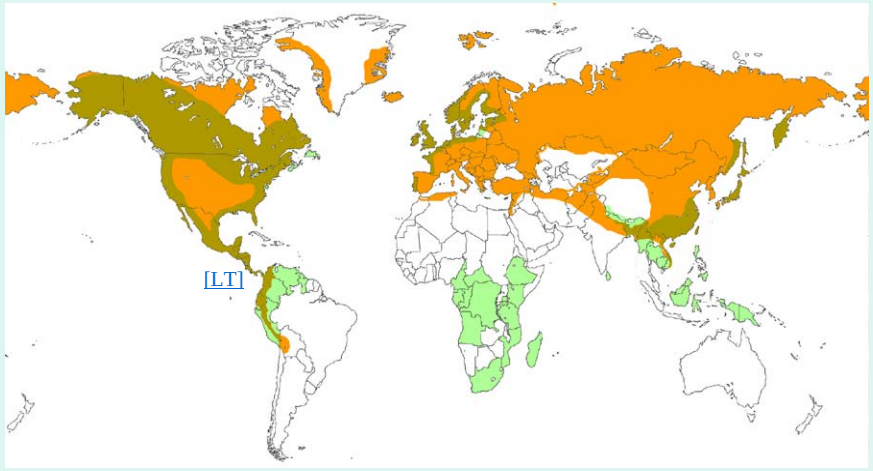
- Our approach
  - Identify selectable markers
  - Develop mutagenesis protocols for mutant selection including transposon systems
  - Develop transformation conditions
  - Establish background for genome sequence
- Get a Genome Sequenced

Why Sequence Ccl3 and EAN1pec?

- Both strains grow well
- Prototrophs
- Knowledge base: physiology, biochemistry, and ecology
- Ability to differentiate (hyphae, spores and vesicles)
- Identification of Intrinsic Markers: antibiotic and metal resistance
- Ability to produce secondary metabolites

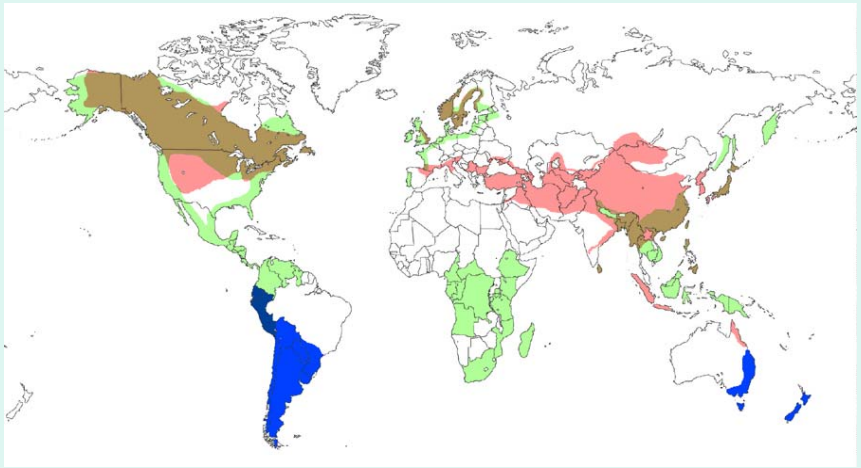

Streptomyces
Kineococcus
Acidothermus
Frankia-CcI3
Frankia-ACN
Frankia-EAN
1000
985
1000
0.01

Present day native distribution of actinorhizal plant hosts.



a, ACN: Betulaceae (orange) Myricaceae (green) and their overlap (khaki).

c, Ccl3: *Casuarina* and *Allocasuarina* of the Casuarinaceae (light blue).



b, EAN :Elaeagnaceae (pink), Myricaceae (green) Rhamnaceae (blue, Tribe Colletieae in South America, Australia and New Zealand). Areas of overlap are brown and dark blue).

What differences are there between Ccl3 and EAN1pec?

Ccl3

- Metabolism
- Member of Group II
- Narrow Host range
- Markers: Kan^R,Gen^R,Kas^R,Nal^R AsO_4^{3-}

EAN1pec

- Diverse metabolism
- Member of Group III (globally distributed)
- Broader Host range Markers:

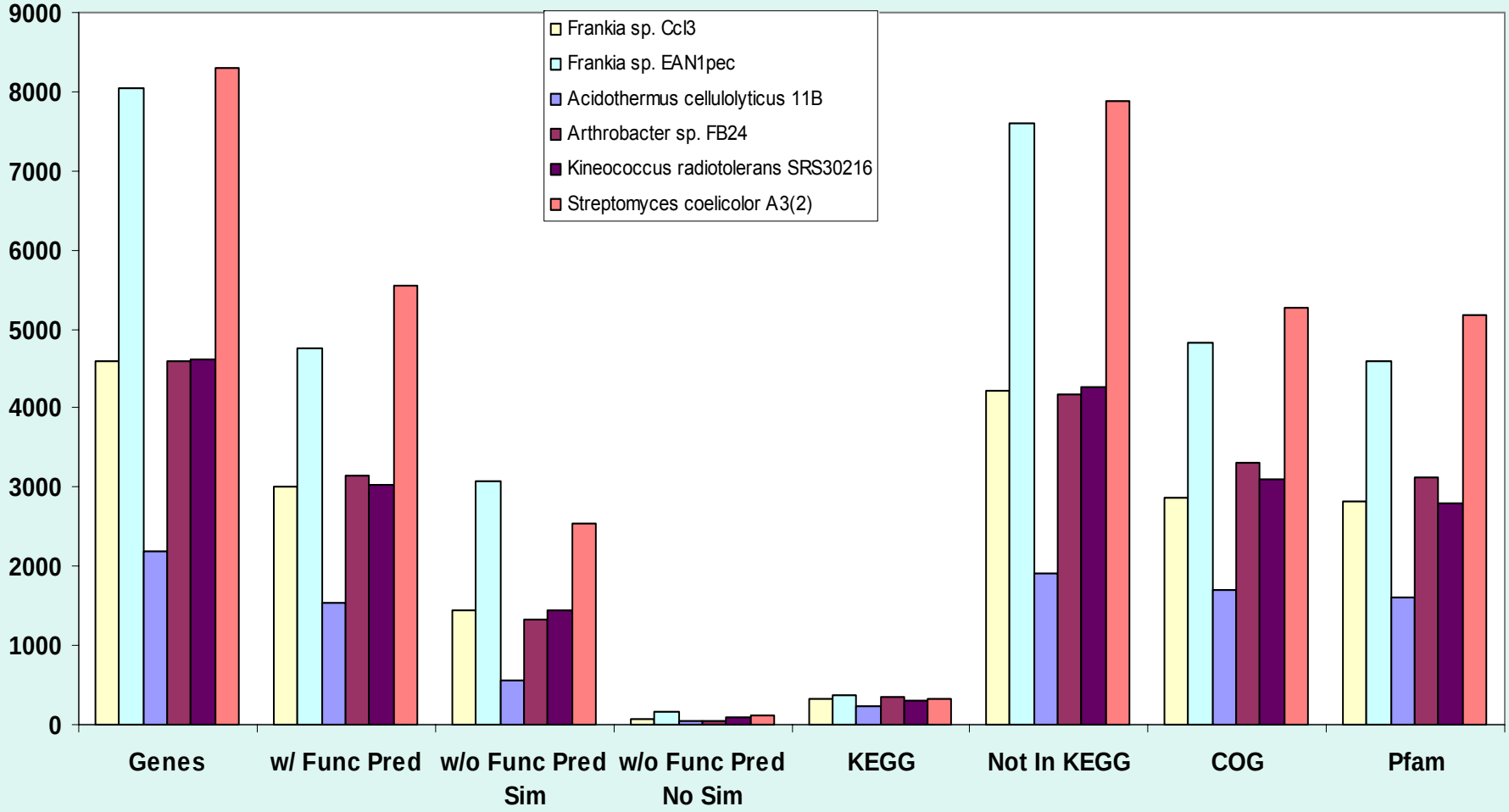
  Nov^R, Lin^R,Kas^R,Nal^R, AsO_4^{3-},Pb^{2+} and CrO_4^{2+}
- Limited genetics

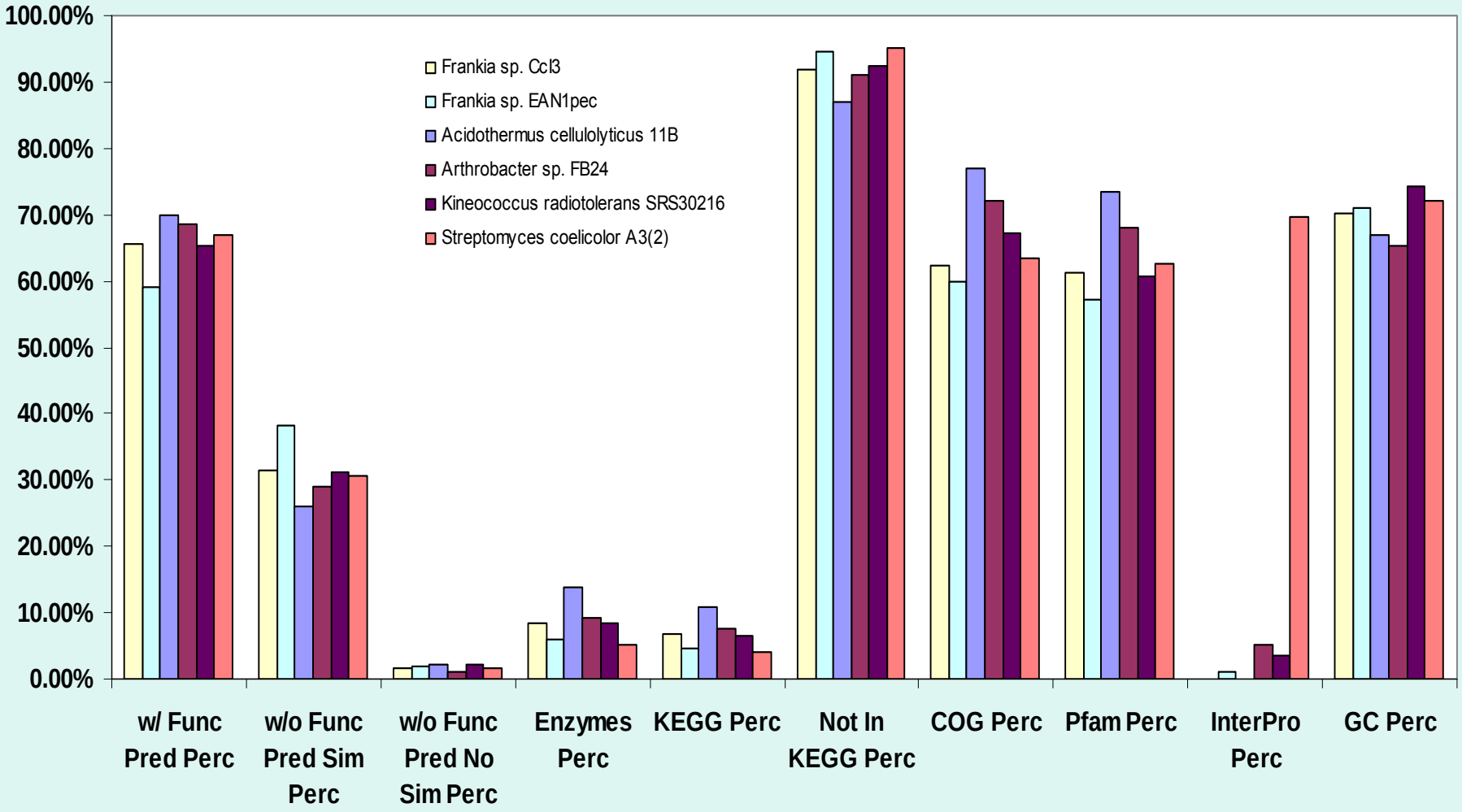
General Aspects of the Project

	Ccl3	EAN1pec
Length	5,433,628	9,081,415
G + C Content	70.1	70.9
Total No. of ORF	4841	7993
rRNA	2	3
tRNA	46	47
Contigs*	1	1
Scaffold*	1	1

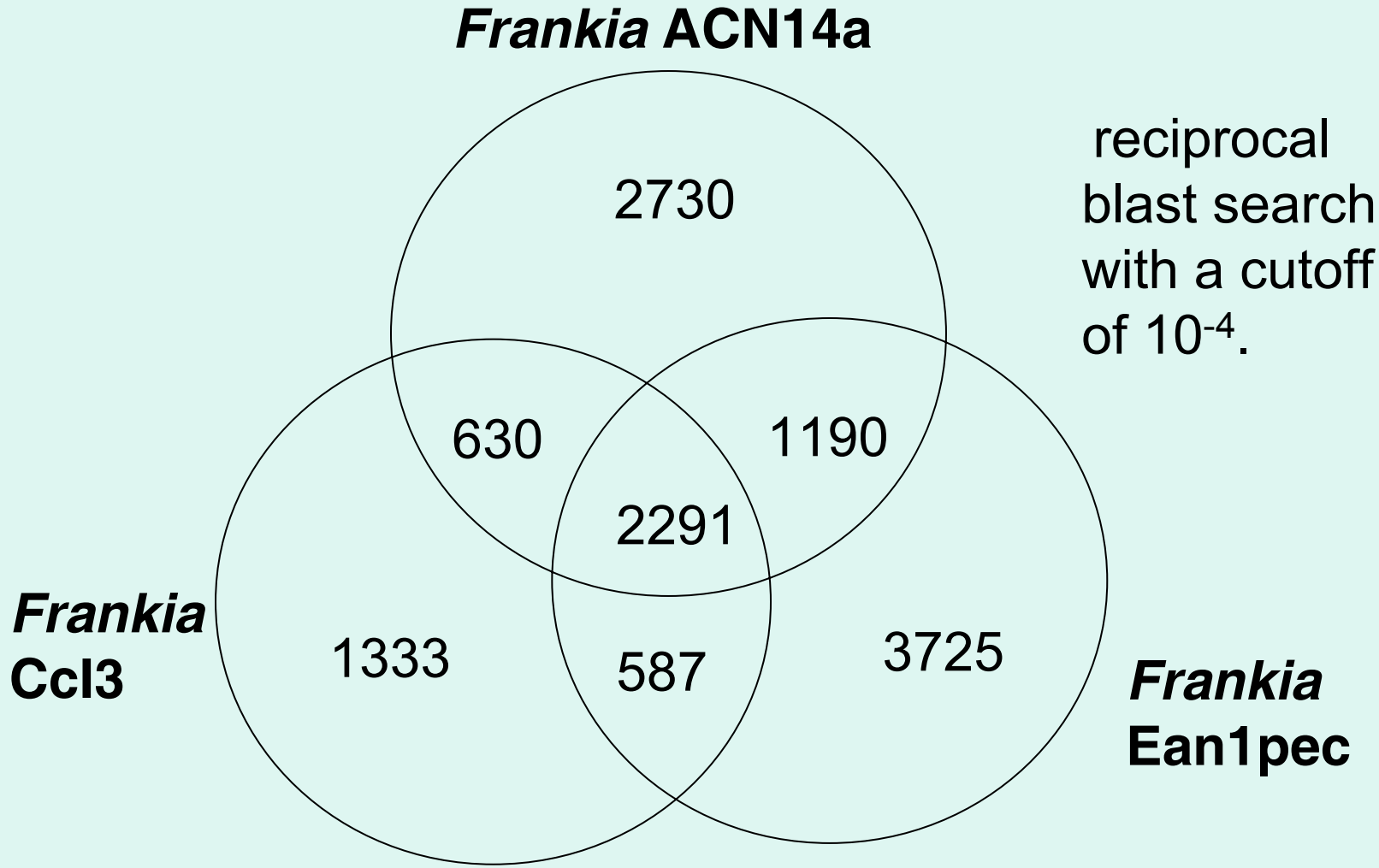
As of 2/6/06

Comparison to other Actinobacteria



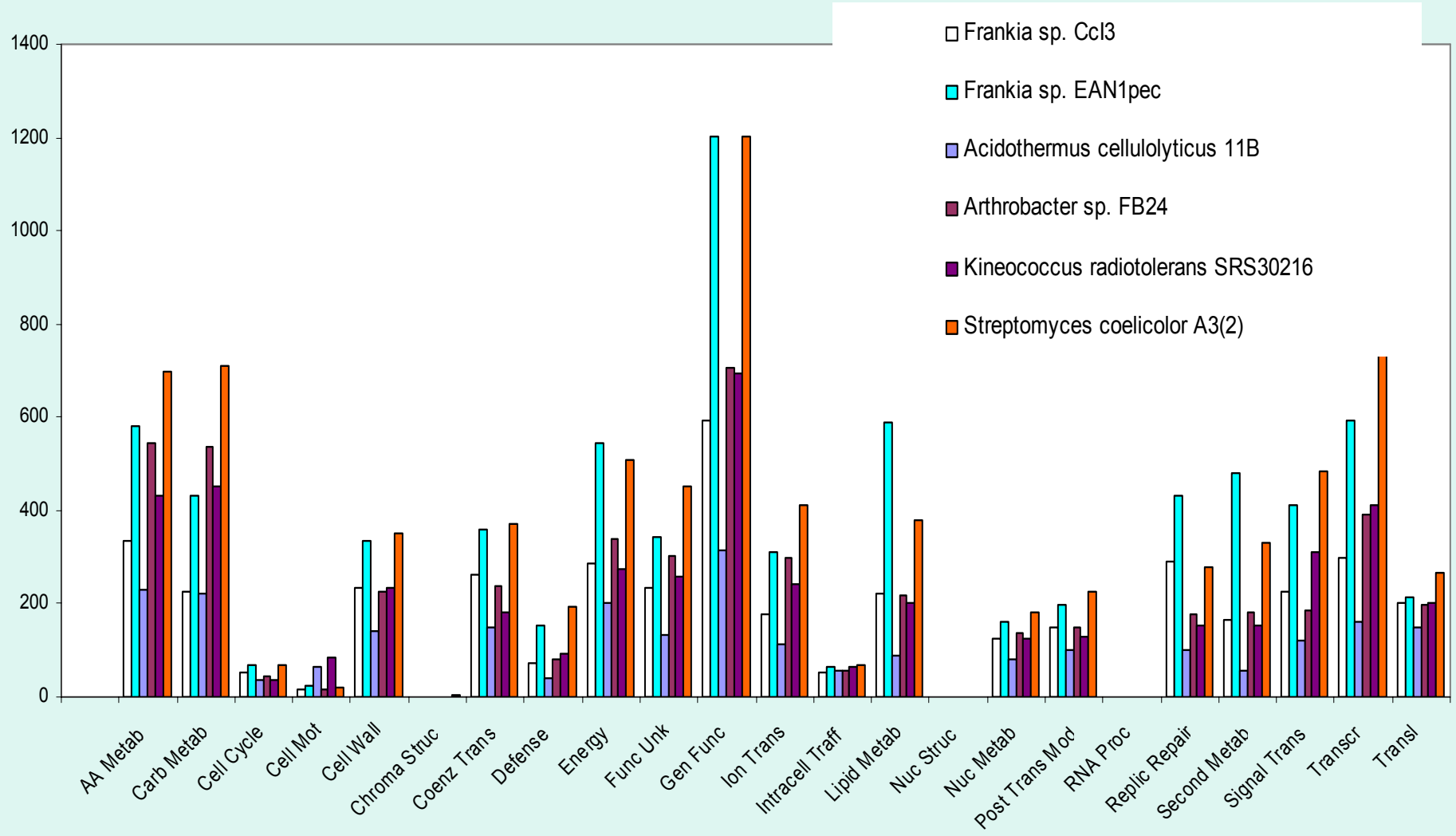

100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Frankia sp. Ccl3
Frankia sp. EAN1pec
Acidothermus cellulolyticus 11B
Arthrobacter sp. FB24
Kineococcus radiotolerans SRS30216
Streptomyces coelicolor A3(2)
w/ Func Pred Perc
w/o Func Pred Sim Perc
w/o Func Pred No Sim Perc
Enzymes Perc
KEGG Perc
Not In KEGG Perc
COG Perc
Pfam Perc
InterPro Perc
GC Perc

Comparison of the CDS

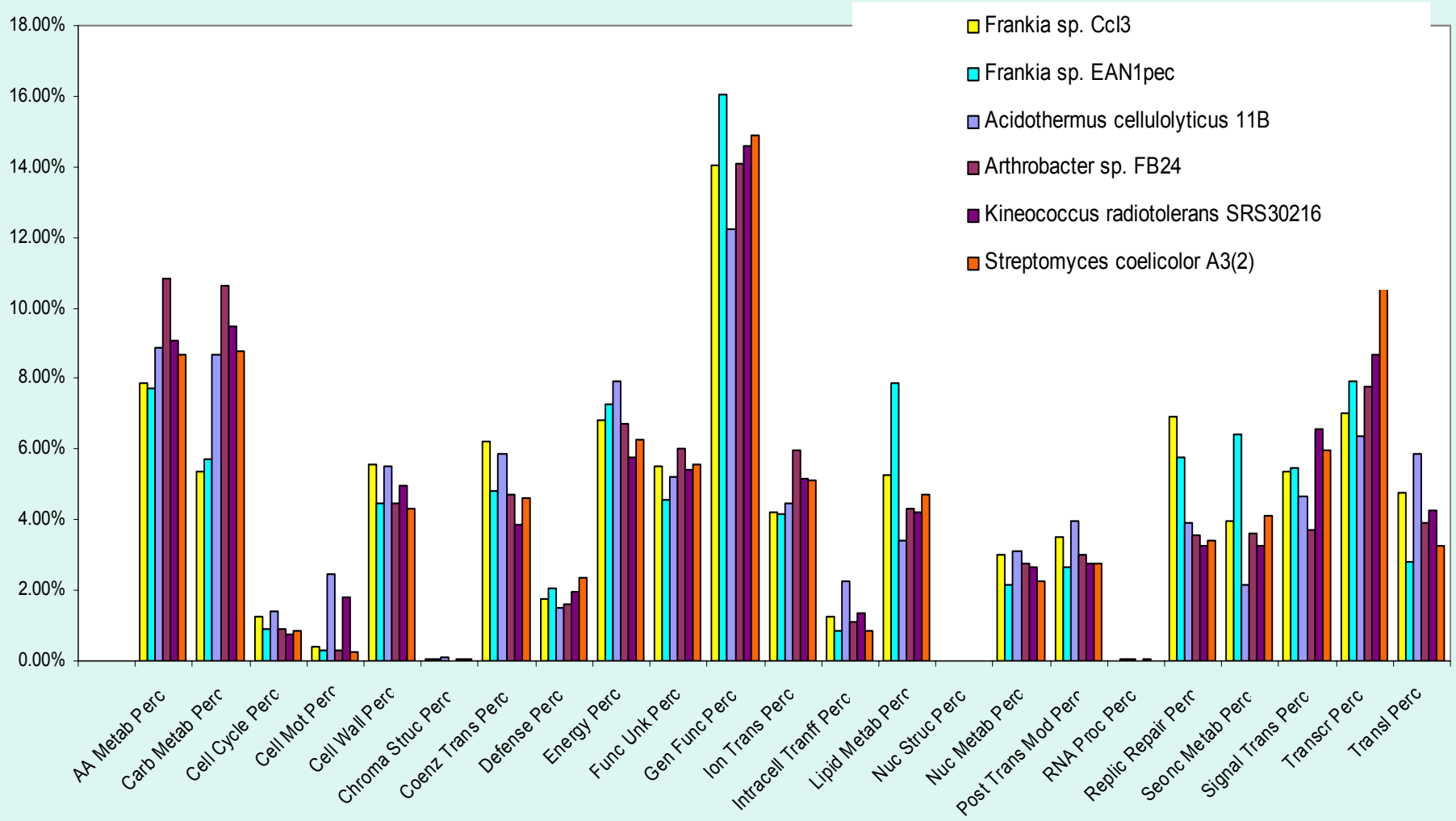


reciprocal blast search with a cutoff of 10^{-4}.

Comparative distribution of ORF function



Comparative distribution on % basis

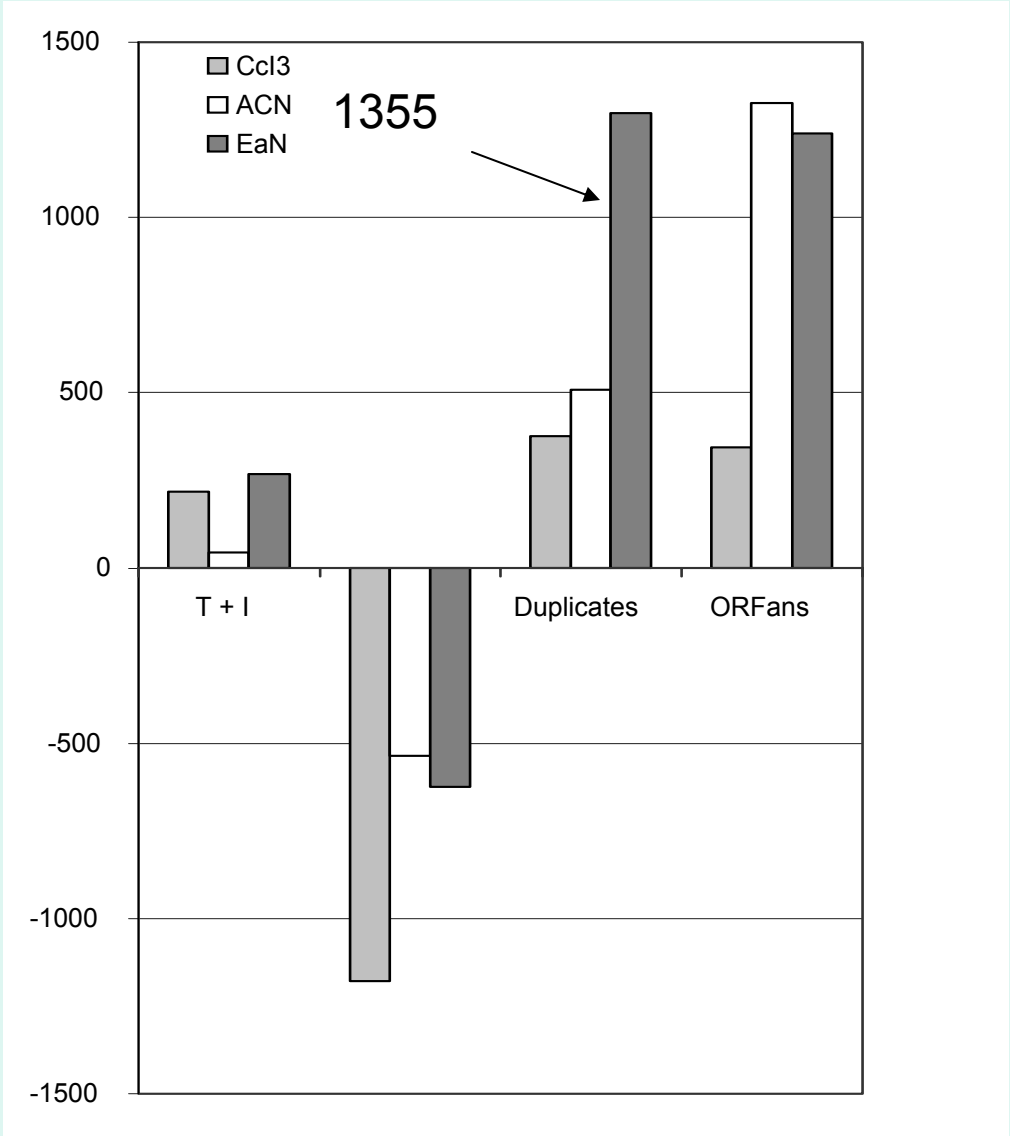


Frankia EAN1pec COG

- Secondary Metabolism (Q)
- Lipid Metabolism (I)
- General Function prediction only (R)

- Energy (C)
- Transcription (K)
- Signal Transduction (T)
- DNA replication, recombination & repair (L)

The EAN Genome is Expanding



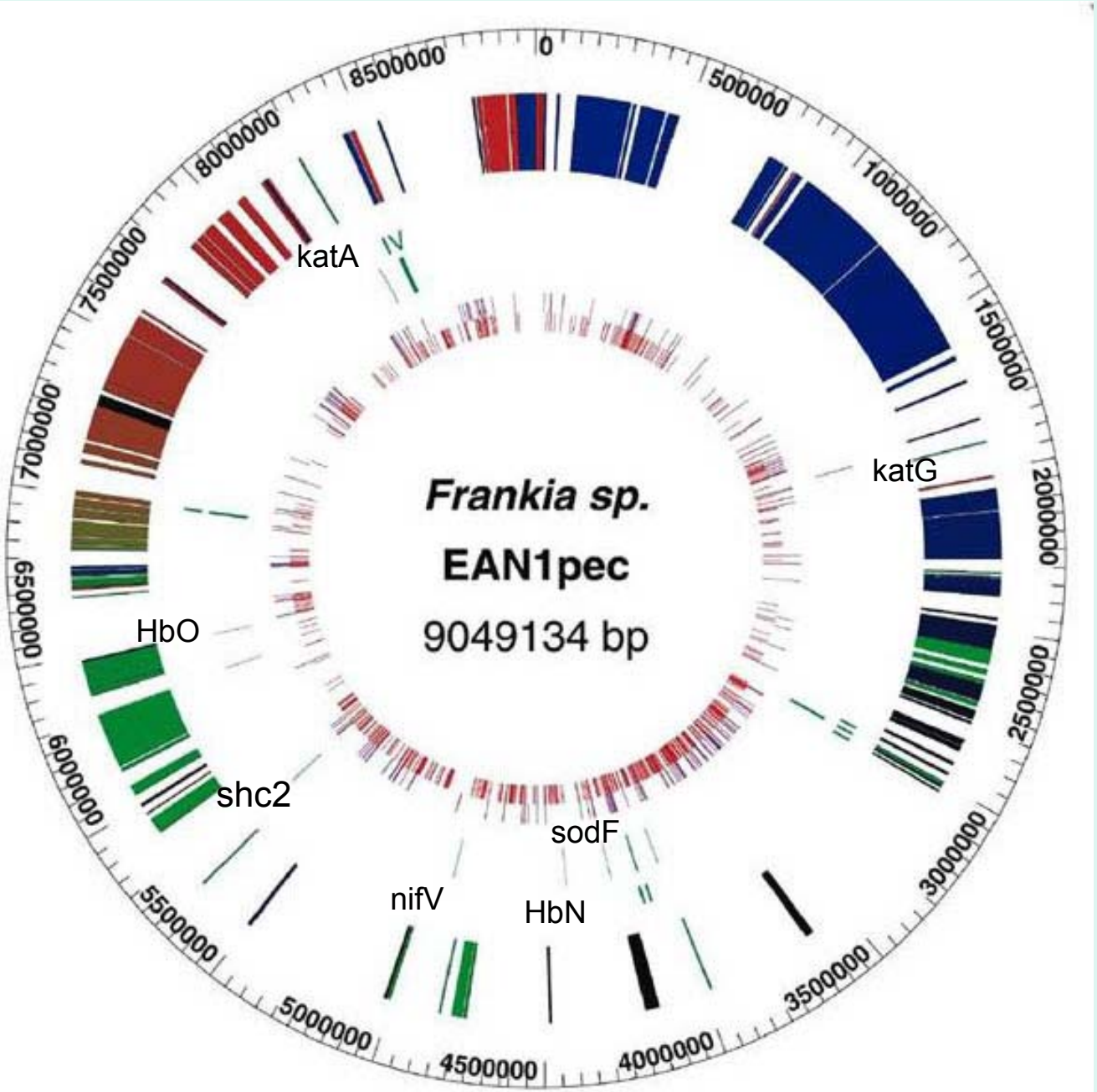
Gene Duplication Level is higher in EAN (18.5% of the ORFs) than ACN (7.5 %) or Ccl3 (9.8 %)

Major families of duplicated genes in *Frankia* EAN

- 49 Short-chain dehydrogenase/reductase SDR
- 48 Integrase, catalytic region
- 32 ABC-type branched-chain amino acid transport systems periplasmic component
- 27 Transposase, IS605 OrfB
- 23 Acyl-CoA dehydrogenase-like
- 22 Cytochrome P450
- 21 Luciferase-like
- 20 Integrase, catalytic region
- 20 Amidohydrolase 2
- 19 Extracellular solute-binding protein, family 5
- 17 Membrane-bound lytic murein transglycosylase B-like
- 14 GGDEF domain
- 13 Putative transposase
- 12 Transposase, IS413ABC transporter related
- 12 Regulatory protein, LuxR
- 11 RNA-directed DNA polymerase
- 11 Phenylpropionate dioxygenase/'related ring-hydroxylating dioxygenases
- 11 Taurine catabolism dioxygenase TauD/TfdA
- 11 Regulatory protein, TetR

Genes identified as potentially involved in Symbiosis

- Hopanoid biosynthesis (Cluster I fppS, sch1, ispH) shc2
- Uptake Hydrogenase biosynthesis (Cluster II; hypA1, hypB1, hupD1, hupS1, hupL1, hypC1, hypD1, hypE1
- Nitrogenase components (Cluster IV; nif)



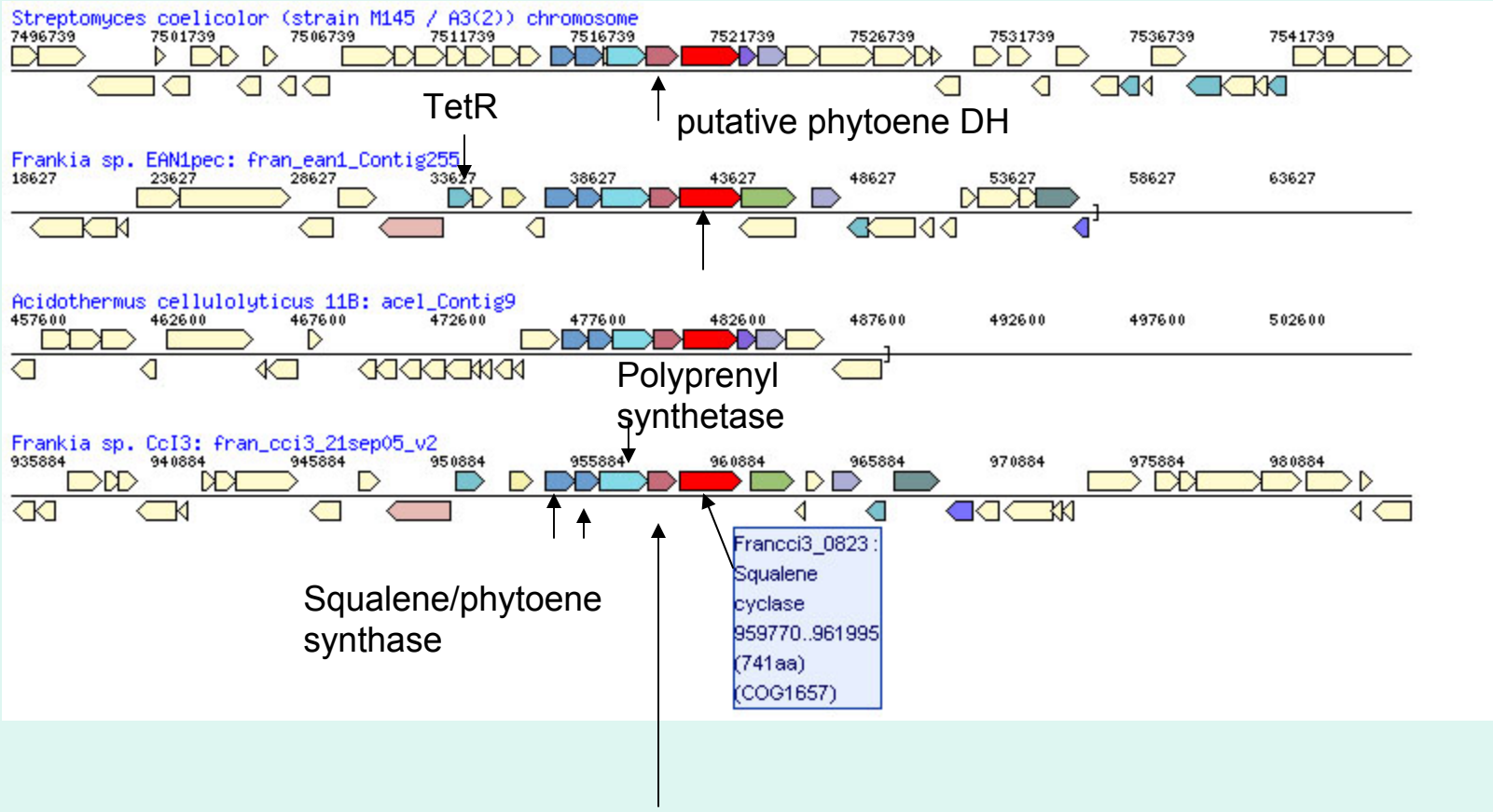
Potential

symbiosis genes are not clustered

- cluster I: shc1 FRAEA6946-6954
- cluster II: hup2 FRAEA4081-4086
- cluster III: hup1, FRAEA2955-2965
- cluster IV: nif, FRAEA8447-8463

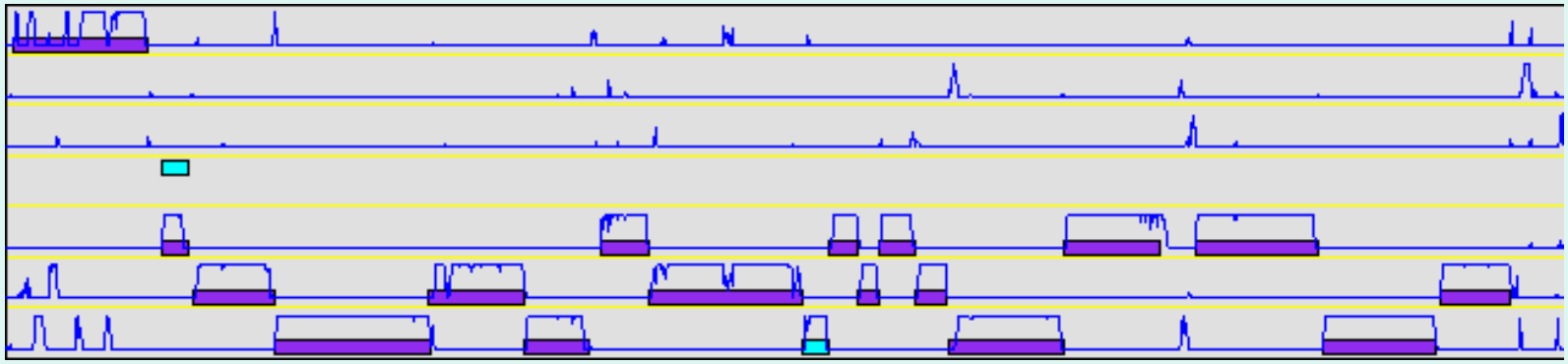
- HbO, FRAEA6420
- HbN, FRAEA4419
- shc2, FRAEA5736
- katA,FRAEA8358
- sodF,FRAEA4204
- nodB FRAEA6279, NifV, FRAEA4890

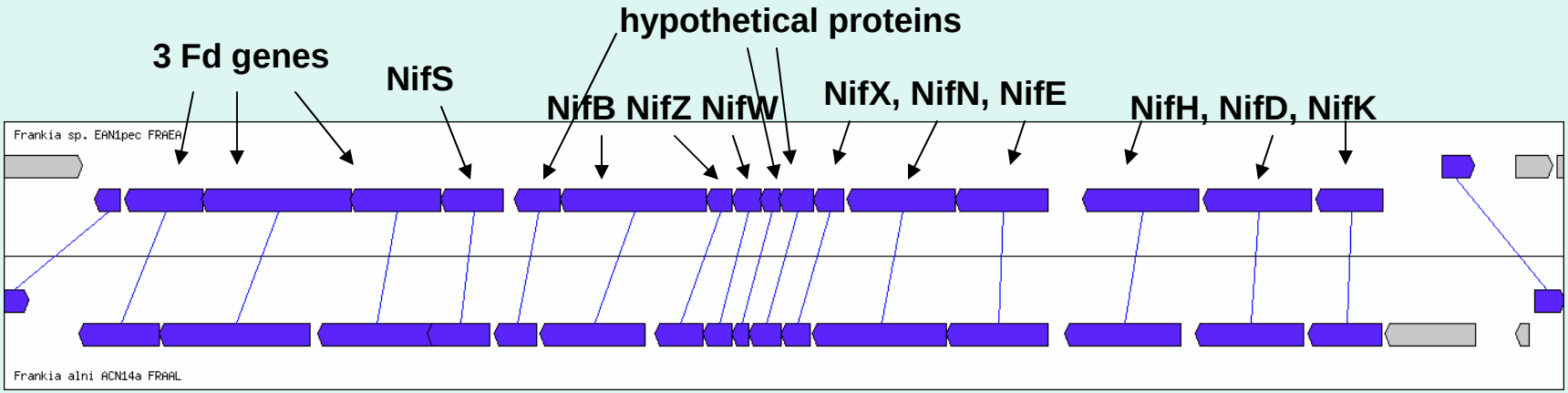
Cluster I: shcI



http://img.jgi.doe.gov/cgi-bin/pub/main.cgi

Nitrogenase Cluster for EAN1pec





NifV is located in another region of the chromosome

MaGe site

Why the large genome (9.1 Mb) for *Frankia* EAN1pec?

Habitat or Life Style?

Habitat?

- many soil dwellers have large genomes (*Streptomyces, Bradyrhizobium, Burkholderia, etc*.
- these “boy-scouts” are always prepared for changing conditions of the soil environment
  - wide array of substrates (uptake systems)
  - need for tight regulation

Metabolism

- Complete Embden-Meyerhof, TCA and Pentose Phosphate pathways
- Incomplete Glycolate Shunt (missing isocitrate lyase)
- wide arsenal of transport genes
- large numbers of genes for short chain dehydrogenase/reductase, dioxygenase, etc.

Transport Genes

- >384 genes ABC Transporters (Major Transport system) also cation ATPase
- Phosphotransfer System PTS (not in other two *Frankia* genomes)
- amino acid permeases (3)
- arabinose permeases (55)
- sugar phosphate permease (7)
- Arsenite efflux pump ACR3 and related permeases
- antiporter and symporter genes
- MSC channel (L and S)

PTS Genes

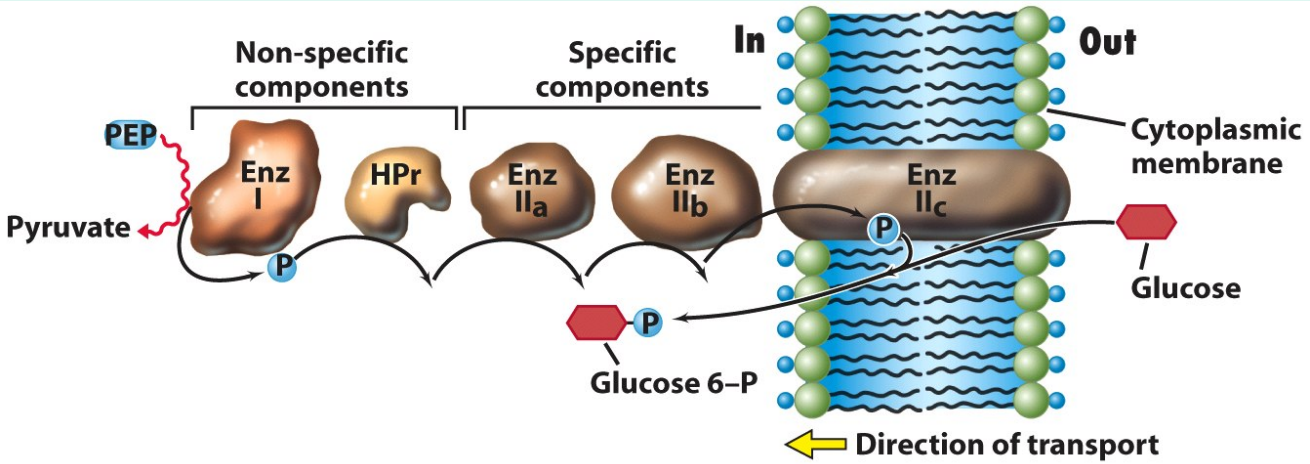


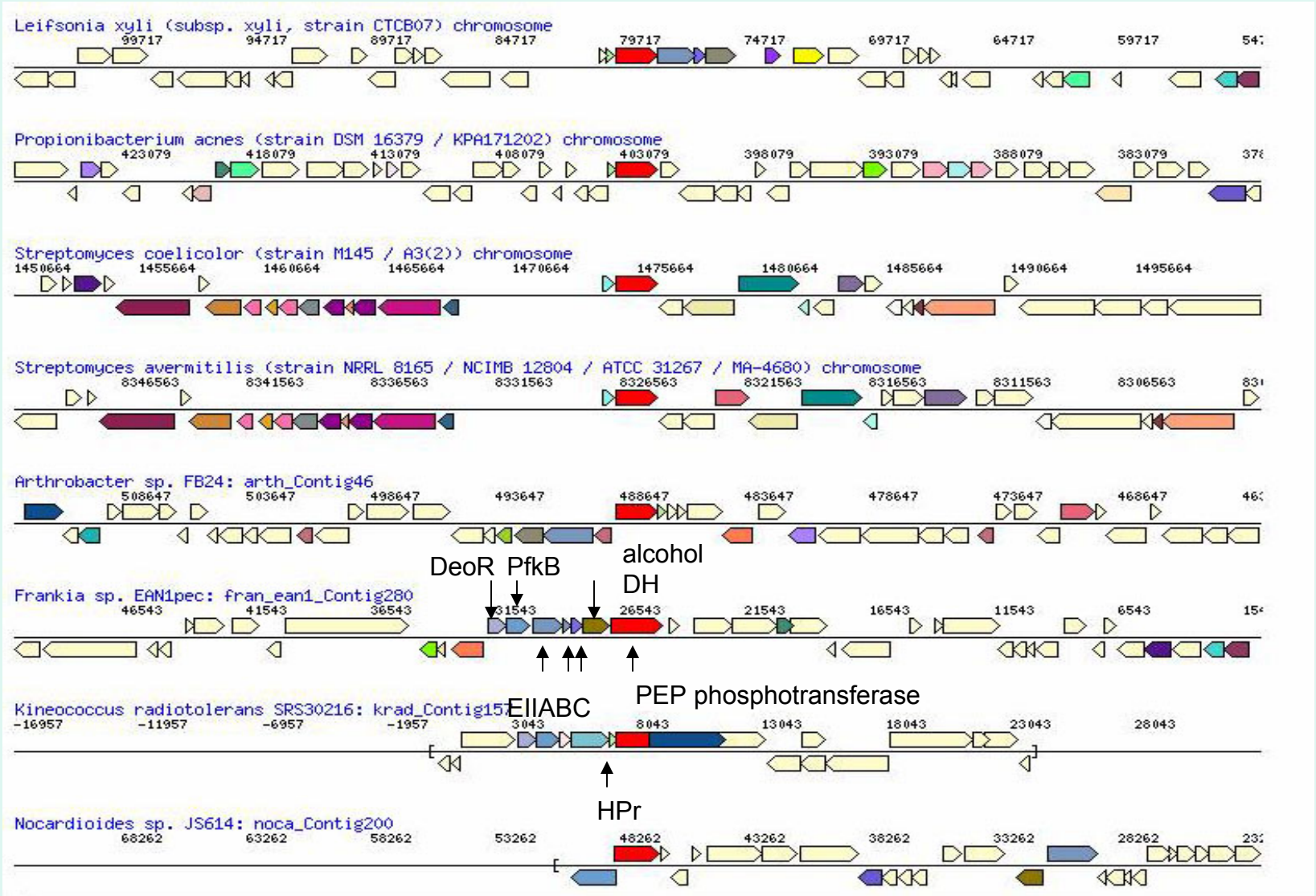
Figure 4-25 Brock Biology of Microorganisms 11/e
© 2006 Pearson Prentice Hall, Inc.

- compare to Streptomyces
- DeoR, carbohydrate kinase(PfkB) IIABC, Zn alcohol DH, PEP phosphotransferase (EI): potential operon?
- fructose, mannitol and sorbitol metabolism is inducible.

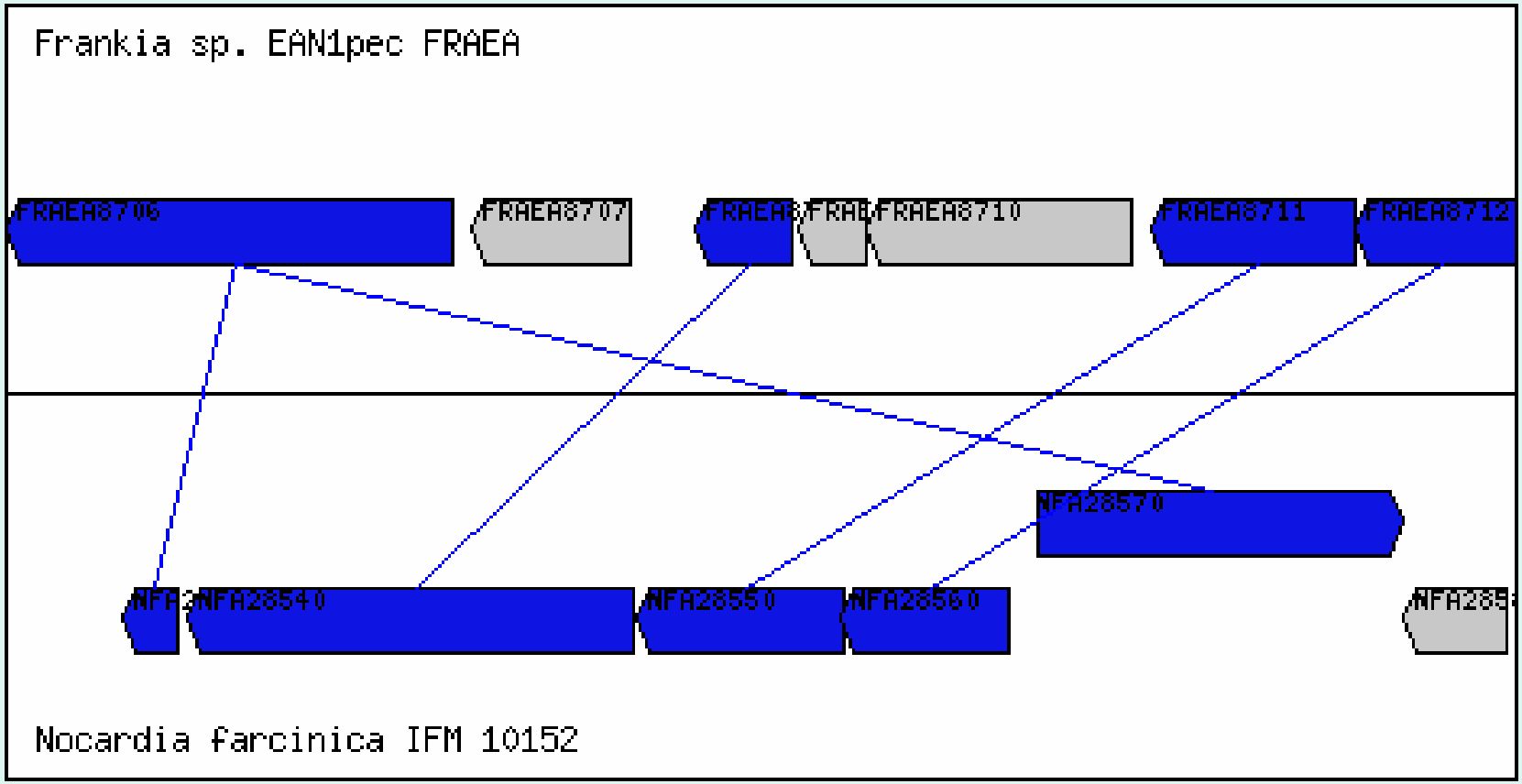
Franean13385
PEP-protein phosphotransferase

- Phosphoenolpyruvate-protein phosphotransferase [Thermobifida fusca YX] Expect = e-143 Identities = 303/568 (53%), Positives = 355/568 (62%), Gaps = 30/568 (5%)
- PTS system, enzyme I [Leifsonia xyli subsp. xyli str. CTCB07] Length = 548 Score = 476 bits (1225), Expect = e-132 Identities = 283/538 (52%), Positives = 336/538 (62%), Gaps = 20/538 (3%)
- Phosphoenolpyruvate-protein phosphotransferase [Kineococcus radiotolerans SRS30216] Length = 557 Score = 468 bits (1203), Expect = e-130 Identities = 287/564 (50%), Positives = 339/564 (60%), Gaps = 28/564 (4%)
- phosphoenolpyruvate-protein phosphotransferase [Streptomyces coelicolor A3(2)] Length = 556 Score = 467 bits (1202), Expect = e-130 Identities = 288/563 (51%), Positives = 350/563 (62%), Gaps = 24/563 (4%)
- putative phosphoenolpyruvate-protein phosphotransferase [Streptomyces avermitilis MA-4680] Length = 556 Score = 459 bits (1182), Expect = e-127 Identities = 280/567 (49%), Positives = 347/567 (61%), Gaps = 32/567 (5%)

PTS genes in other Actinobacteria



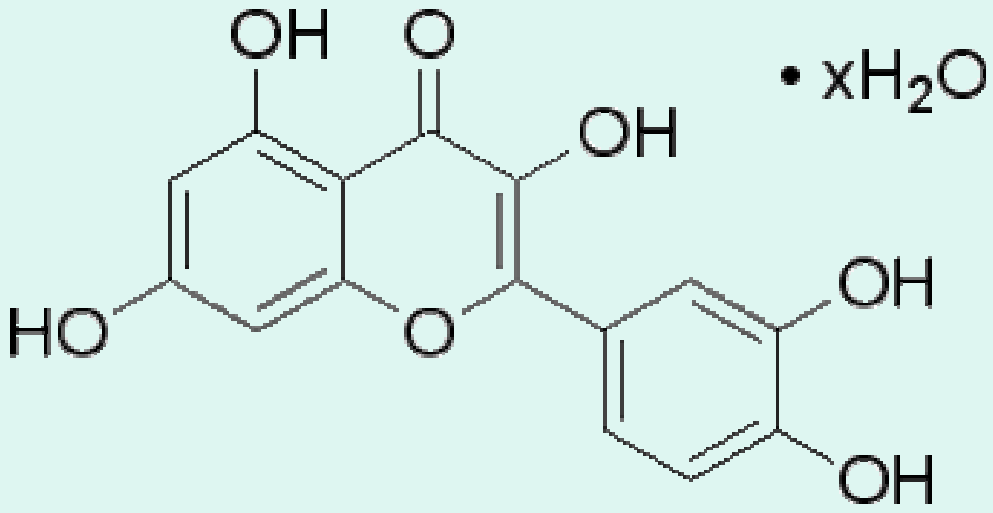
Syntony with other actinobacteria PTS genes



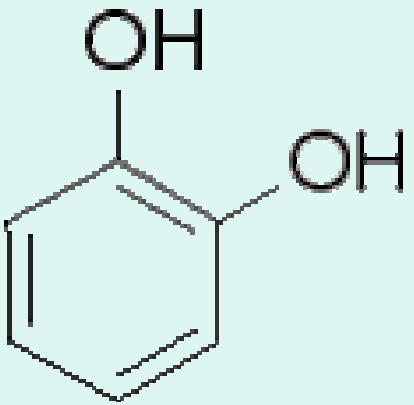
General Metabolic Enzyme Groups

	ACN	CCl3	EAN
esterase/lipase	75	19	50
cellulase	2	0	2
aldehyde DH	30	13	41
alcohol DH	18	9	35
dioxygenase	26	3	40
short chain DH/reductase	38	7	101

Is *Frankia* EAN1pec versatile?

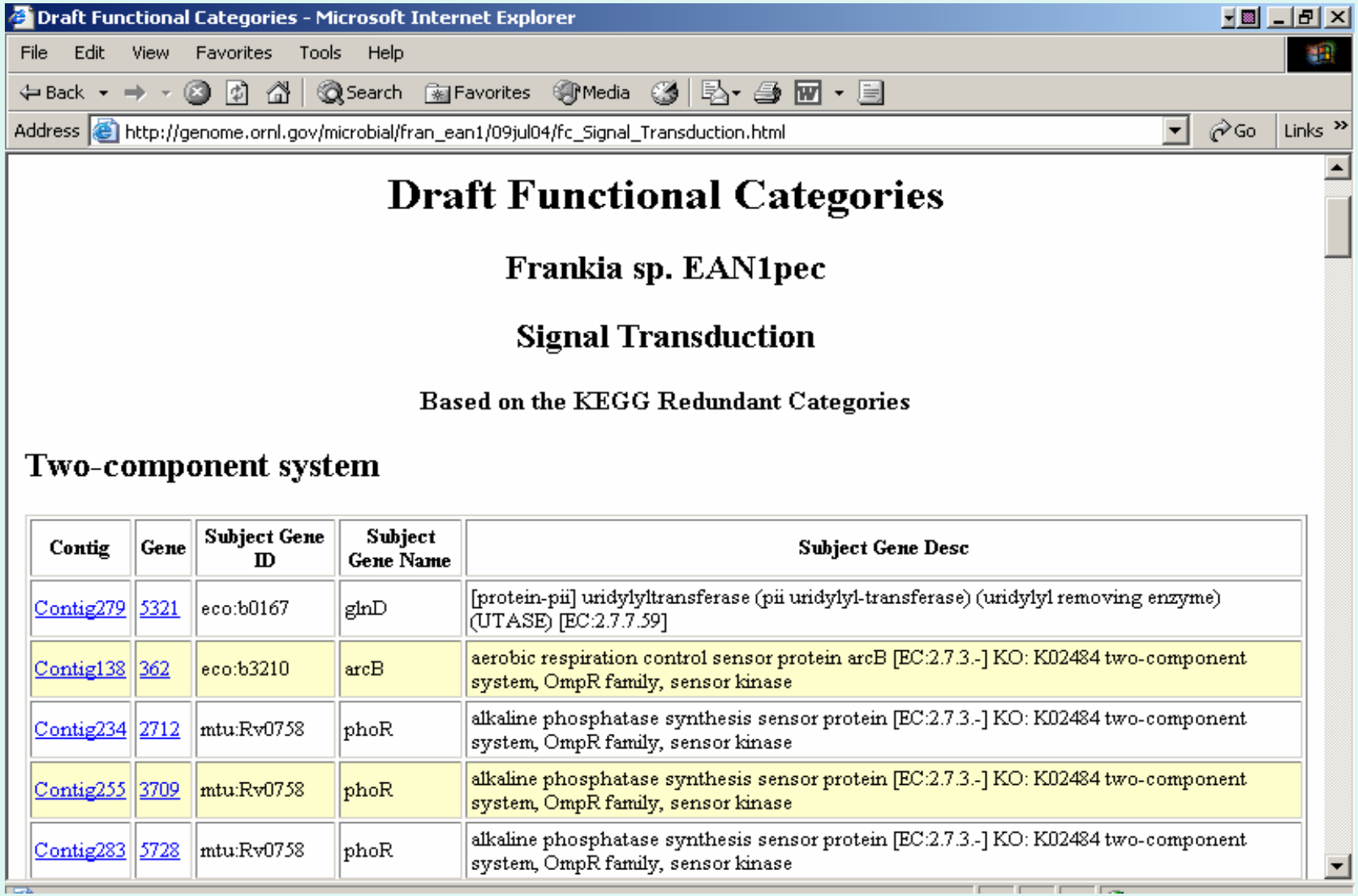


Quercetin



Catechol

Signal Transduction (Two Component)



Draft Functional Categories

Frankia sp. EAN1pec

Signal Transduction

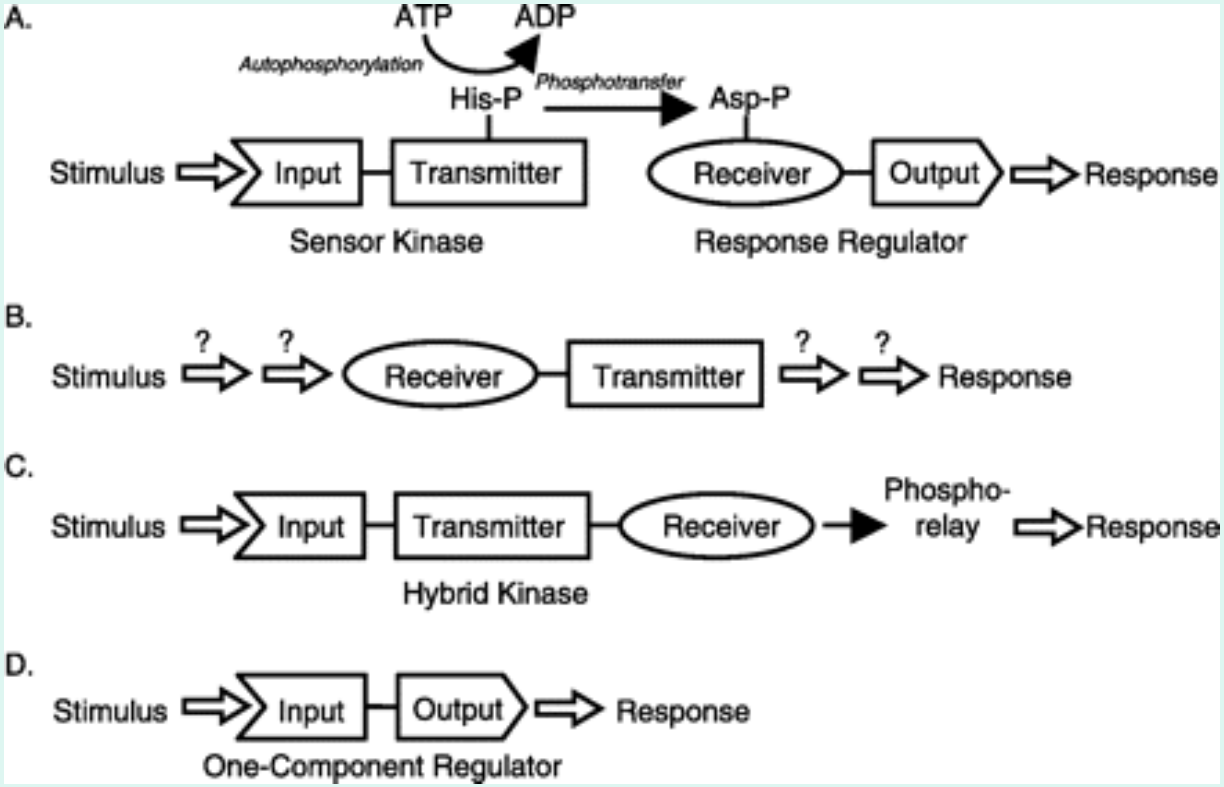
Based on the KEGG Redundant Categories

Two-component system

Contig	Gene	Subject Gene ID	Subject Gene Name	Subject Gene Desc
Contig279	5321	eco:b0167	glnD	[protein-pii] uridylyltransferase (pii uridylyl-transferase) (uridylyl removing enzyme) (UTASE) [EC:2.7.7.59]
Contig138	362	eco:b3210	arcB	aerobic respiration control sensor protein arcB [EC:2.7.3.-] KO: K02484 two-component system, OmpR family, sensor kinase
Contig234	2712	mtu:Rv0758	phoR	alkaline phosphatase synthesis sensor protein [EC:2.7.3.-] KO: K02484 two-component system, OmpR family, sensor kinase
Contig255	3709	mtu:Rv0758	phoR	alkaline phosphatase synthesis sensor protein [EC:2.7.3.-] KO: K02484 two-component system, OmpR family, sensor kinase
Contig283	5728	mtu:Rv0758	phoR	alkaline phosphatase synthesis sensor protein [EC:2.7.3.-] KO: K02484 two-component system, OmpR family, sensor kinase

Two-Component Systems

- Large numbers of these genes present in EAN

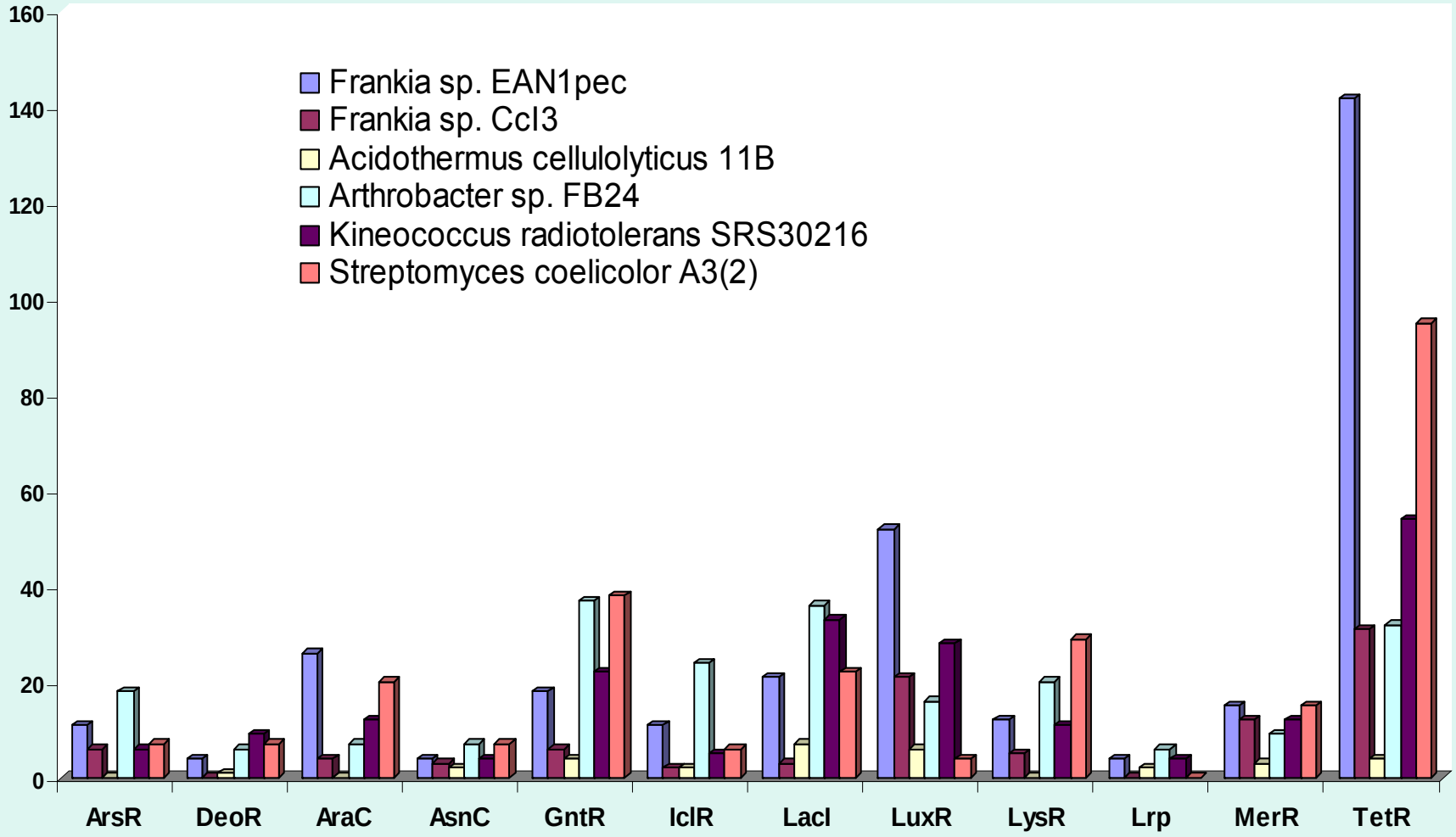


Other Major Regulatory mechanisms

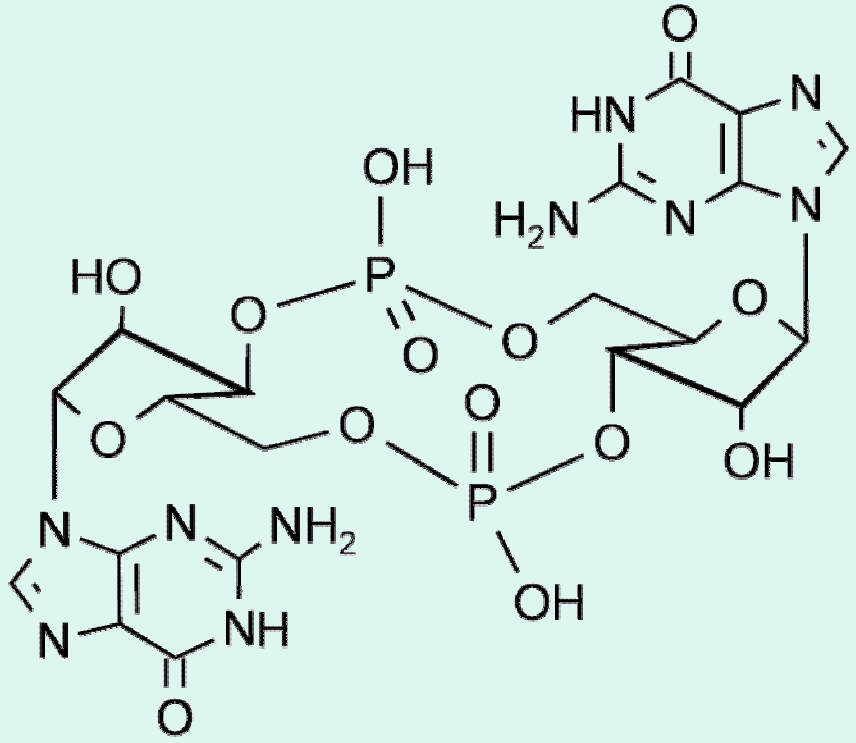
- Many DNA binding proteins

- Sigma Factors
- Anti-sigma Factors
- Anti-sigma Factor Antagonists

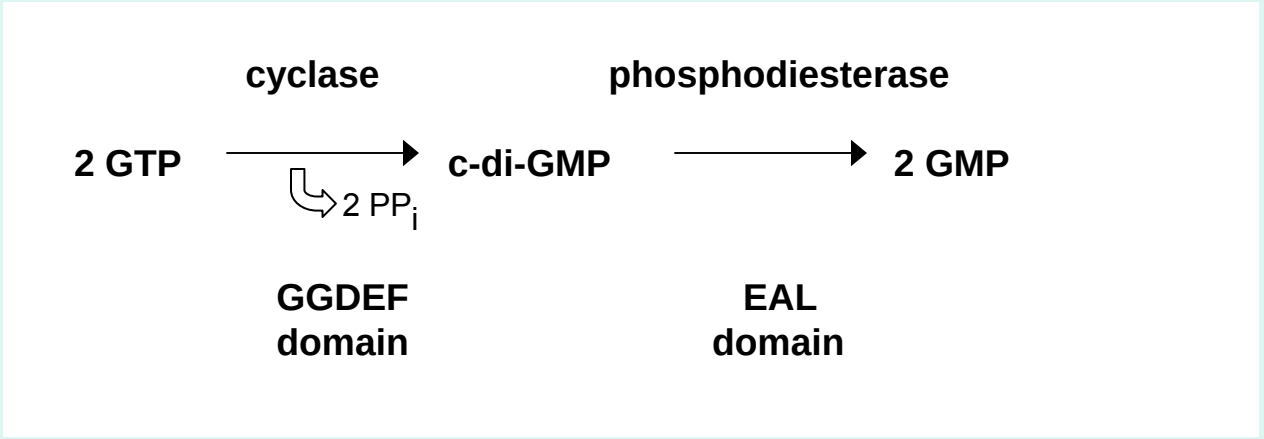
DNA Regulatory Proteins



Di-cyclic guanosinemonophosphate (c-di-GMP)

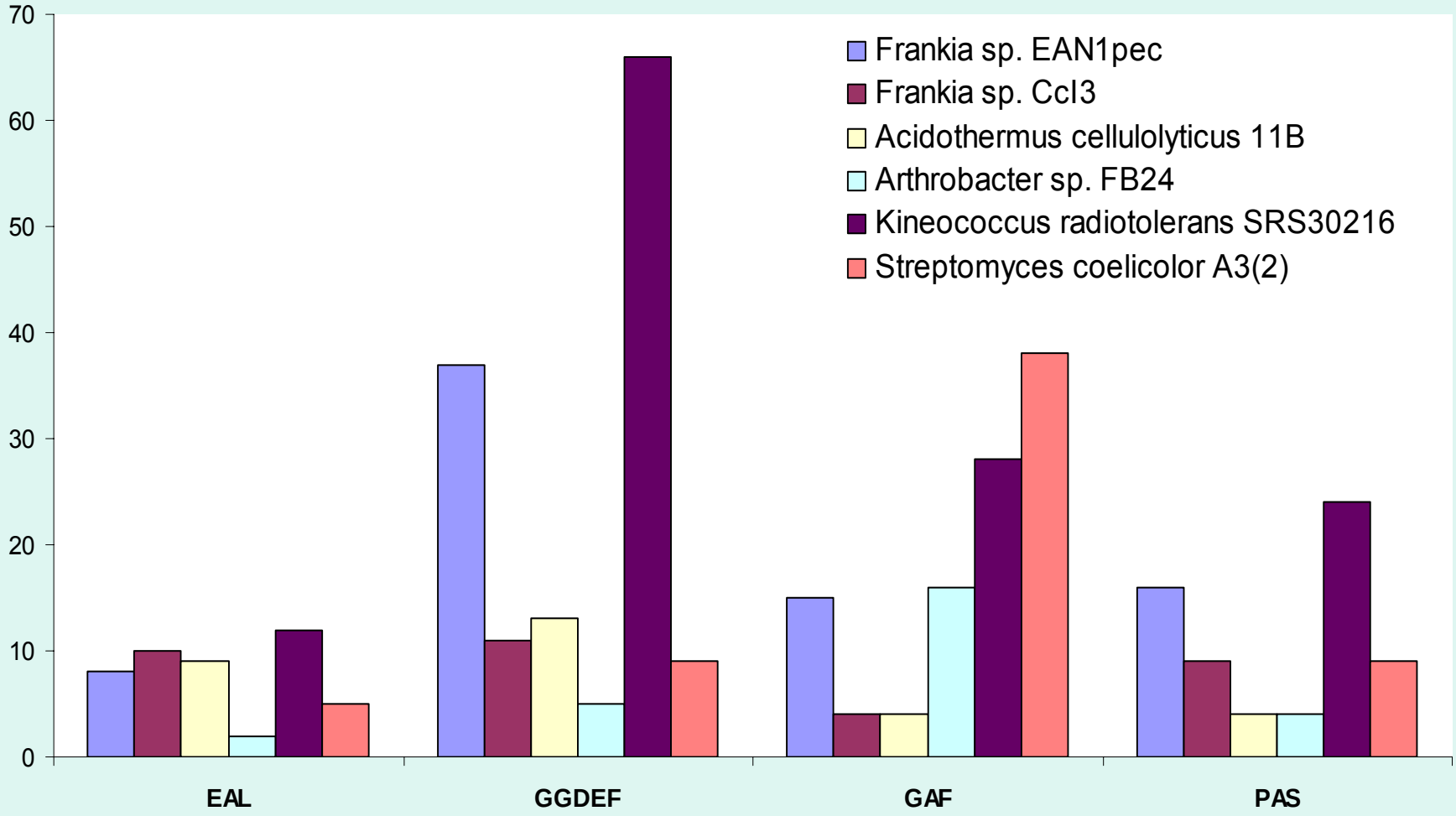


- Bacterial secondary messenger
- functions in:
  - changing surface properties
  - development
  - motility
  - conserved regulator of bacterial adhesion to plant and animal surfaces



- **GGDEF** synthetase cyclic di-GMP
- **EAL** cyclic di-GMP hydrolase
- **PAS** oxygen sensor domain
- **GAF** Domain present in phytochromes and cGMP-specific phosphodiesterases

Domain distribution

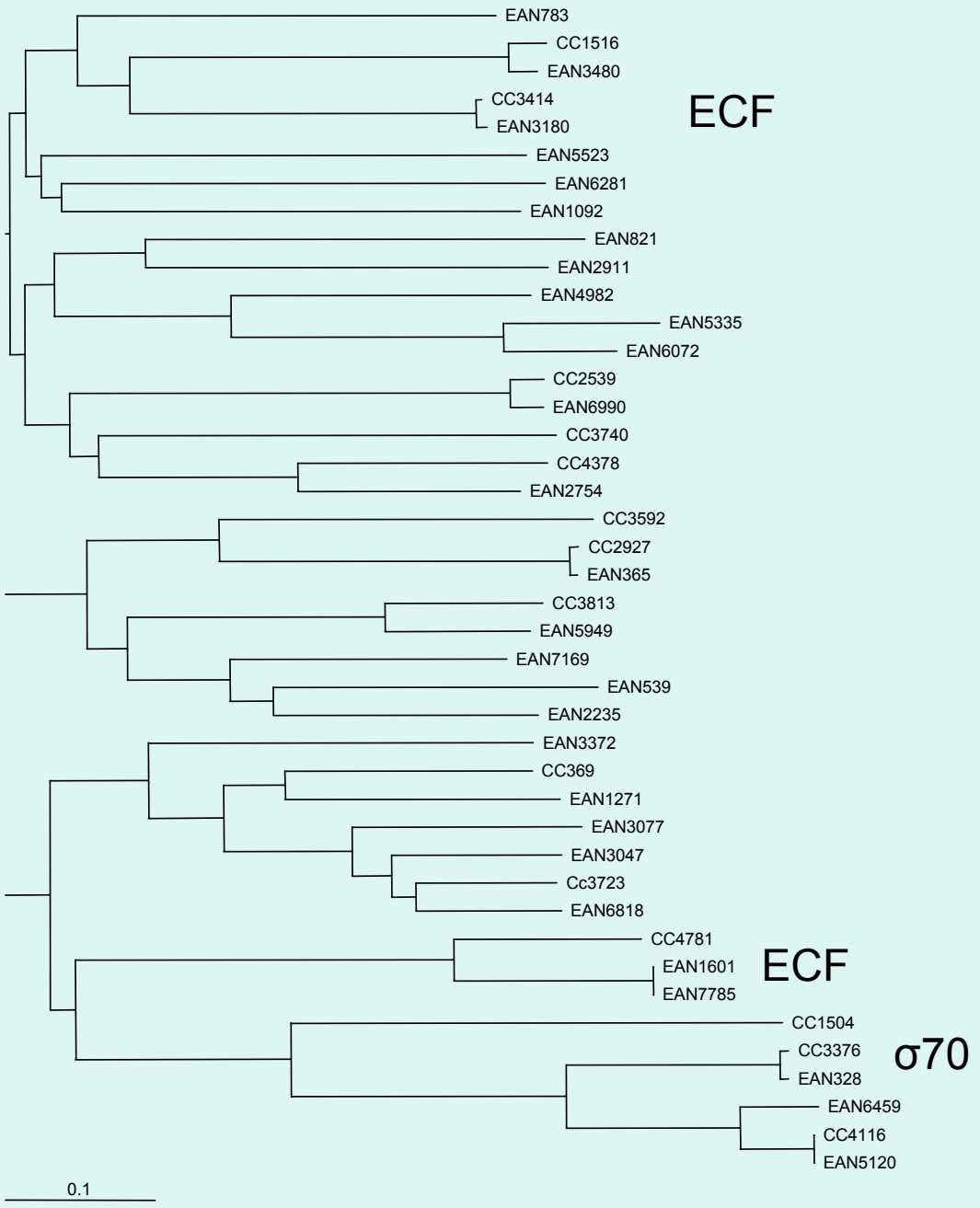


Sigma Factors

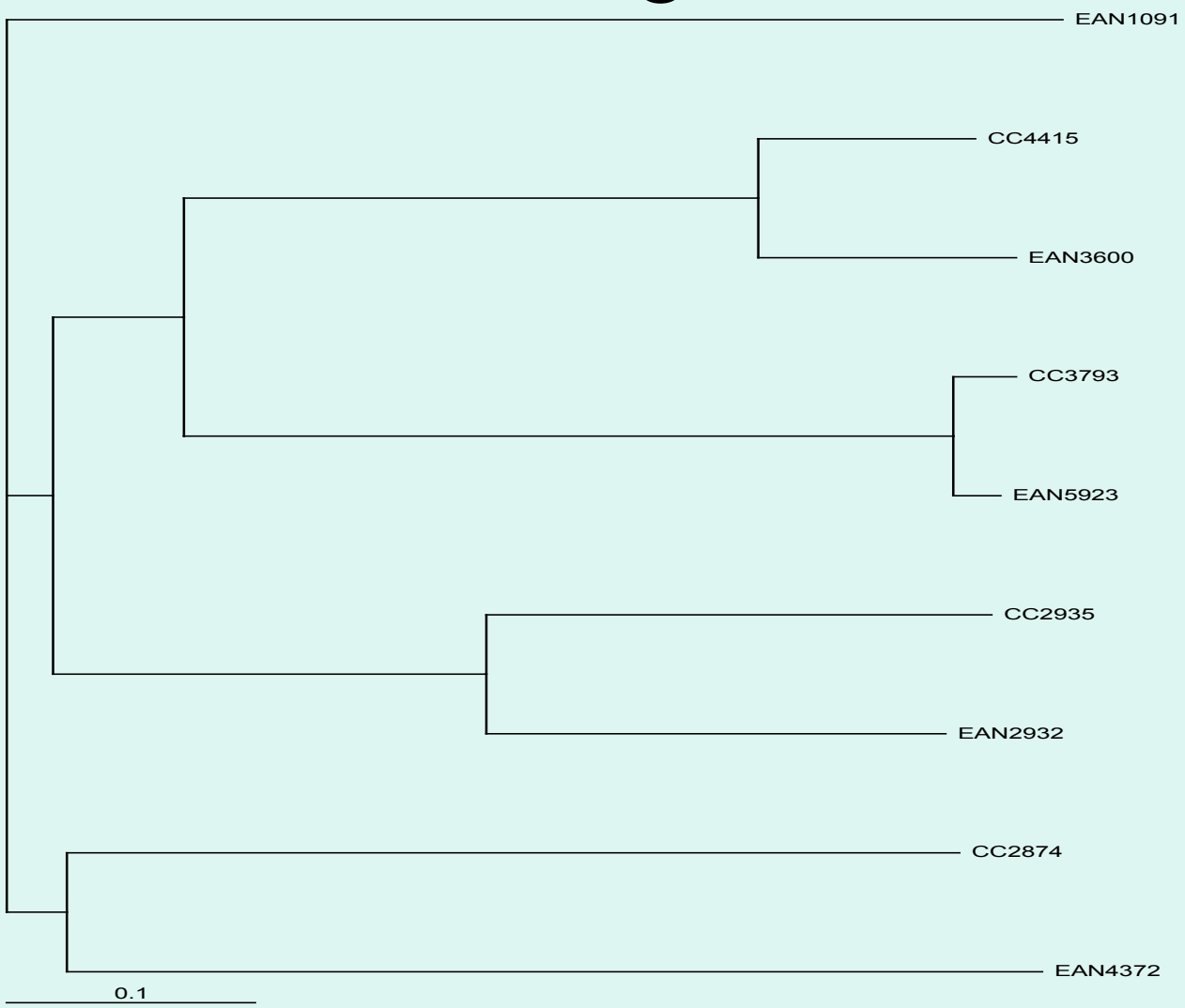
14 Ccl3

28

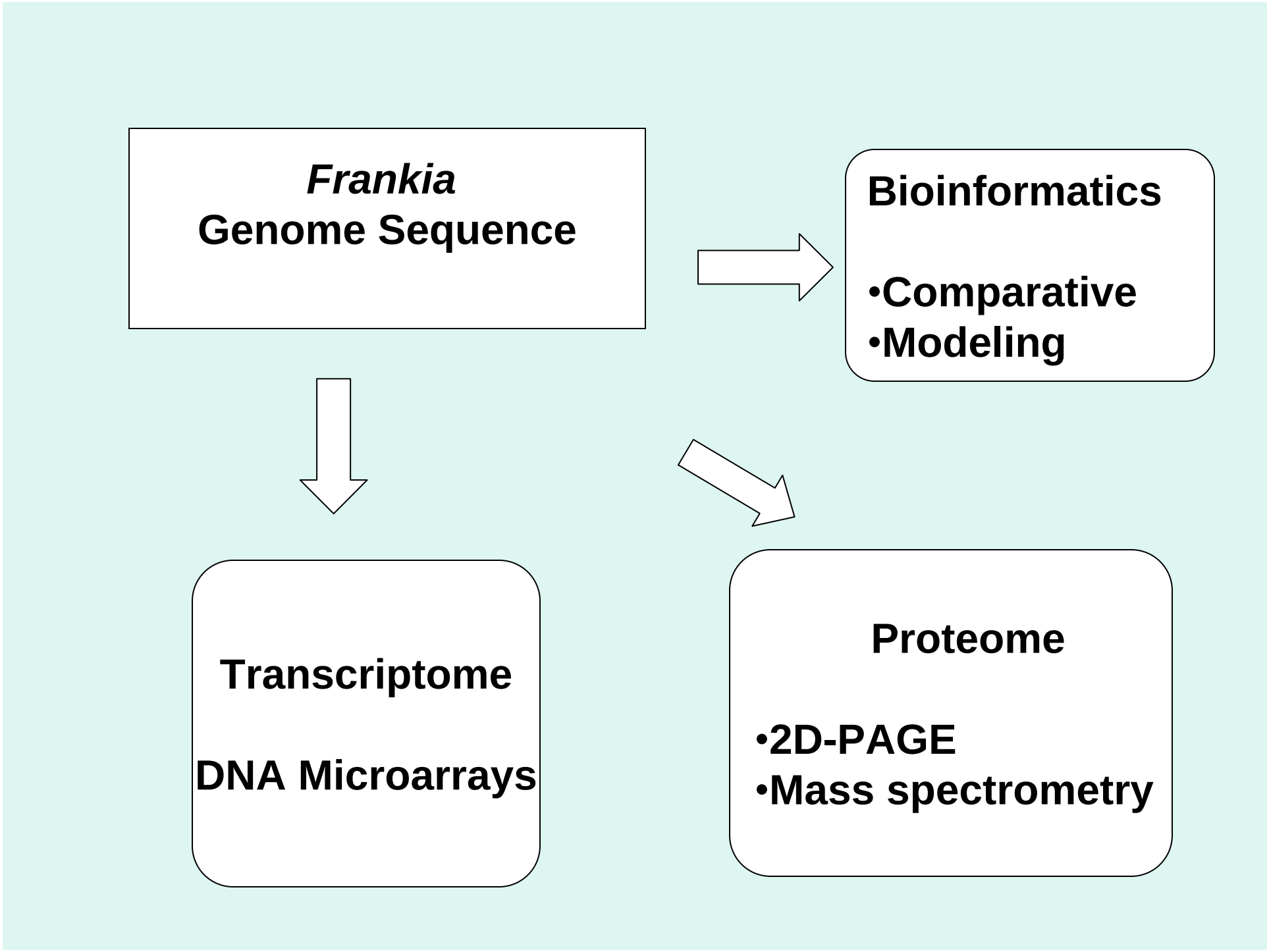
EAN1pec



Anti-Anti-sigma factors



Common to all three


Frankia
Genome Sequence
Bioinformatics
•Comparative
•Modeling
Transcriptome
DNA Microarrays
Proteome
•2D-PAGE
•Mass spectrometry

Vesicle Development controlled by:

a. N status
b. Oxygen
c. Mo & Fe
d. Calcium
e. Temperature

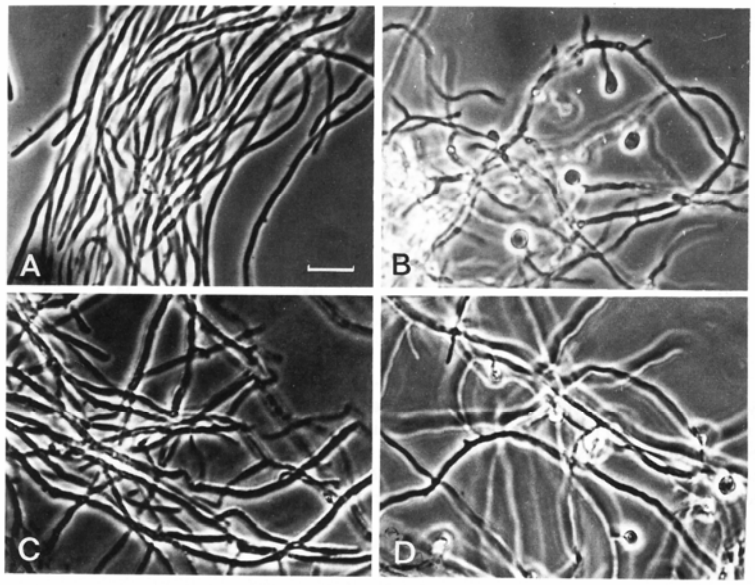


Fig. 2 A – D
Phase-contrast photomicrograph of *Frankia* strain CPI incubated for 4 d at 25°C. **A** Medium with NH_4^+ as nitrogen source; **B** Medium with N_2 as nitrogen source; **C** Medium with N_2 as nitrogen source and 1.0 mM EGTA; **D** Medium with N_2 as nitrogen source, 1.0 mM EGTA and 5.0 mM $CaCl_2$; Bar represents 10 µm

What is the protein profile under these conditions?

Vesicle-specific proteins?

Vesicle Development and Function

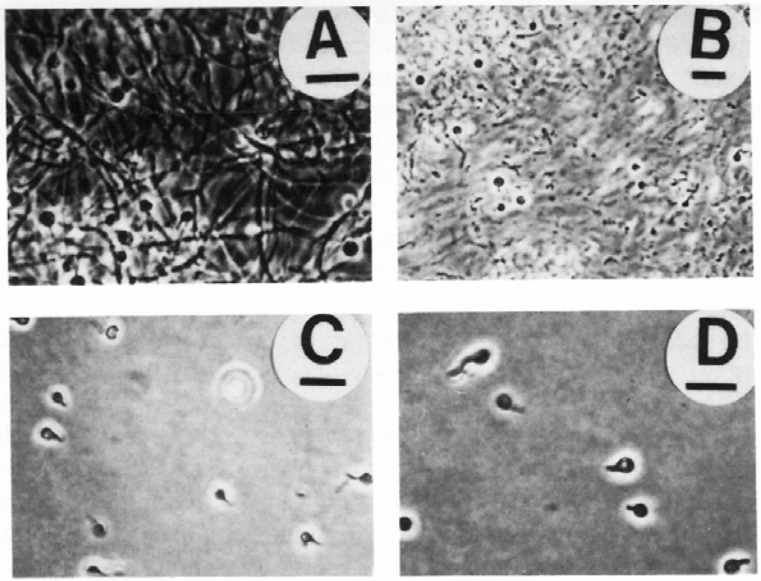


FIG. 1. Photomicrographs of *Frankia* sp. strain EAN at different stages of vesicle purification. (A) Whole cells grown with N_2 as the nitrogen source. (B) Culture after one passage through a French pressure cell at 10,000 lb/in^2. (C) Vesicles purified by a series of low-speed centrifugations. (D) Vesicles purified by isopycnic density gradient centrifugation.

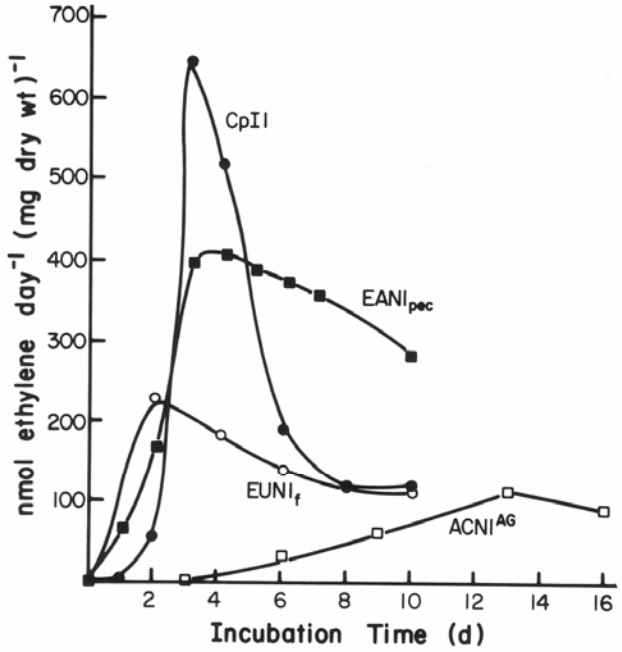
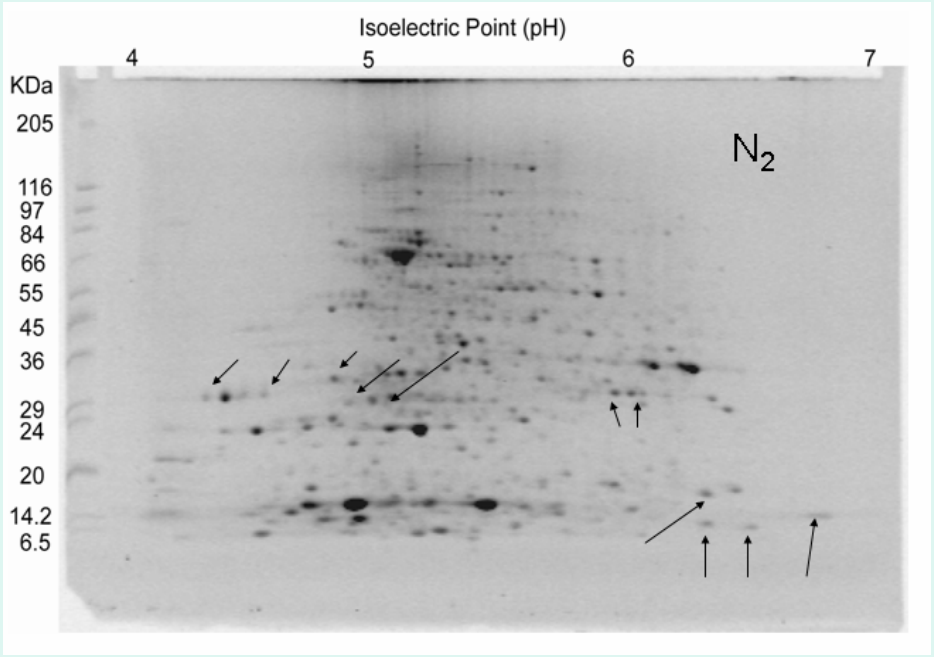
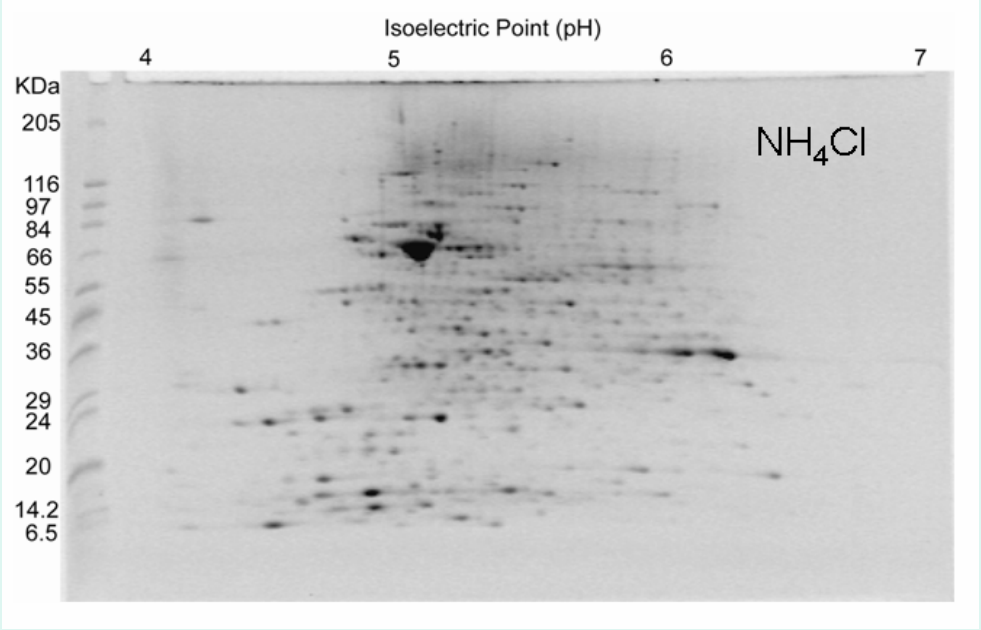


Fig. 1. Depression of nitrogenase by four *Frankia* strains following transfer to media with N_2 as nitrogen source. Cells were grown at their optimum growth temperatures on medium containing NH_4^+ as nitrogen source and succinate as carbon source except for EUN where fructose was used. Washed cells were inoculated into NH_4^+-free medium with the same carbon sources. Strains CPI, EAN and EUN were incubated at 25°C and ACN at 20°C. At varied time intervals, samples were removed and assayed for nitrogenase activity by the acetylene reduction procedure. Activity is expressed as nmol ethylene produced/day (mg dry wt)$^{-1}$

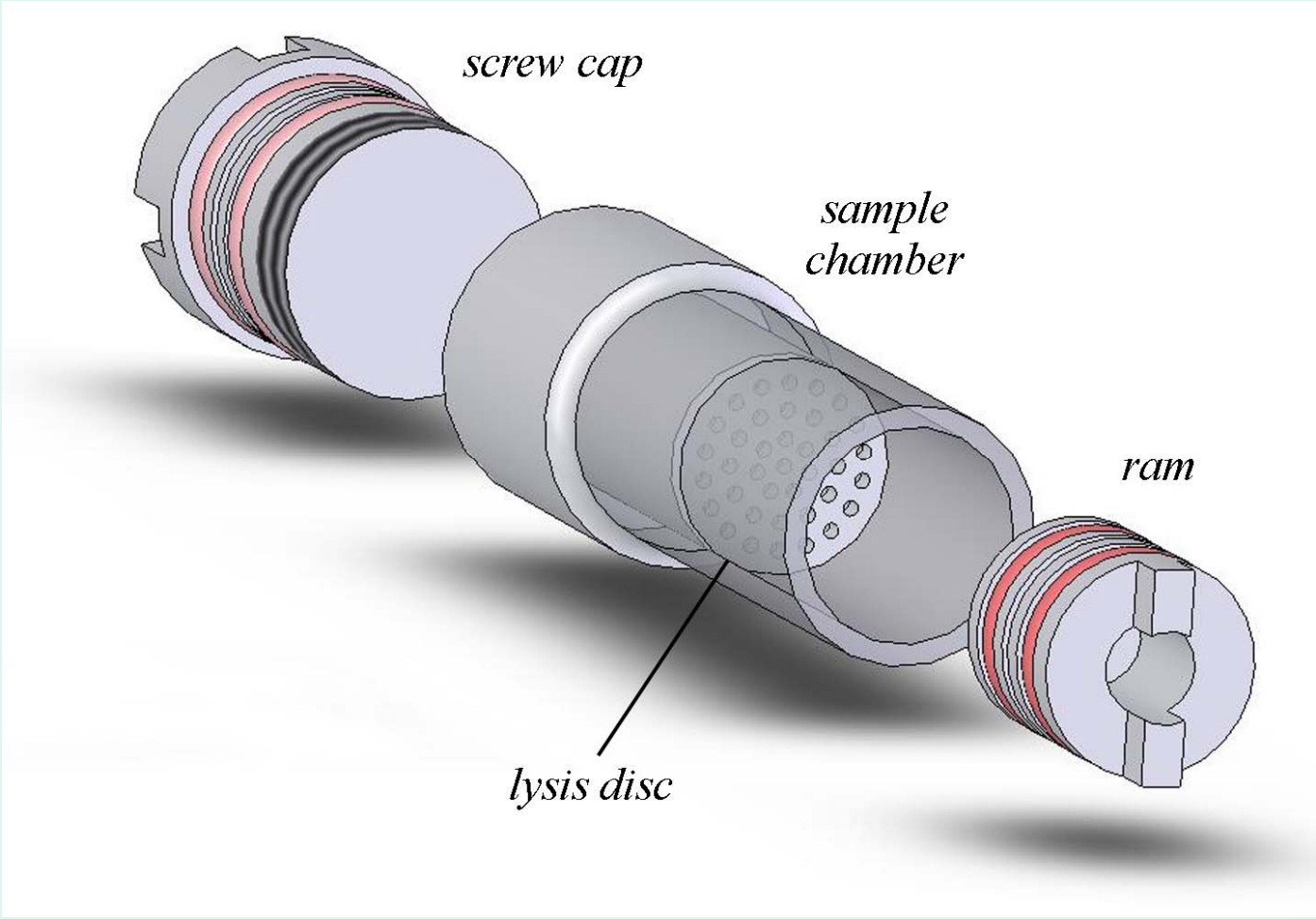
Purified Vesicles





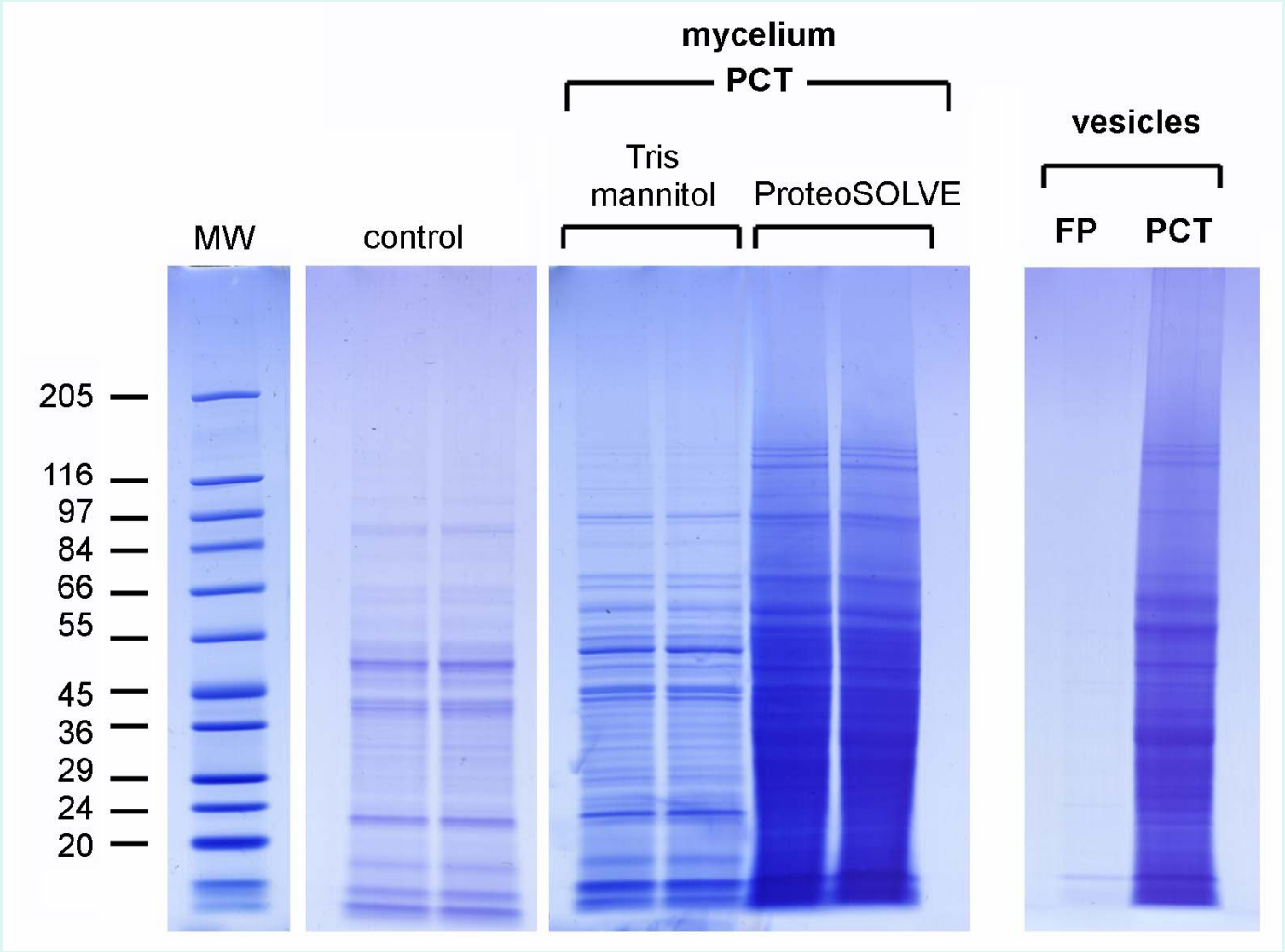
Proteome profile of *Frankia* Ccl3 grown under N_2 or NH_4Cl conditions. Arrow point out N_2-grown specific proteins.

The PULSE Tube used in Pressure Cycling Technology facilitates high efficiency lysis of cells and subcellular components



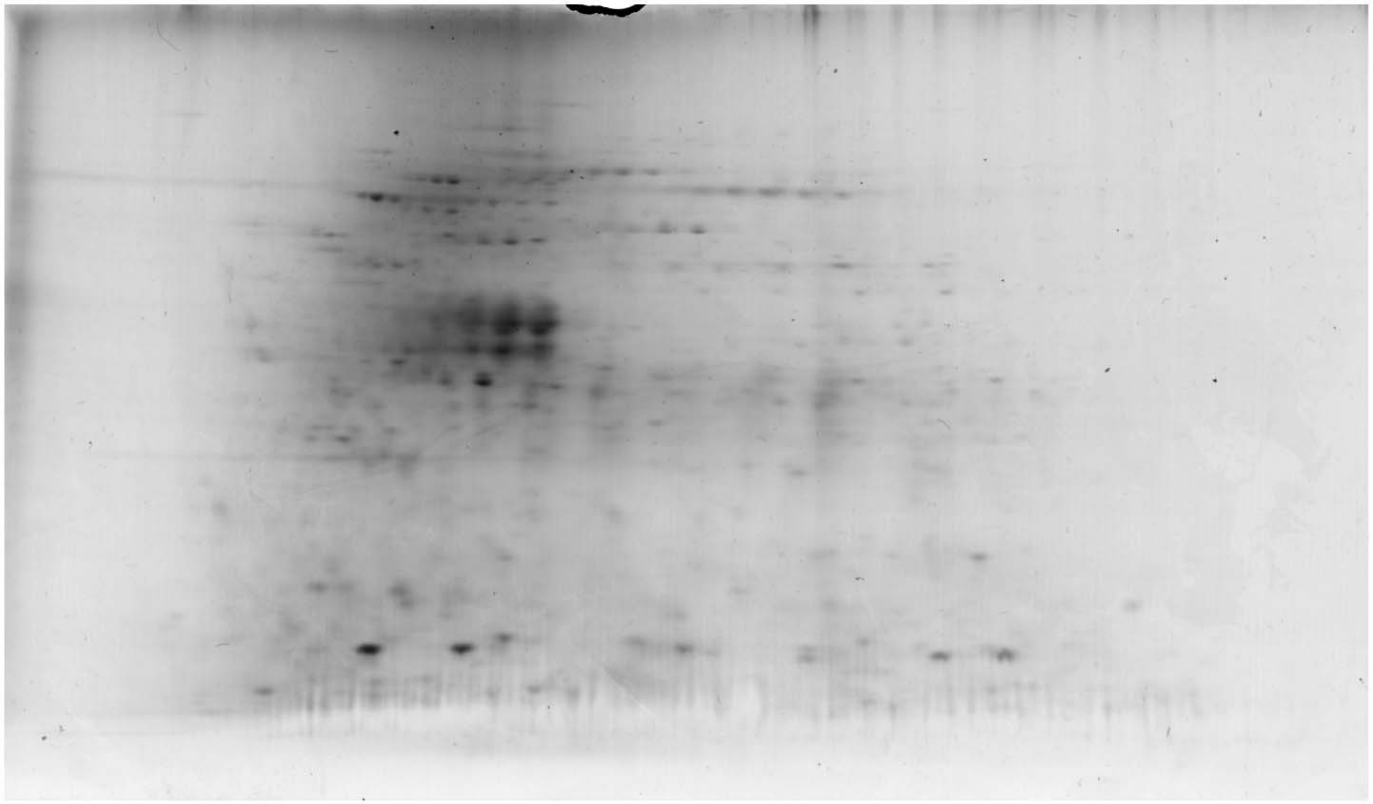
PBI Pressure BioSciences Inc.

Isolation of proteins from *Frankia* mycelium and vesicles by PCT



PBI Pressure BioSciences Inc.

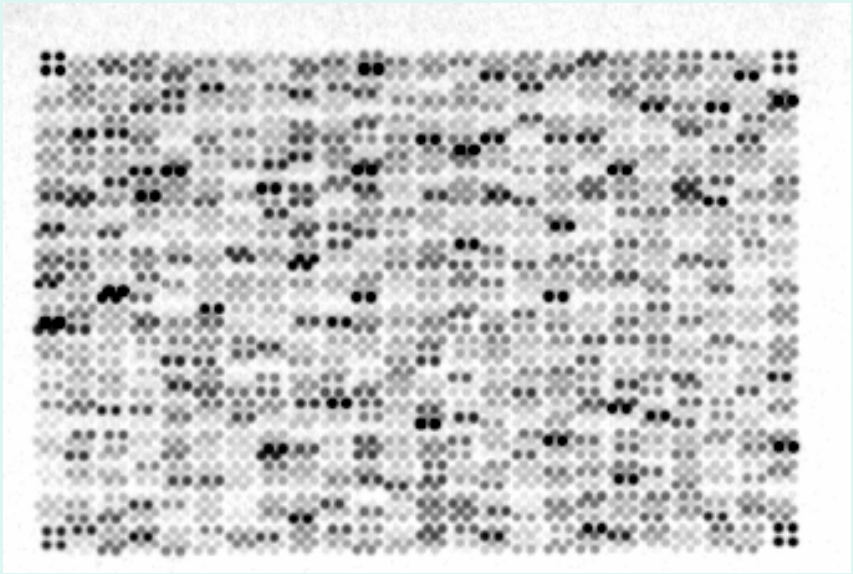
Two-dimensional gel electrophoresis of vesicle proteins isolated by PCT



PBI Pressure BioSciences Inc.

Transcriptome Studies

- A pilot project for Strains Ccl3 and EAN1pec high-density DNA arrays
- 12 perfect match probes/gene
- 12 mismatch probes/gene



Perspectives

- Gene Duplications
- Comparative Genomics
- Evolutionary aspects
- Proteomic Profiles
- Transcriptome Profiles (DNA arrays)
- Genetics

Acknowledgements

This work was supported by:
USDA Hatch grant 486; USDA 2003-01127;
NSF EF-0333173; DOE Microbial Genome Program

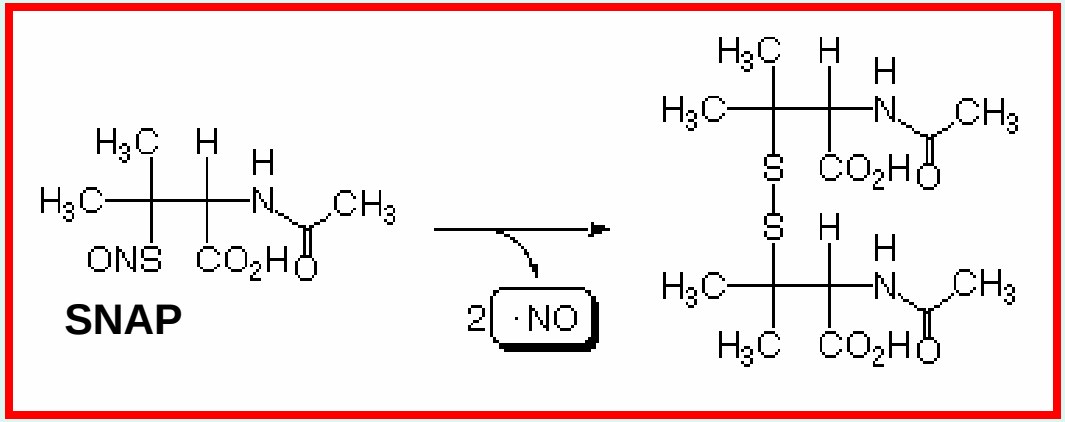
Secondary Metabolites

- Polyketide synthases
- Non-ribosomal peptide synthetases
- CDA synthase
- EAN Red Pigment

TrHb Function with Nitric Oxide: (Cultures +/- NO)

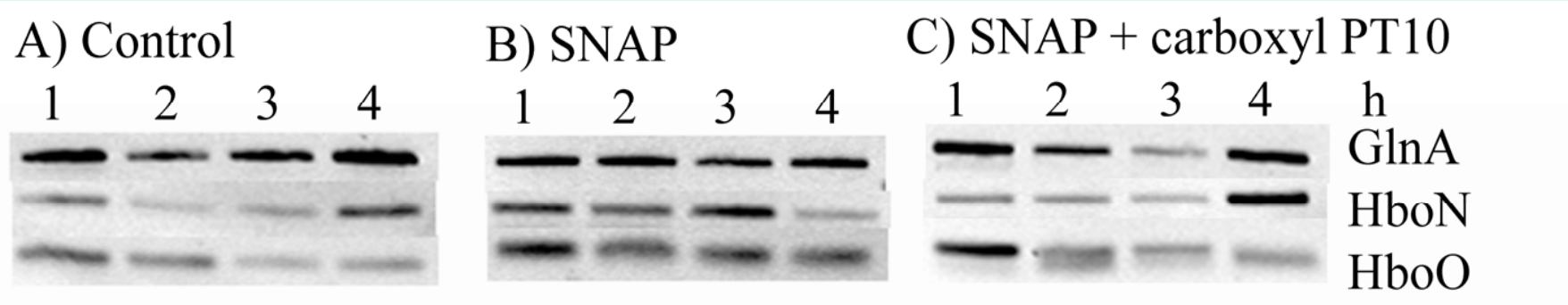
- Prediction: ↑ expression of trHbN w/ NO
  - *Mycobacterium tuberculosis* model

- SNAP: spontaneous NO donor

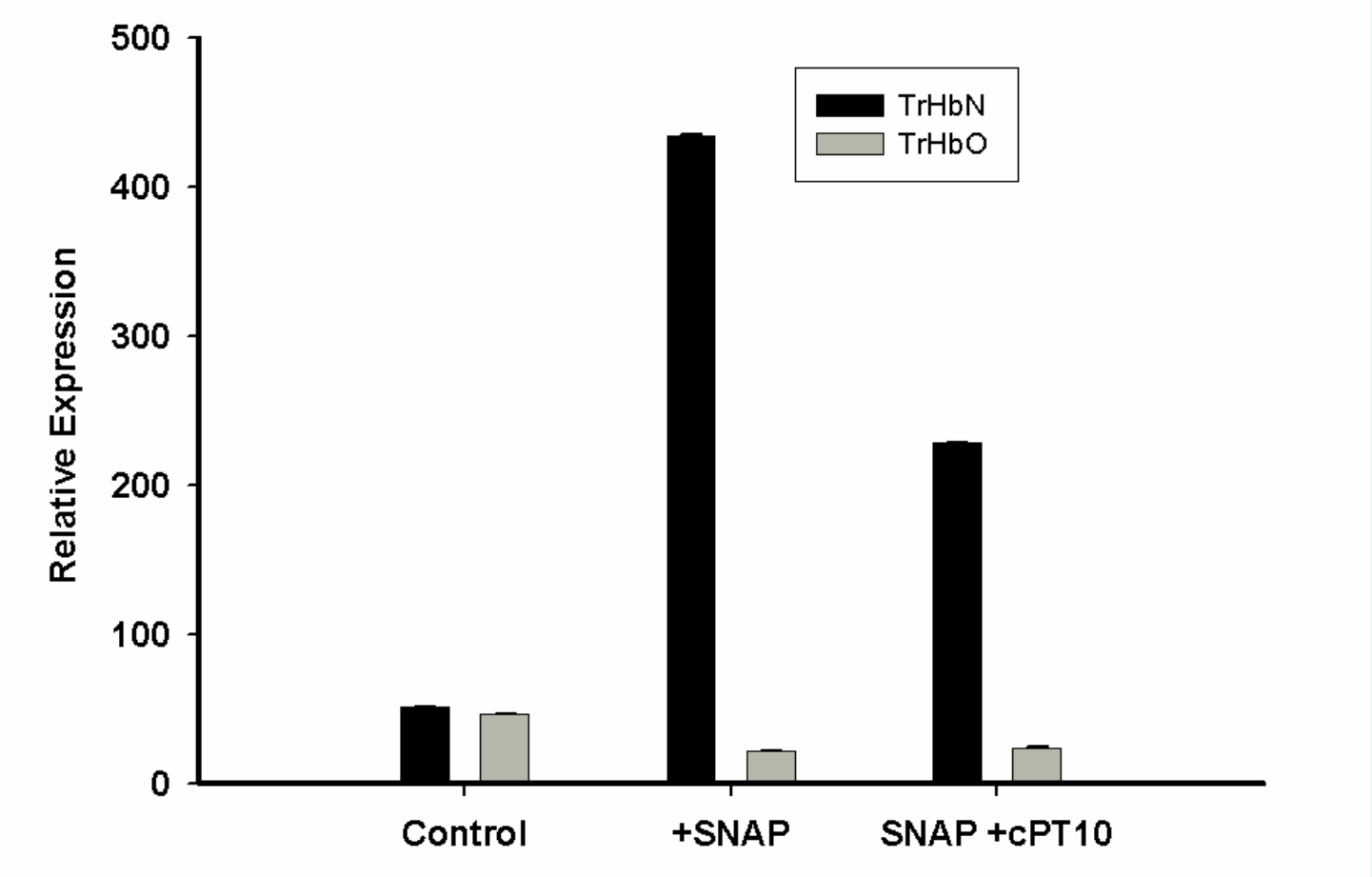


www.dojindo.com/newprod/ 1/no/nodonors/snap.html

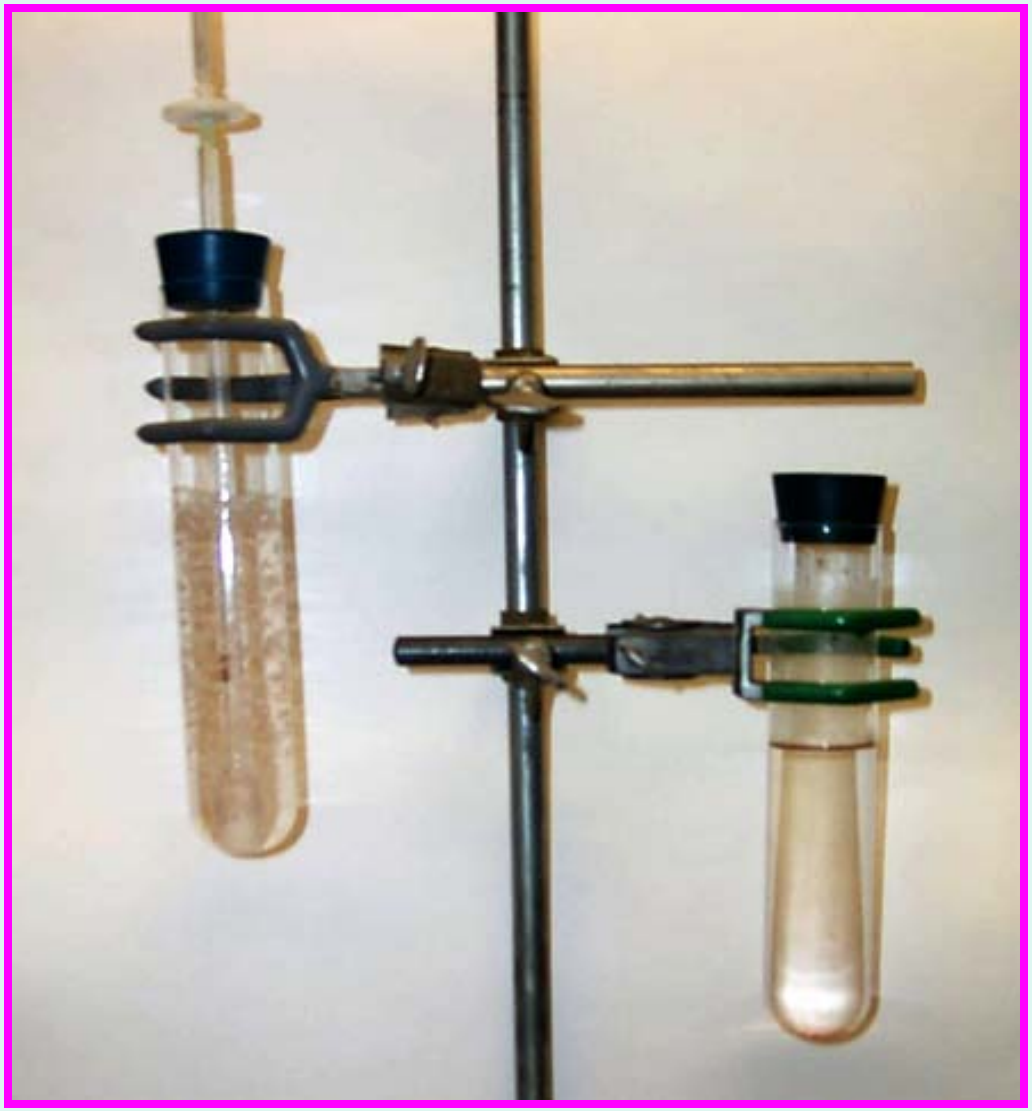
Transcription Analysis: RT-PCR



Transcription Analysis: qPCR (SYBR Green)



TrHb Function in Respiration:
(Aerated vs. Static Cultures)



Prediction: ↑ expression of trHbO in static culture

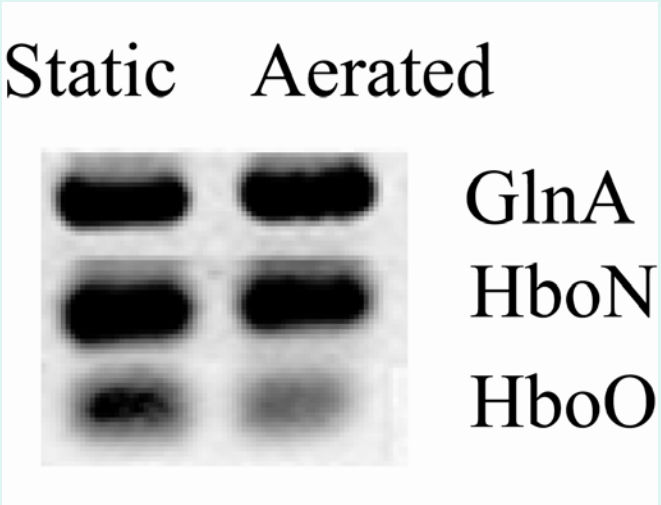
Mycobacterium tuberculosis model

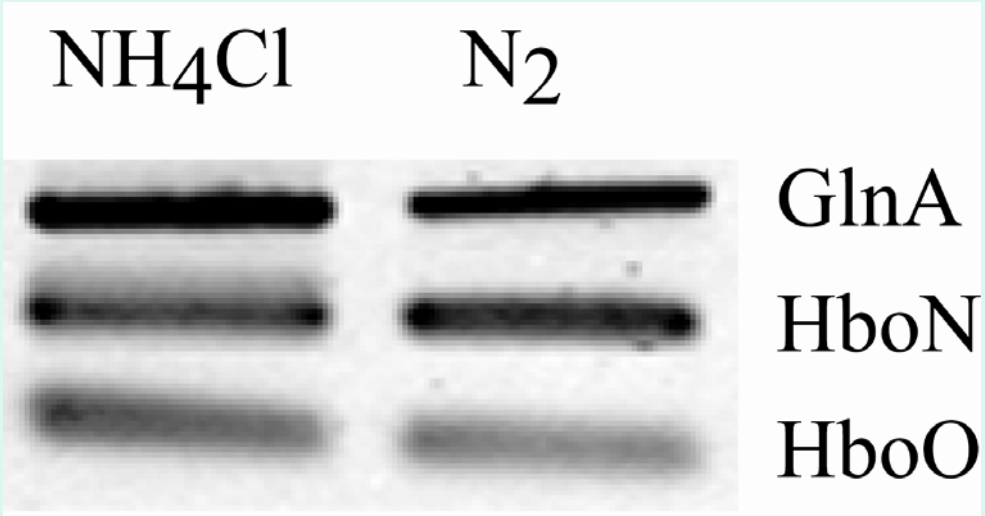
Aerated: ↑ 0_2

-bubbled with atm 0_2)

Static: ↓ 0_2

Transcription Analysis: RT-PCR





HboO expression is up-regulated under hypoxic conditions

Nitrogen status did not significantly affect expression