

# Multiple Sequence Alignment of the mexB gene in *Pseudomonas* *aeruginosa*

Ally Watkins

# Introduction

- *Pseudomonas aeruginosa*
  - Gram negative
  - Non-spore forming
  - Rod
  - Versatile
- Multidrug resistance -> Challenging pathogen to treat
- Investigating multidrug resistance gene to determine
  - Functional importance of resulting protein
  - Evolutionary relationships between species

# Finding Data

National Library of Medicine  
National Center for Biotechnology Information

Gene Gene Pseudomonas aeruginosa Advanced

Full Report Send to: Download Datasets

**mexB multidrug resistance protein MexB [ Pseudomonas aeruginosa PAO1 ]**

Gene ID: 877852, updated on 6-Oct-2023

**Summary**

Gene symbol: mexB  
Gene description: multidrug resistance protein MexB  
Locus tag: PA0426  
Gene type: protein coding  
RefSeq status: PROVISIONAL  
Organism: Pseudomonas aeruginosa PAO1 (strain: PAO1)  
Lineage: Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas

**New Try the new Gene table**  
**New Try the new Transcript table**

**Genomic context**

Sequence: NC\_002516.2 (473191..476331)

**NC\_002516.2**  
Genomic Context describing neighboring genes

**Genomic regions, transcripts, and products**

Genomic Sequence: NC\_002516.2

Go to reference sequence details  
Go to nucleotide: Graphics FASTA GenBank

Pseudomonas\_protein.fas

```
>P_aeruginosa
MSKFFIDRPIFAWVIALVIMLAGGLSILSLPVNQYPAIAPPAIAVQSYPGASAETVQDTVVQVIEQQMN
GIDNLRYISSESNSDGSMTITVTFEQGTDPIAQVQVNKLQLATPLLPQEVRQRQGIRVTKAVKNFLMV
GVVSTDGSMTKEDLSNYIVSNIQDPLSRTKGVGDFQVFSGSQYSMRIWLDPAKLNSYQLTPGDVSSAIQAQ
NVQISSGQLGGLPAVKGQQLNATIIGKTRLQTAEQFENILLKVNPDGSQLVRLKDADVLGLGGQDYSINAQ
FNGPASGIAIKLATGANALDTAKAIRQTIANLEPFMPQGMKVVPYDTPVVSASIHEVVKTLGAEILL
VFLVMYLFLQNFRATLIPTIAVPPVLLGTFGVLAAGFSINTLTMFGMVLAIGLLVDDAIVVVENVERVM
AEEGLSPREAARKSMGQIQGALVGIAVMVSAVFLPMAFFGGSTGVYRQFSITIVSAMALSVIVALILTP
ALCATMLKPIEKGDHGEHKGGFFGWFNRMFLSTTHGYERGVASILKHRAPIYLLIVVAGMIWMFTRIP
TAFLPDEDQGVLFQAQVTPPGSSAERTQVVDSMREYLLEKESSVSSVFTVTGFNFAGRQSSGMAFIM
LKPWEERPGGENSVFELAKRAQMHHFFSKDAMVFAFAPPSSVLELGNATGFDLFLQDQAGVGHEVLLQARN
KFLMLAAQNPALQRVRPNMGSDEPQYKLEIDDEKASALGVSLADINSTVSIAWGSSYVNDFIDRGRVKRV
YLQGRPDARMNPDDLSKWYVRNDKGEMVPFNATGKWEYGSPKLERYNGPAMEILGEPEPAPGLSSGDAM
AAVEEIVKQLPKGVGYSWTGLSYEERLSGSQAPALYALSLLVVFLCLAALYESWSIPFSVMLVVPLGVIG
ALLATSMRGLSNDVFFQVGLLTTIGLSAKNAILIVEFAKELHEQGKGIVEAAIEACRMRLRPIVMTSLAF
ILGVVPLAISTGAGSGSQHAIGTGVIGGMVTATVLAIFWVPLFYAVSTLFKDEASKQQASVEKGQ
```

Gene and amino acid sequence copied and pasted into separate .fas files

# Finding Close Relatives

National Library of Medicine  
National Center for Biotechnology Information

Log in

BLAST® » blastp suite

Home Recent Results Saved Strategies Help

blastn blastp blastx tblastn tblastx Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more... Reset page Bookmark

**Enter Query Sequence**

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>NP\_249117.1 multidrug resistance protein MexB [Pseudomonas aeruginosa PAO1]  
MSKFFIDRPIFAWVIALVIMLAGGLSILSPVNQYPAIAPPAlAVQSYPGASAETV  
QDTVVQVIEQQMN

Query subrange [?](#)  
From   
To

Or, upload file [Choose File](#) No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

**Choose Search Set**

Databases  Standard databases (nr etc.) [New](#)  Experimental databases [Try experimental clustered nr database](#)  For more info see [What is clustered nr?](#)

Compare  Select to compare standard and experimental database [?](#)

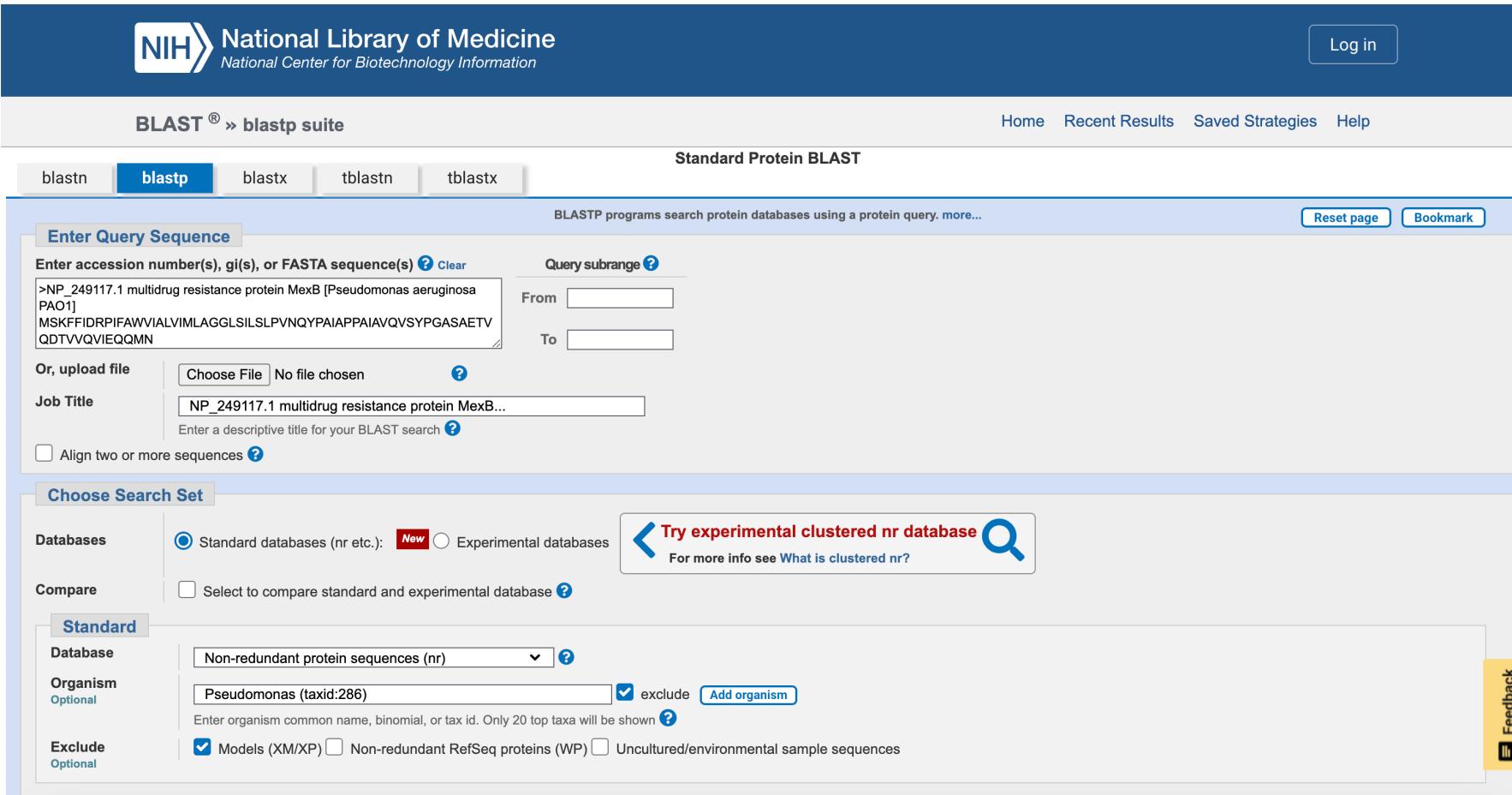
**Standard**

Database Non-redundant protein sequences (nr) [?](#)

Organism **Pseudomonas (taxid:286)**  exclude [Add organism](#)  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude **Models (XM/XP)**  Non-redundant RefSeq proteins (WP)  Uncultured/environmental sample sequences

Feedback



BLASTn and BLASTp run excluding *pseudomonas*

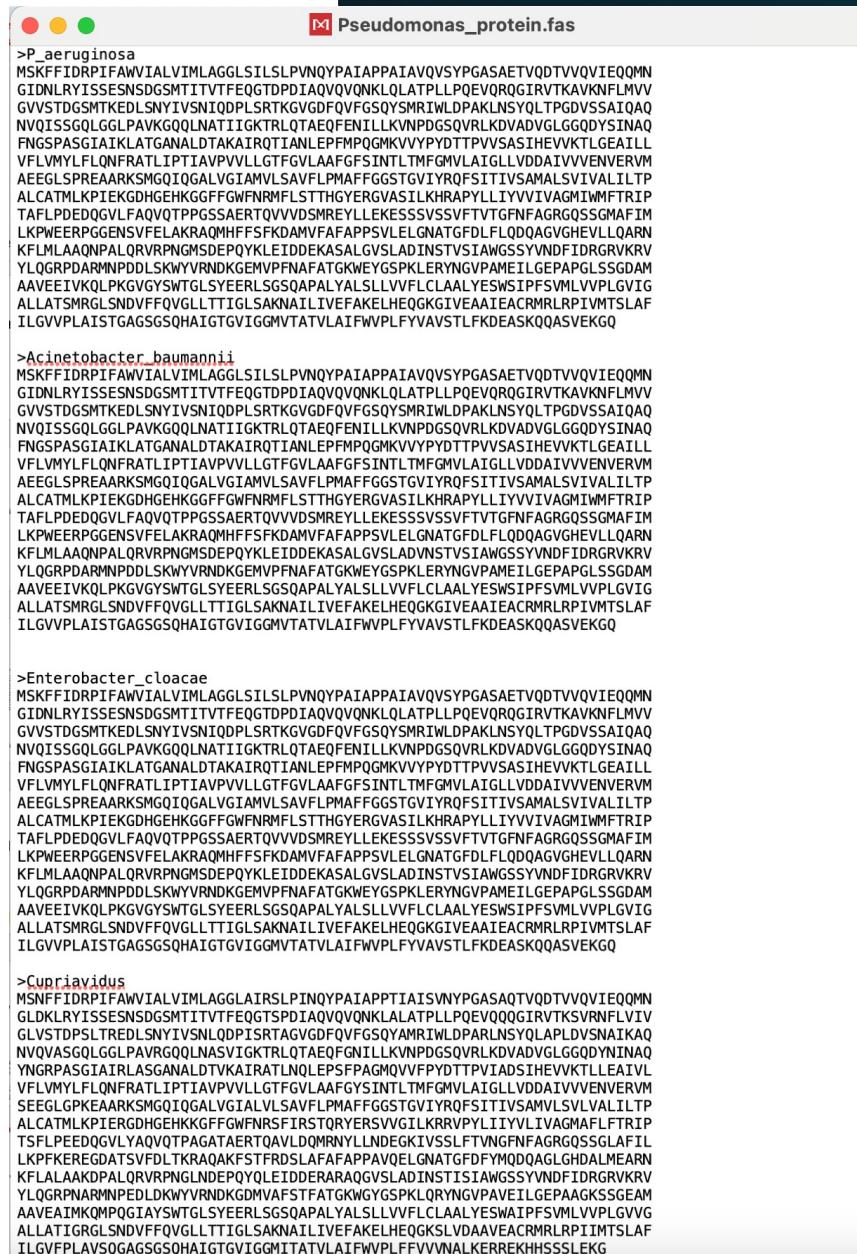
# Retrieving Sequences

- Top species from unique genera were selected
- FASTA with nucleotide or amino acid species was restricted to range

The screenshot shows a protein sequence retrieval interface. At the top, there is a search bar with dropdown menus for 'Protein' and 'Advanced'. Below the search bar, a message states: 'This record is a non-redundant protein sequence. Please [read more here](#)'. A 'Send to:' button is followed by a 'Change region shown' dropdown. This dropdown has two options: 'Whole sequence' (radio button) and 'Selected region' (radio button, which is selected). It also includes input fields 'from: begin' and 'to: 1045' and a 'Update View' button. To the right of this, a large box displays the protein sequence information. The title is 'MULTISPECIES: efflux RND transporter permease subunit [Cupriavidus]'. It shows the NCBI Reference Sequence: WP\_144196336.1. Below the title is the sequence identifier >WP\_144196336.1:1-1045. The sequence itself is a long string of amino acids. On the right side of the sequence box, there are links for 'See 2 more title(s)', 'See all Identical Proteins(IPG)', 'Range 1: 1 to 1045', 'GenPept', and 'Graphics'. Below the sequence box, there is a table with columns: Score, Expect, Method, Identities, and Positives. The values are: 1797 bits(4655), 0.0, Compositional matrix adjust., 861/1045(82%), and 962/1045(92%). The 'Score' row also includes a link 'Query 1'.

# Preparing .fas files

- The nucleotide and protein .fas files were updated to include the top 3 species from each respective BLAST



```
>P_aeruginosa
MSKFFIDRPIFAWIALVIMLAGGLSILSLPVNQYPAIAPPかいAVQSYPGASAETVQDTVVQVIEQQMN
GIDNLRYISSESNSDGSMITITVTFEOGTDPIAQVQVQVQNLQLATPLPQEVQRQGIRVTKAVKNFLMV
GVVSTDGSMTKEDELSNYIVSNIDQPLSRTKGVDQFVFGSQYSMRLWDPAKLNSYQLTTPGDVSSAIQAQ
NVQISSGQLGLPALKQQLNATIIGKTRLQTAEQFENILLKVNPDGSQVRLKDADVGLGGQDYSINAQ
FNGSPASGIAIKLATGANALDTAKAIRQTIANLEPFMPQGMKVVPYDTPVVSASIHEVKTLEAII
VFLVMYLFQNFRATLIPTIIAVPVLLGTGFVLAAGFSGINTLTMFGVLAIGLLVDDAIVVVENVERVM
AEEGLSPREARKSMGIOQGALVGIAVLSAVFLPMAFFGGSTGVYIYRQFSITIVSAMALSIVALILTP
ALCATMLKPIEKGDHGEHKGGFFGWNRMFSLTTGYERGVASILKRAPYLLYYVIVAGMIMMFTTRIP
TAFLPEDOGVLFAQVTPPGGSAERTQVVDMSREYLLKEESSSSVFTVTFGNFAGRGQSSGMFIM
LKWEERPGGENSFELAKRAQMHHFSFKDAMVFAFAPPVLELGNTAGFDLFLQDQAGVGHEVLLQARN
KFLMAAQNPALQRVRPNMGSDEPQYKLEIDDEKASALGVSLADINSTSIAWGSYVNDFIDRGRVKRV
YLOGRPDARMNPDDLSKWYVRNDKGEMVPFNAFATGKWEYGPKLERYNGVPAMEILGEPAPGLSSGDAM
AAVEETVKOLPKGVGVYWTGLSYEERLSSGQAPALYALSLVVFLCLAALEYESWIPFSVMLVPLGVIG
ALLATSMRGLSNDVFFQVGLLTTIGLSAKNAILIVEFAKELHEQKGIVEAAIEACRMRLRPIVMTSLAF
ILGVVPLAISTGAGSGSQHAIGTGVIGGMVTATVLAIFFWPLFYVAVSTLFKDEASKQQASVEKGQ

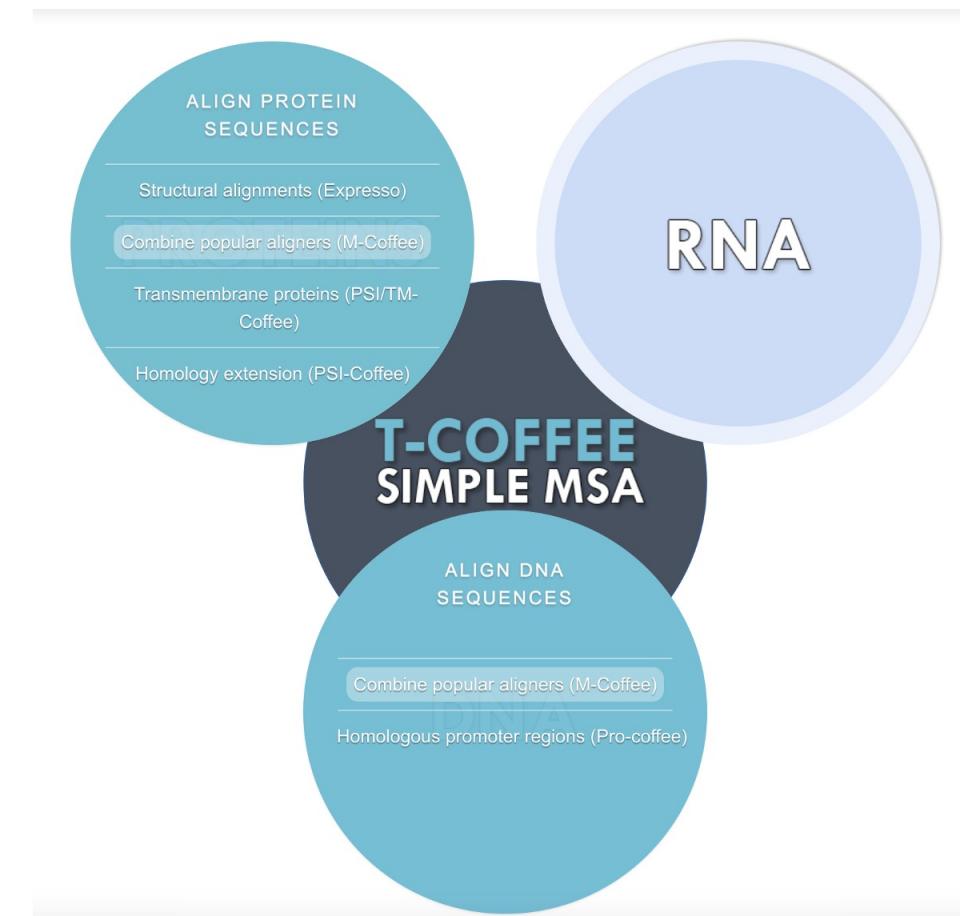
>Acinetobacter_baumannii
MSKFFIDRPIFAWIALVIMLAGGLSILSLPVNQYPAIAPPかいAVQSYPGASAETVQDTVVQVIEQQMN
GIDNLRYISSESNSDGSMITITVTFEOGTDPIAQVQVQVQNLQLATPLPQEVQRQGIRVTKAVKNFLMV
GVVSTDGSMTKEDELSNYIVSNIDQPLSRTKGVDQFVFGSQYSMRLWDPAKLNSYQLTTPGDVSSAIQAQ
NVQISSGQLGLPALKQQLNATIIGKTRLQTAEQFENILLKVNPDGSQVRLKDADVGLGGQDYSINAQ
FNGSPASGIAIKLATGANALDTAKAIRQTIANLEPFMPQGMKVVPYDTPVVSASIHEVKTLEAII
VFLVMYLFQNFRATLIPTIIAVPVLLGTGFVLAAGFSGINTLTMFGVLAIGLLVDDAIVVVENVERVM
AEEGLSPREARKSMGIOQGALVGIAVLSAVFLPMAFFGGSTGVYIYRQFSITIVSAMALSIVALILTP
ALCATMLKPIEKGDHGEHKGGFFGWNRMFSLTTGYERGVASILKRAPYLLYYVIVAGMIMMFTTRIP
TAFLPEDOGVLFAQVTPPGGSAERTQVVDMSREYLLKEESSSSVFTVTFGNFAGRGQSSGMFIM
LKWEERPGGENSFELAKRAQMHHFSFKDAMVFAFAPPVLELGNTAGFDLFLQDQAGVGHEVLLQARN
KFLMAAQNPALQRVRPNMGSDEPQYKLEIDDEKASALGVSLADINSTSIAWGSYVNDFIDRGRVKRV
YLOGRPDARMNPDDLSKWYVRNDKGEMVPFNAFATGKWEYGPKLERYNGVPAMEILGEPAPGLSSGDAM
AAVEETVKOLPKGVGVYWTGLSYEERLSSGQAPALYALSLVVFLCLAALEYESWIPFSVMLVPLGVIG
ALLATSMRGLSNDVFFQVGLLTTIGLSAKNAILIVEFAKELHEQKGIVEAAIEACRMRLRPIVMTSLAF
ILGVVPLAISTGAGSGSQHAIGTGVIGGMVTATVLAIFFWPLFYVAVSTLFKDEASKQQASVEKGQ

>Enterobacter_cloacae
MSKFFIDRPIFAWIALVIMLAGGLSILSLPVNQYPAIAPPかいAVQSYPGASAETVQDTVVQVIEQQMN
GIDNLRYISSESNSDGSMITITVTFEOGTDPIAQVQVQVQNLQLATPLPQEVQRQGIRVTKAVKNFLMV
GVVSTDGSMTKEDELSNYIVSNIDQPLSRTKGVDQFVFGSQYSMRLWDPAKLNSYQLTTPGDVSSAIQAQ
NVQISSGQLGLPALKQQLNATIIGKTRLQTAEQFENILLKVNPDGSQVRLKDADVGLGGQDYSINAQ
FNGSPASGIAIKLATGANALDTAKAIRQTIANLEPFMPQGMKVVPYDTPVVSASIHEVKTLEAII
VFLVMYLFQNFRATLIPTIIAVPVLLGTGFVLAAGFSGINTLTMFGVLAIGLLVDDAIVVVENVERVM
AEEGLSPREARKSMGIOQGALVGIAVLSAVFLPMAFFGGSTGVYIYRQFSITIVSAMALSIVALILTP
ALCATMLKPIEKGDHGEHKGGFFGWNRMFSLTTGYERGVASILKRAPYLLYYVIVAGMIMMFTTRIP
TAFLPEDOGVLFAQVTPPGGSAERTQVVDMSREYLLKEESSSSVFTVTFGNFAGRGQSSGMFIM
LKWEERPGGENSFELAKRAQMHHFSFKDAMVFAFAPPVLELGNTAGFDLFLQDQAGVGHEVLLQARN
KFLMAAQNPALQRVRPNMGSDEPQYKLEIDDEKASALGVSLADINSTSIAWGSYVNDFIDRGRVKRV
YLOGRPDARMNPDDLSKWYVRNDKGEMVPFNAFATGKWEYGPKLERYNGVPAMEILGEPAPGLSSGDAM
AAVEETVKOLPKGVGVYWTGLSYEERLSSGQAPALYALSLVVFLCLAALEYESWIPFSVMLVPLGVIG
ALLATSMRGLSNDVFFQVGLLTTIGLSAKNAILIVEFAKELHEQKGIVEAAIEACRMRLRPIVMTSLAF
ILGVVPLAISTGAGSGSQHAIGTGVIGGMVTATVLAIFFWPLFYVAVSTLFKDEASKQQASVEKGQ

>Cupriavidus
MSNFFIDRPIFAWIALVIMLAGLAIERSLPIQYPAIAPPかいAVQSYPGASAETVQDTVVQVIEQQMN
GLDKLRYISSESNSDGSMITITVTFEOGTDPIAQVQVQVQNLALATPLPQEVQQQGIRVTKSVRNFLVIV
GLVSTDPSLTREDLSNYIVSNIDQPLSRTKGVDQFVFGSQYANMRIWLDPARLNSYQALPDVSNAIKAO
NVQVASGQLGLPALKQQLNAsVIGKTRLQTAEQFENILLKVNPDGSQVRLKDADVGLGGQDYNINAQ
YNGRPASGIAIRLASGANALDTVKAIRATLNQLEPSFPAGMQVVFYDTPVIAOYIHEVKTLEAII
VFLVMYLFQNFRATLIPTIIAVPVLLGTGFVLAAGFSGINTLTMFGVLAIGLLVDDAIVVVENVERVM
SEEGLGKEAARKSMGIOQGALVGIAVLSAVFLPMAFFGGSTGVYIYRQFSITIVSAMVLSVVALILTP
ALCATMLKPIEKGDHGEHKGGFFGWNRNSFIRSTORYERSVGILKRRPYLIYYVIVAGMAMFLFTRIP
TSFLEEDOGVLYAQVTPAGATAERTQAVLDQMRNYLNDEGKIVSSLFTVNGFNFAGRGQSSGLAFIL
LKPFKEREQDGTASVFDLTKRAQKFSTRDLSLAFAFAPPVQELGNATGDFYMQDQAGLGHDALMEARN
KFLALAACKDPLQRVRPNGLNDEPQYQLEIDDERAARQGVSLADINSTISIAWGSYVNDFIDRGRVKRV
YLOGRPNARMNPDELDKWWYVRNDKGDMVAFSTFATGKWEYGPSPKLQRYNGVPAVEILGEPAGKSSGEAM
AAVEAIMKOMPQGIAYSWTGLSYEERLSSGQAPALYALSLVVFLCLAALEYESWIPFSVMLVPLGVIG
ALLATIGRGLSNDVFFQVGLLTTIGLSAKNAILIVEFAKELHEQKGSLVDAAVEACRMRLRPIVMTSLAF
ILGVVPLAVSQAGSGSQHAIGTGVIGGMVTATVLAIFFWPLFYVAVSTLFKDEASKQQASVEKGQ
```

# Creating the Alignments

- Used T-Coffee Simple Multiple Sequence Aligner
  - Input .fas files to align *P. aeruginosa* with closely related species
  - Performed for DNA and Protein
  - M-Coffee: Combines results between other aligners showing agreement



# T-Coffee

- Color coding indicates level of agreement between aligners
- Score tells us consistency across each aligner from 1-100
  - <50 = unreliable
- Normal for DNA to appear more distantly related

## MSA

The multiple sequence alignment result as produced by T-coffee.

T-COFFEE, Version\_11.00 (Version\_11.00)

Cedric Notredame

SCORE=306

\* BAD AVG GOOD

\* P aeruginosa : 31

C campinensis : 36

R pickettii : 31

S stutzeri : 30

cons : 30

P aeruginosa ATGTCGAAGTTTTCATGTAG-GCCCATTTCTCGTGGGTGATCG--CCTTGGTAT-CATG

C campinensis TCCCTGA-----GGCGTTACCAACAGAAAGAACCGGCCAGAGAACGATGGCAG

R pickettii ATGTCATTCTTATCGATCG-GCCCATCTTCGCTGGGTGATCG--CCTTGGTAT-CATG

S stutzeri -----CACAGAAACAGCGGCCACCCAGAACGATCACCG

cons :

\*\*\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa CTC-----GCAGGGCGCTCTCGATCTCACTGCTCGGGTAACAGTACCCGGCATCGCCC

C campinensis CACCGTGGCGTGATCGCCGCGATCACCGGGTACCGATGGCGT-----GC---T

R pickettii CTG-----GCAGGGCGCTCTCGATCGCCGCGATCACCGGGTACCGATGGCGT-----GC---G

S stutzeri CACGGTGGCGTGACCATCCGCCGATCACCGGGTACCGATGGCGT-----GC---G

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa GCGGCCATCGCGTGCAGGTGAGCTACC-CGGGCCTCGGAGACGGTGAGGACACGGTGG

C campinensis GGCCTCCAGCGCCCTTCCGACACCGCAACGGGAACACGGGAGATGACGCCAGCAGG

R pickettii GCCACCATCGCCATCTCGCAACTACC-CGGCTGCTCGGAGACGGTGAGGACACGGTGG

S stutzeri AACTGCTCGCCGCCCGCTGGCATGGCCAGCGGCCACACGGCAGAGTGAGGACACGGTGG

cons :

\*\*\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa TCCAGGTGATCGAGCAGCATG-AACGGGATCGACAATCTCGCTACATCTCTCGGAGAGTAAC

C campinensis TCATGATGATGGGT-----CGCAGGGCATCGGCACGC-CTCCAGGGCTGATCGAC-CAGAGCTCTGC

R pickettii TGCAAGGTGATCGAGCAGCATG-AACGGGCTGGACGCCGCTGCTACATCTCTCGGAGAGTAAC

S stutzeri TCATGATGATCGGA-CGCAGGGCATCGGCCAAGC-CTGACCGCTCCCTCGAA-CAGACCATGC

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa TCCG-ACG-GCAGCA-TGACCATCACCGTGACCT--TCG--AACAGGGC-ACCGACC--CCG

C campinensis CTGCTGCTGCAACTCTTGGGAACTCCAGCATGGGACAGCTGGCTCTGGGACAGGGCATGG

R pickettii GCGC-ACG-GCAACA-TGACGATCACCGTGACCT--TCG--AGCAAGGC--ACCAATC-CGG

S stutzeri CGTCGCGGATCACAGATCTTGGCAACTCGACAATCAGATGGCTTCTCGCTGACGACCATGC

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa ACATCGGCGGAGGTG-CAGAACAGCTGC--AACTGGGCCACCC-----CG-CTACTGCCG

C campinensis TCGTCG-CAGGGGACCTTGGAAAGAACCTGCTGGGAAAGCAGGGCCCAACTCTGGGACAGC

R pickettii ATATGGCCAGGTG-CAGAACAGCTGC--CCCTGGGACAGC-CTCTGGGACAGC

S stutzeri TTGGTCAG-CAGGGGACCTGGAAGAACACGTCGGTGTGCAAGGCCACGGGGCTGGCTGGCAGATG

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa AGGAAGTGCAGGCC-AGGG-GA--TCGGT-GTGAACAGG-CGGTAAGAACCTCTCATGGTG

C campinensis GGCGCACAC-GCCGAGCGGACCAACAGCATCCAGGAAACCGGAATGCCAGCTT--TCAT

R pickettii AGGAAGTGCAGCAGC-AAGG-CA--TCGGT-GTGAACAGG-CGGTAAGAACCTCTCATGGTG

S stutzeri GGCGCATGGACCAACGGACAGCATCCAGGAAAGGGATGCCAGCTC

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa TCGGTGTTCTTCC-ACCGACGGCAGCATGACCAAGGAAGACCTGTCGAACATCTGGTCAAC

C campinensis ACAGCGGCCAGGCACAG-GAACACACAGCAGGGACAGCGCTAGAGCCGCCCTGG--

R pickettii TGGGCTGTTCTTCC-ACCTGACCCAAAGTCACCGGGGAAGACCTGTCGAACATCTGGTCAAC

S stutzeri ACAGCGGCCAGGCAGAG-GAACACACAGCAGGGACAGCGCTAGAGCCGCCCTGG--

cons :

\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa ATCCAGG-ACCCACTCTC-GCGGACCAAGGGCTCGTGACTCTCAGGTGTT--CGGCT-C-

C campinensis AACCCGACAGCGCTCTCG--TAGA-CAGGGGGCTCGACAGGAGAACCTGTCGAATGCCCT

R pickettii ATCCAGG-ACCCGCTCTC-GCGGACCTACGGGGCTGGGACCTTCAGGTGTT--CGGCT-C-

S stutzeri CGCGGAGGGAGCTCTCG--AAGGA-AAGCCGGTCCACGACACGCCGATCCAGGCCAGCT

cons :

\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

P aeruginosa -----GCAGTACTCGA

C campinensis -----GGCGCA-TC-T

R pickettii -----CGAGTACCG-C

S stutzeri -----GCTGCGCCAGCGCTCGA-C

cons :

\* \* \*

T-COFFEE, Version 11.00 (Version\_11.00)

Cedric Notredame

SCORE=998

\* BAD AVG GOOD

\* P aeruginosa : 99

Acinetobacter b : 99

Enterobacter cl : 99

Cupriavidus : 99

cons : 99

MSKFIDPFWAVLALVIMLAGLGLSILSLPVNOYPATAPPAZAVOYGPAGAETVQDVTVOQ

# T-Coffee Alignments

- Low level of agreement for DNA alignment
    - All results scored < 50
  - Different DNA sequences may encode same protein

MSA  
The multiple sequence alignment result as produced by T-coffee.

T-COFFEE, Version\_11.00 (Version\_11.00)  
Cedric Notredame  
SCORE=306

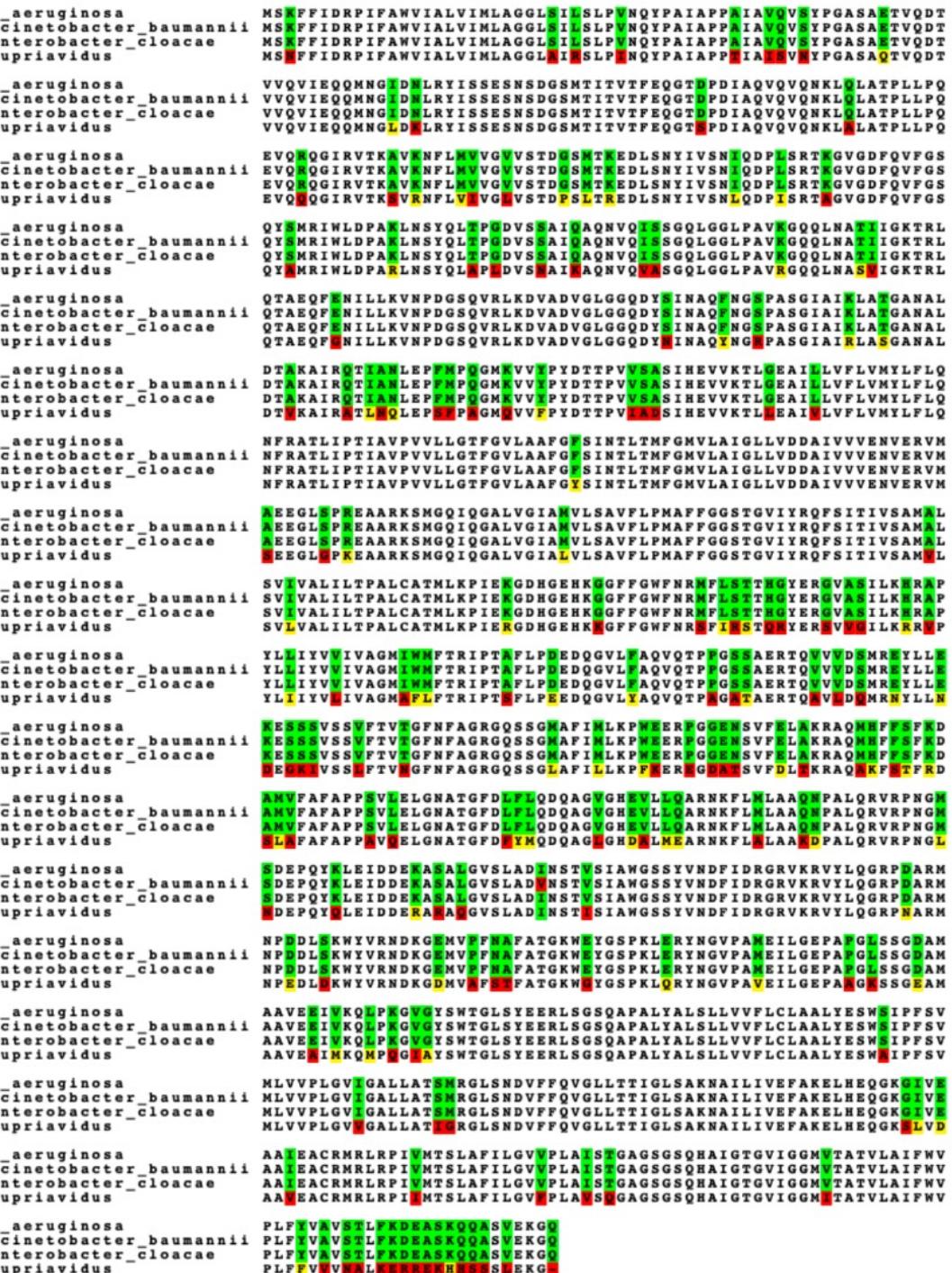
	BAD	AVG	GOOD
P aeruginosa	31		
C campinensis	36		
R pickettii	31		
S stutzeri	30		
S stutzeri	30		
P aeruginosa	ATGTGCGAAGTTTTCATGATAG-GCCCATTTGGTGGGTGATCG--CCTTGATGAT-CATG		
C campinensis	TCCTTGA-----GGCGGTTACCAACAGCAAGAACAGCGGCACCCAGAAAGATGGCAG		
R pickettii	ATGCCAATTCTTATCGATCG-GCCCATCTCGGTGGGTGATCG--CCTTGATGAT-CATG		
S stutzeri	CACGAAAGAACAGCGGCACCCAGAAAGATCACCG		
cons		* * * * * * * * * *	
P aeruginosa	CTC-----GCGGCCGCGCTGCGATCTCAGTCTGGCGTCAACCAGTACCCGCCATGCC		
C campinensis	CACCGTGGCGTGTGATCATGCGCCGATCACGGCGTACCGATGCGT-----GC---T		
R pickettii	CTG-----CACGGTGGCGTGTGATCCGCGGCGTGCACAGCTGATTCGGCCATGCC		
S stutzeri	GC-----CACGGTGGCGTGTGACCATCCCGCGATCACGGCGTACCGATGCGT-----GC---G		
cons		*	
P aeruginosa	GCGGCCATGCGCTGCAAGGTGACTTACCGGGCCTCGGCGAGACGGTGAGGACACGGT		
C campinensis	GCTCTGGCGAGCGGGCGCTTGCGACACCGCAGGAAACACGCCGAGGATGACGCCAGGAG		
R pickettii	GCCACCATGCGCATCTCGTCACTACCGGTGCTCGGCGAGACGGTGAGGACACGGT		
S stutzeri	AACTGCGCTGCGCCGCGCCGTCGGCCATGCCGACCGCAGGAGTGAAGGCCAGGAG		
cons		*	
P aeruginosa	TCCAGGTGATCGAGCAGCAGATG-AACGGGATCGACAATCTGGCTACATCTCTGGAGGTAAC		
C campinensis	TCATGATGATGGGT-CGAGCGGCATCCGGCGACGC-CTTCAACGGCTGATCAG-CAGACTCTGG		
R pickettii	TGCGATGATCGAGCAGCGTC-AACGGCTGAGCGCTGGCTACATCTCTGGAGGACAC		
S stutzeri	TCATGATGATCGGA-CGAGCGGCATCGGCCAAG-CCTGACCGCTGAA-CAGACCATGCG		
cons		*	
P aeruginosa	TCCG-ACG-GCAGCA-TGACCATCACCGTGACTCTCG--AACAGGGC-ACCGACC--CCG		
C campinensis	TCGTCGCTGCGAATCTCTGGGAACTCCGAGGATCGGCTTCTGGGCAACGGCCATGG		
R pickettii	GCG-ACG-GAACAA-TGACATCGCTGACCTCTCG--AGCAAGGGC-AACCAATC-CGG		
S stutzeri	CGTCGCGGTACAGATCTGGGAACCTCGACAATCAGAATGGCTTCTCGCTGAGGCGATGG		
cons		*	
P aeruginosa	ACATGCCCAAGGTCAGGTG-CAGAACAAAGCTGC--AACTGGGCCACCC--CG-CTACTGCGC		
C campinensis	TCGTCAG-CAGGCCACCTGGAAGAACACGTCGTTGACAAGCCCCGGCAATCGCCGAGCAGC		
R pickettii	ATATGGCCCAAGGTCAGGTG-CAGAACAAAGCTGC--CCCTGGCGCAGC--CC-CTGTGTC		
S stutzeri	TGGTCAG-CAGGCCACCTGGAAGAACACGTCGTTGTCAGGCCAACCTGGGGTGGTCAGCTG		
cons		*	
P aeruginosa	AGGAAGTGCAGGCC-AGGG-GA--TCCGG-GTGACCAAGG-CGGTGAAGAACCTCTCATGGTG		
C campinensis	GGCCGACAC-GCCGAGCGCACCCACAGCATCACCAGAACGGAACTGCCAGCTT---TCAT		
R pickettii	AGGAAGTGCAGGCC-AAGG-CA--TCCG-GTGACCAAGT-CGTCGCAACTCTCTGGTCATCG		
S stutzeri	GGCCGATGAG-GCCCAAGGCCAACCCAGCATCACCGCAAGGGAACTGCCAGCTT---TCGT		
cons		*	
P aeruginosa	TGGGTGTTTCC-ACCGAGGGCAGCATGACCAAGGAAGACCTGTCGAACATCTGTTTCAAC		
C campinensis	ACAGCGGGCCAGGACACAG-GAACACCAACAGCAGGGACAGCGCTGAGGCCGCCCTGG--		
R pickettii	TGGGCTGATTTCC-ACTGACCCCAAGGTACCGGGGAACCTCTGCAACTACATCTGTTGAGT		
S stutzeri	ACAGCGGGCCAGGACAG-GAACACCAACAGCAGGGAGAGCGCTGAGGCCGCCCTGG--		
cons		*	
P aeruginosa	ATCCAGG-ACCCACTCTC-GCGGACCAAGGGCTCGGTGACTTCAAGGTGTT--CGGCT-C--		
C campinensis	AACCCGAGCAGCCCTCTG---TACGA-CAGGGCGTCAACGAGTAACTGCCCT		
R pickettii	ATCCAGG-ACCCGCTCTC-GCCGACCTACGGGGCTGGCGACTTCAAGGTGTT--CGGCG-C-		
S stutzeri	CGCGGAGAGGAGCTCTCG--AAAGA-AAGCCGCTCACGACACGCCATCCCAGCGCAGCT		
cons		*** * * *** * * * * * *	
P aeruginosa	-----GCAGTACTCGA		
C campinensis	-----GGCGCA-TC-T		
R pickettii	-----CGAGTAGCG-C		
S stutzeri	GCTGCGCCAGGGCTCGA-C		
cons		*	

# T-Coffee Alignments

- High level of agreement for protein alignment
    - All results score = 99
  - Protein is conserved across species

# Visualizing Protein Results

- Output files downloaded from T-Coffee
- Result.fasta\_aln file opened in boxshade to generate high quality output of MSA
  - Red = Different from consensus
  - Green = Identical to consensus
  - Yellow = Similar to the consensus



# Creating .mas

- MEGA Alignment Sequence file = .mas file
- Alignment created on MEGA using protein .fas file
- Results downloaded as .mas file
- This file is used on MEGA to create pairwise distance and phylogenetic trees

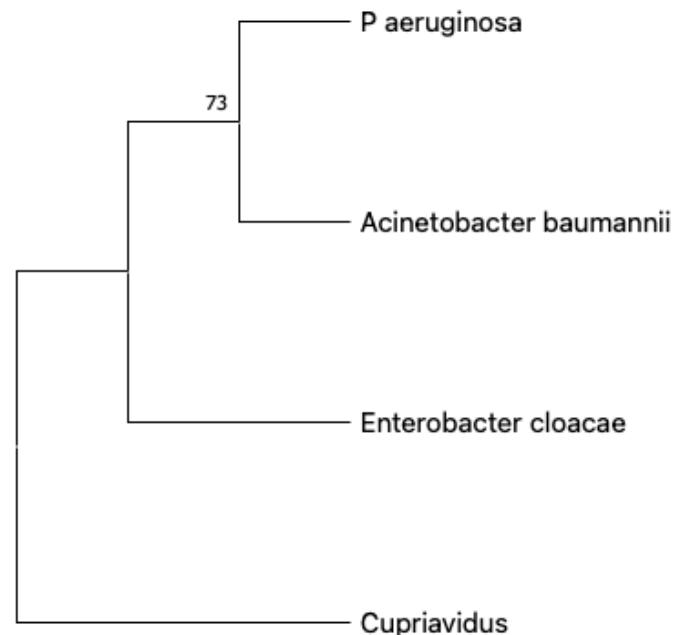
# Pairwise Distances

	1	2	3
1. <i>P aeruginosa</i>			
2. <i>Acinetobacter baumannii</i>	0.0009564802		
3. <i>Enterobacter cloacae</i>	 0.0000000000	0.0009564802	
4. <i>Cupriavidus</i>	0.1936776600	 0.1948397752	0.1936776600

- Estimation of evolutionary divergence between sequences
- Number corresponds with differences in amino acid sequences
- *Pseudomonas aeruginosa* and *Enterobacter cloacae* are most closely related
- *Acinetobacter baumannii* and *Cupriavidus* are least closely related

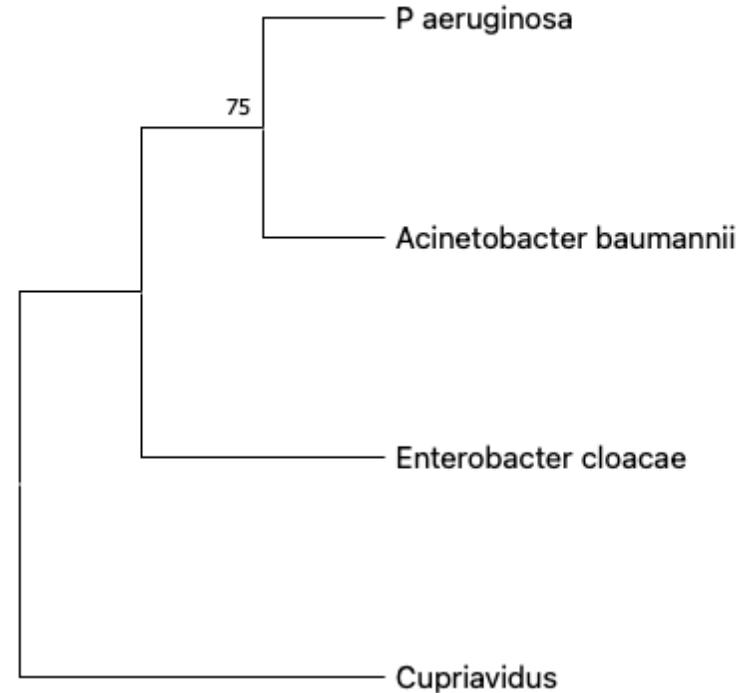
# Minimum Evolution Tree: Bootstrap 500

- Tree created that minimizes amino acid changes
- Good option for only 4 taxa
- Bootstrap score > 70 indicates tree is well supported



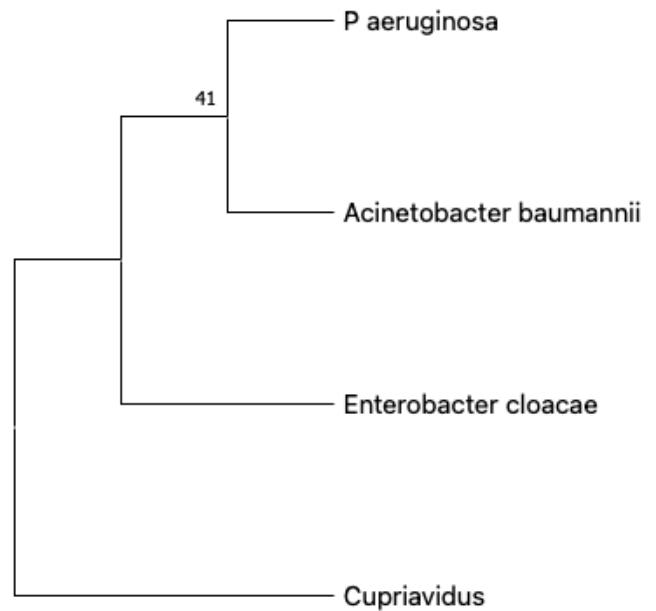
# Neighbor Joining Tree: Bootstrap = 500

- Based on pairwise distances
- Can help choose between differing trees
- Bootstrap score > 70 indicates tree is statistically well supported

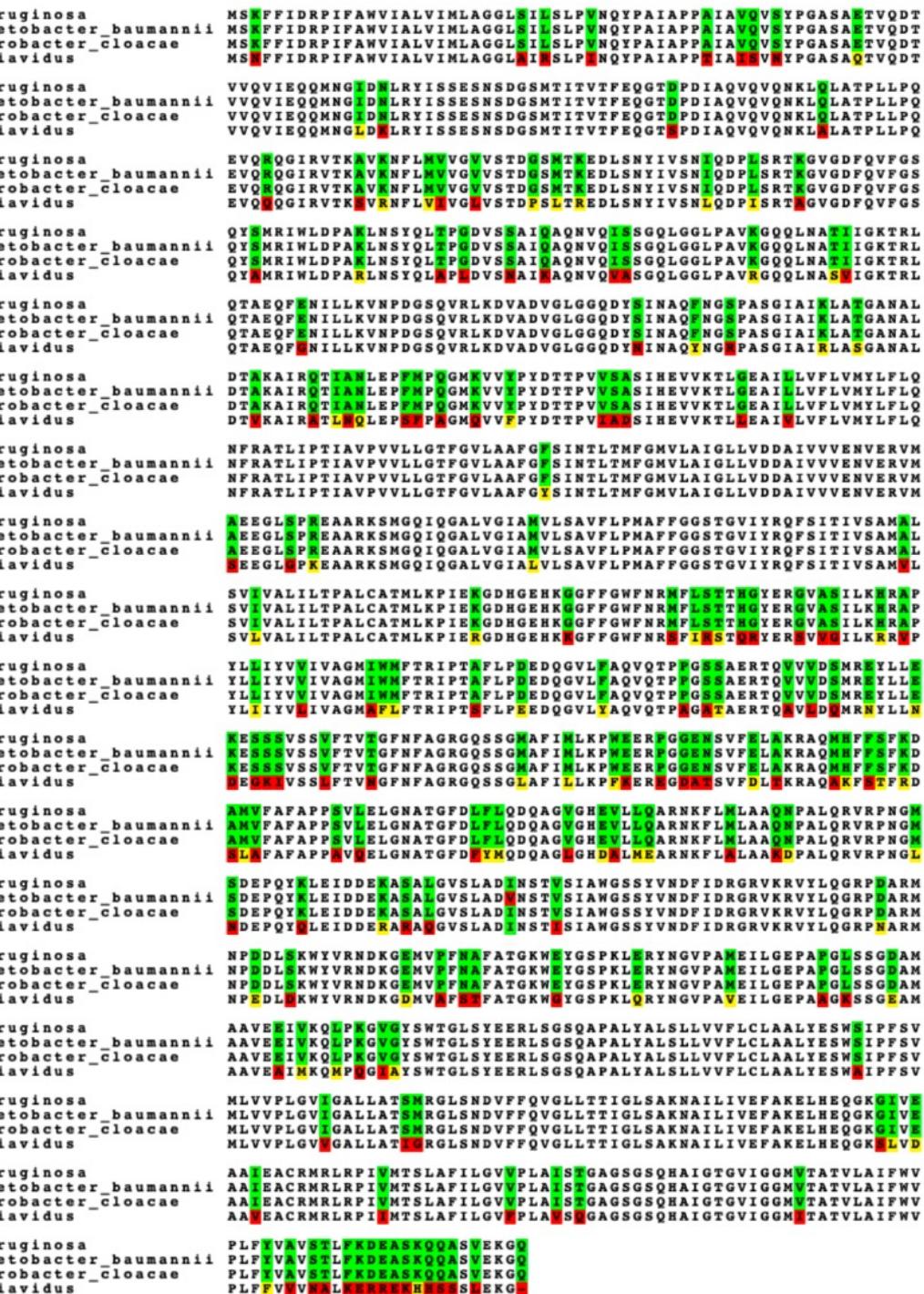


# Maximum Likelihood Tree: Bootstrap = 500

- Considers all available data
- Computationally intensive, but worked for 4 taxa
- Bootstrap score < 50 indicates tree is unresolved



*Cupriavidus* was only one  
on boxshade that ever did  
not match the consensus



# Results

- Multidrug resistance protein MexB is highly similar between
  - *Pseudomonas aeruginosa*
  - *Acinetobacter baumannii*
  - *Enterobacter cloacae*
- MexB is present with slight differences in amino acid sequence in *Cupriavidus*
- The gene mexB is conserved across all 4 species
  - Differences in nucleotide sequence -> same protein product
  - Differences could be helpful to decipher evolutionary relationships
- The protein must have functional importance in resistance