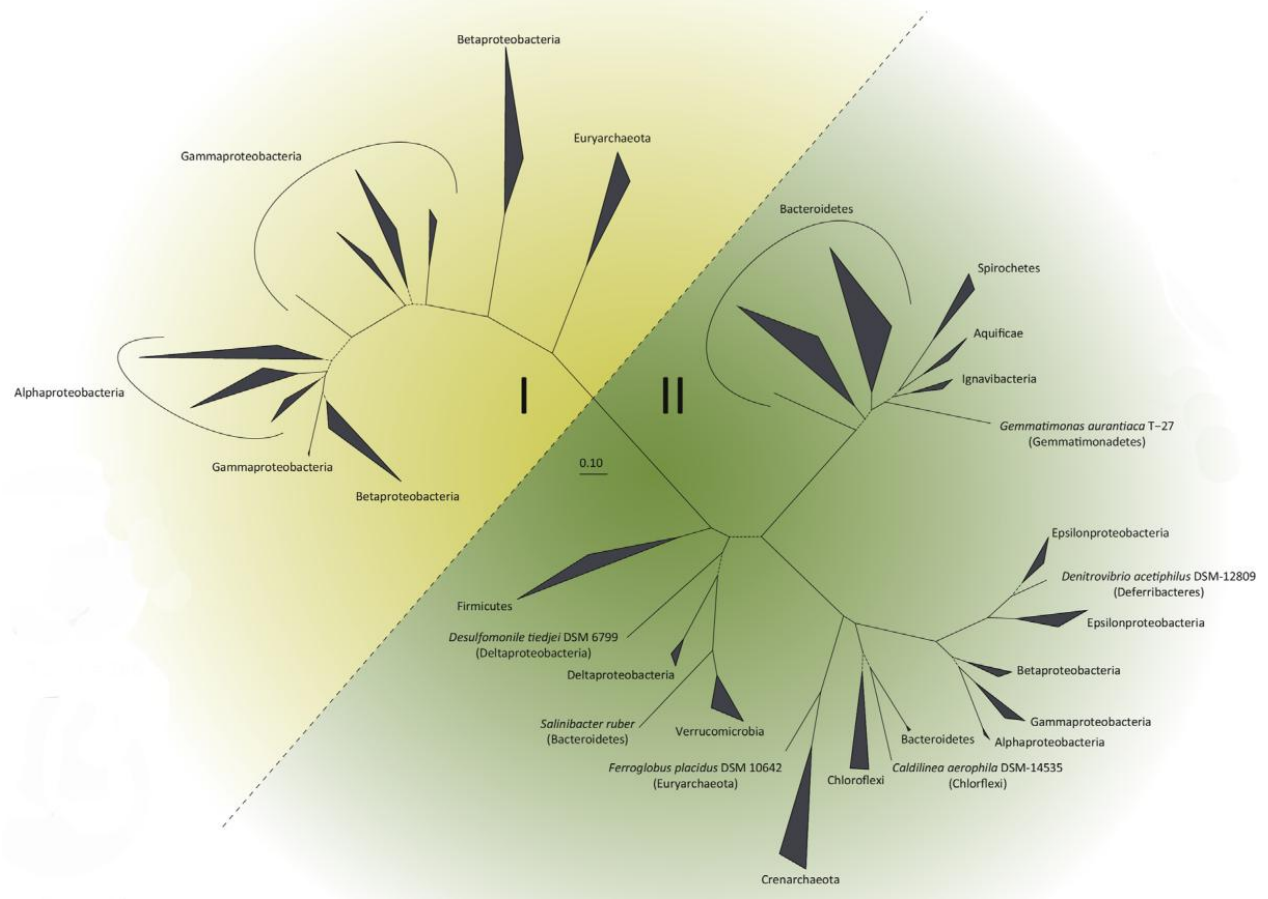


Análisis filogenético de la óxido nitroso reductasa (*nosZ*): establecimiento del perfil de la proteína

Estudiante: David Lázaro Gimeno

Consultor/a; Paloma María Pizarro Tobías

Profesor de la asignatura: Antoni Pérez Navarro



Hallin *et al.* 2018

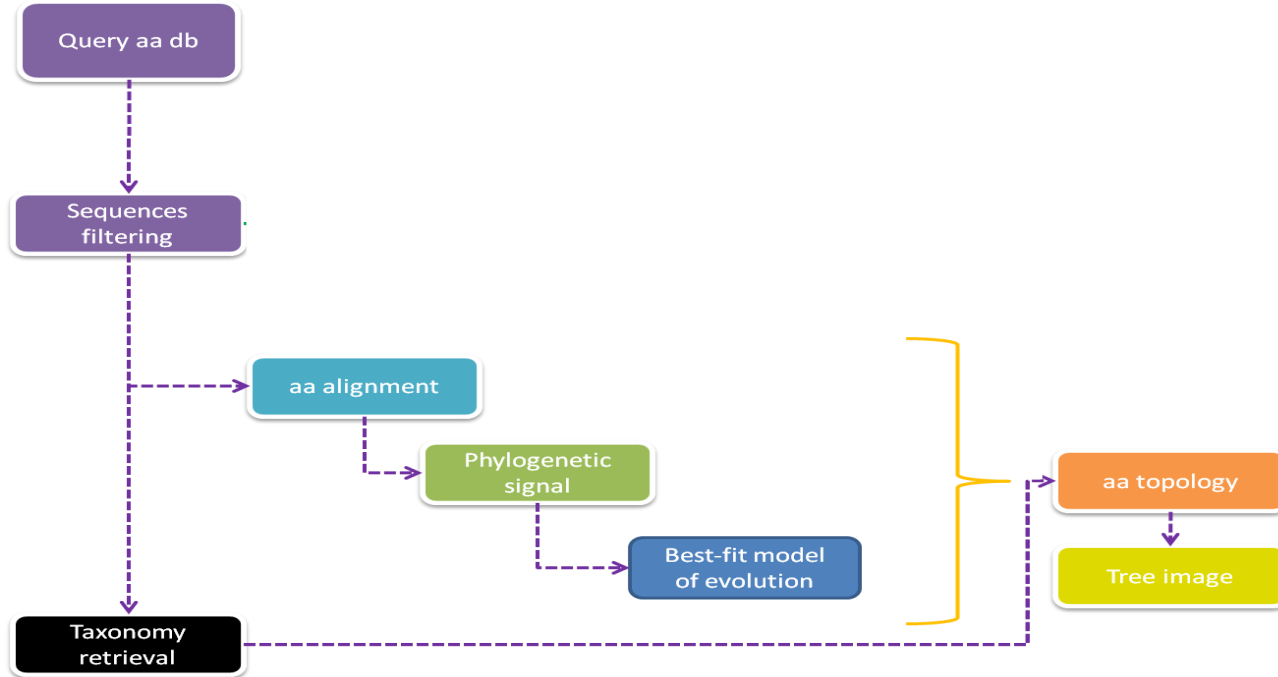
Objetivos generales

- 1) Realizar un **análisis filogenético** de las secuencias *nosZ* mediante **máxima verosimilitud** (ML) para secuencias de aminoácidos e **inferencia bayesiana** (Bayesian) en secuencias de nucleótidos.
- 2) Establecer **el perfil de la enzima** óxido nítrico reductasa.
- 3) **Programación de scripts** que automaticen partes de proceso.

Objetivos específicos

- 1) Descarga y filtrado de las secuencias correspondientes al gen *nosZ*.
- 2) Identificación de secuencias y filtrado de información redundante, así como información que suponga contaminación o ruido.
- 3) Alineamiento de las secuencias mediante algoritmos específicos para nucleótidos o aminoácidos.
- 4) Evaluación de la existencia de señal filogenética.
- 5) Realización de la filogenia mediante métodos de ML y Bayesian.
- 6) Evaluación de las topologías obtenidas.
- 7) Establecimiento del perfil de la proteína.
- 8) Programación de script que automatice algunos de los pasos más significativos del proceso.

Workflow



Obtención secuencias aminoácidos

```
$ esearch -db protein -query "nosZ" | efetch -format fasta > nosz_prot.fasta
```

44.607

¿Qué nos sobra?
Utilicemos herramientas del terminal

```
44600 >CAF74886.1 nitrous-oxide reductase, partial [Pseudomonas stutzeri]
44601 >CAG26676.1 nitrous oxide reductase [Wolinella succinogenes]
44602 >BAC55278.1 nitrous oxide reductase, partial [Marinobacter sp. HS9]
44603 >BAC55277.1 nitrous oxide reductase, partial [Marinobacter sp. HB7]
44604 >BAC55276.1 nitrous oxide reductase, partial [Marinobacter sp. HS7]
44605 >BAC00875.1 NosD [Pseudomonas sp. MT-1]
44606 >BAC00874.1 nitrous oxide reductase [Pseudomonas sp. MT-1]
44607 >BAC00873.1 NosR [Pseudomonas sp. MT-1]
```

```
$ grep "oxide reductase" nosz_prot.fasta > lista1.txt
```

```
$ grep -Eiv "sp.|uncultured|unclassified| culture| \[bacterium |proteobacteria| endosymbiont |
symbiont|endosymbiont |\proteobacterium|division|gamma|raceae|bacterium]| bacterium |
partial | nosD| nosY |norL |nosR " lista1.txt > lista2.txt
```

26.194

4.583

```
$ gawk -F ' ' '{print $1}' lista2.txt > lista3_1.txt
```

```
$ gawk -F '[]' ' '{print $2}' lista2.txt > lista3_2.txt
```

```
paste -d ' ' lista3_1.txt lista3_2.txt | sort -u -k2,3 > lista3.txt
```

1.105

lista4.txt

+ revisión y limpieza manual
última

1.103

```
$ gawk -F ' ' '{print $2}' lista4.txt | sort -u > lista5.txt
```

451

Número de géneros únicos

Aminoácidos

```
>CUJ97397.1 [Achromobacter agrifaciens]
>WP_135527249.1 [Achromobacter agilis]
>WP_175138605.1 [Achromobacter anhimicus]
>WP_008166018.1 [Achromobacter arsenitoxidans]
>MAD09157.1 [Achromobacter cycloclastes]
>WP_180402934.1 [Achromobacter deleyi]
>GFN28924.1 [Achromobacter denitrificans]
>WP_191579308.1 [Achromobacter insolitus]
>CAB3688965.1 [Achromobacter insuavis]
>CAB3845145.1 [Achromobacter mucicola]

>QRD42730.1 [TAT-dependent nitrous-oxide reductase [Sulfitobacter mediterraneus]
MSAEETRRKGFRRGVLTAGAAAAGAGLASFVPSGAKPAAAGASKVNLPAGLDPYGFNWSQGTGE
>WP_202683945.1 [TAT-dependent nitrous-oxide reductase [Skermanella mucosa]
MRDENEKPTIDRRHLGGTAKVAALAGLGTALGGAAVTTAGGSFSGTALAASGTTKQSAEVKPGDLD
>WP_202685600.1 [TAT-dependent nitrous-oxide reductase [Skermanella mucosa]
MSDETRKVEGASISDRRELLGRTARLALAGLTGAGGGAAATLVGSLAPTPVAAEGGGATSHLEPG
>WP_202683945.1 [TAT-dependent nitrous-oxide reductase [Skermanella mucosa]
MRDENEKPTIDRRHLGGTAKVAALAGLGTALGGAAVTTAGGSFSGTALAASGTTKQSAEVKPGDLD
>WP_202661524.1 [TAT-dependent nitrous-oxide reductase [Fabriziicola sp. KV823]
MTIQEFPKPGISRRMLLGSAGTAAIAGTLGGRAAGLALGASPASAAAEAGSFEVAPGQLDEYYV
MTIQEFPKPGISRRMLLGSAGTAAIAGTLGGRAAGLALGASPASAAAEAGSFEVAPGQLDEYYV
```

1.103

44.607

fastafetcher.py

python fastafetcher.py -f nosz_prot.fasta -k filtro.txt -o limpio.fasta

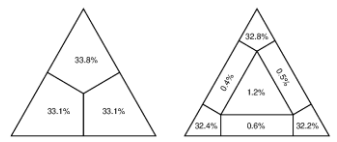
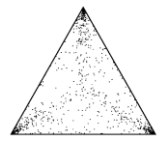
```
>QRD42730.1 [TAT-dependent nitrous-oxide reductase [Sulfitobacter mediterraneus]
MSAEETRRKGFRRGVLTAGAAAAGAGLASFVPSGAKPAAAGASKVNLPAGLDPY
>WP_202685600.1 [TAT-dependent nitrous-oxide reductase [Skermanella mucosa]
MRDENEKPTIDRRHLGGTAKVAALAGLGTALGGAAVTTAGGSFSGTALAASGTTKQSAEVKPGDLD
>WP_202594399.1 [Sec-dependent nitrous-oxide reductase [Pedobacter glucosilyticus]
MKFNCLVVAAGITVVTAMGFGCKPCKVQSEADGAAEKVYVAPGKDELYMFASGDFSG
```

1.103

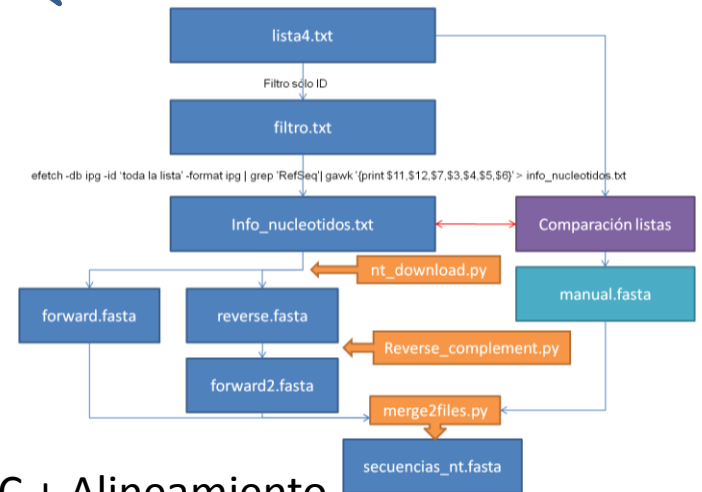
QC + Alineamiento

Protest

AIC → BIC → WAG



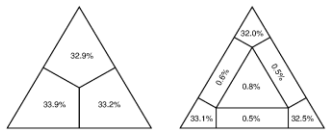
Nucleótidos



QC + Alineamiento

jModeltest

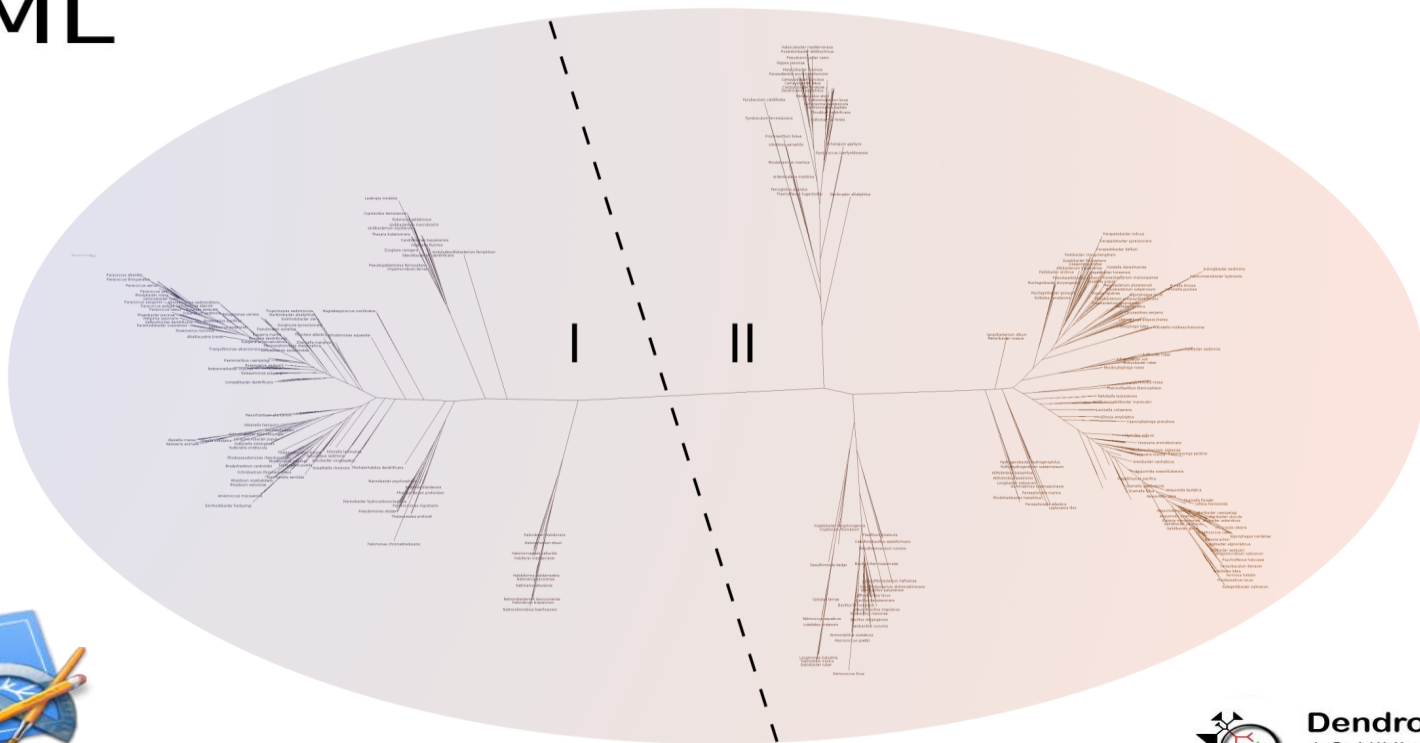
AIC → BIC → GTR+I+G



275 secuencias

Topología aminoácidos

RAxML



Dendroscope 3

by Daniel H. Huson
with contributions from Benjamin Albrecht,
Philippe Gambette, Leo van Iersel,
Celine Scornavacca and others.

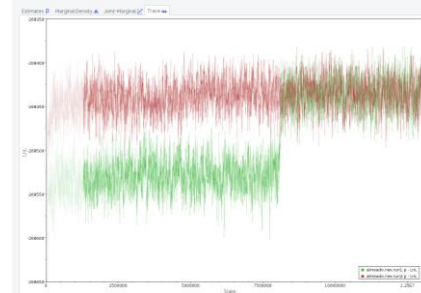
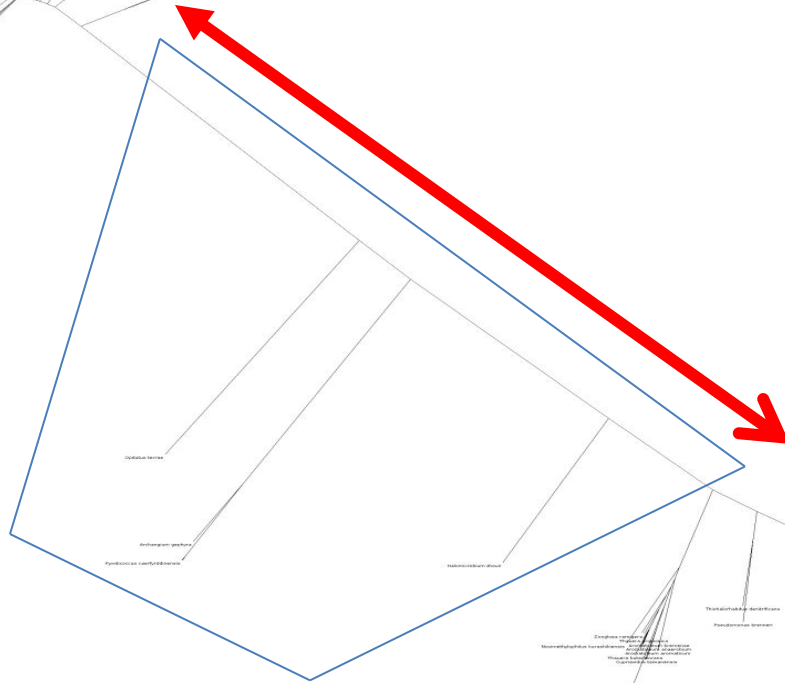
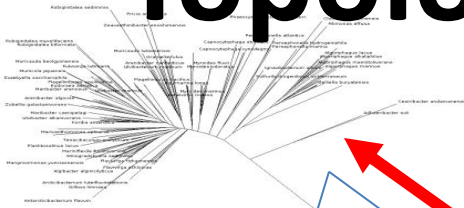
www-ab.informatik.uni-tuebingen.de/software/dendroscope

Topología nucleótidos

→

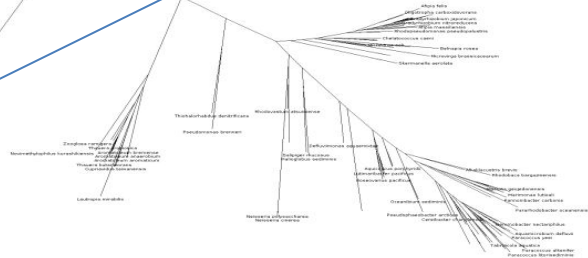
||

LBA

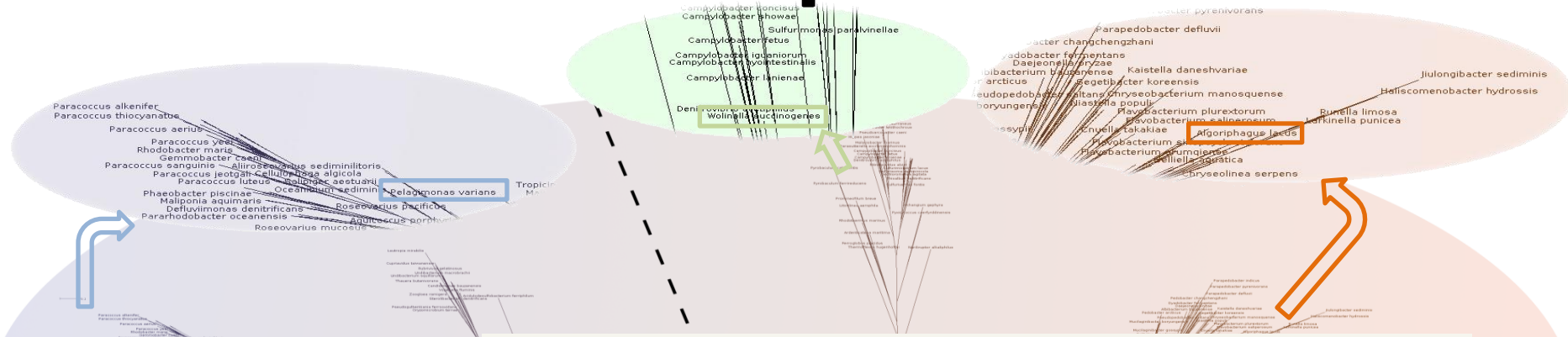


MrBayes

I



Perfil de la proteína



Phyre²

Top model

Model (left) based on template [c2iwbB](#)

Top template information

PDB header: oxidoreductase
Chain: B; **PDB Molecule:** nitrous oxide reductase;
PDB title: inhibitor-bound form of nitrous oxide reductase from *Achromobacter cycloclastes* at 1.7 angstrom resolution

Confidence and coverage

Confidence: 100.0% **Coverage:** 92%

589 residues (92% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

[3D viewing](#)
[Interactive 3D view in JSmol](#)
 For other options to view your downloaded structure offline see the [FAQ](#)

Image coloured by rainbow N → C terminus

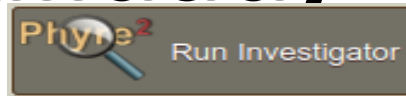
Model dimensions (Å): X:65.381 Y:93.702 Z:59.800

Perfil de la proteína (calidad)

Phyre²



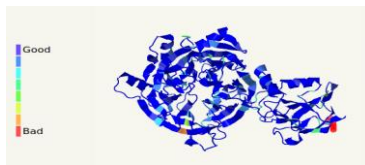
Nitro oxido reductasa
Resolución : 1,7 Å



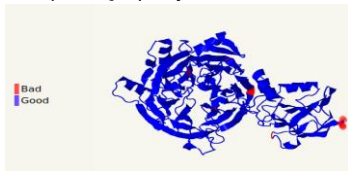
Pelagimona varians (Clado I)



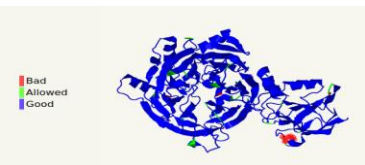
a) ProQ2 quality assesment



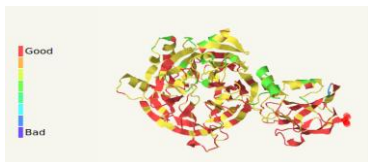
b) Clashes



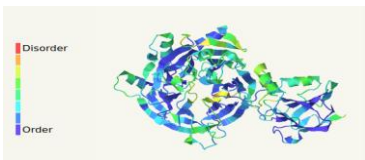
c) Rotamers



d) Ramachandran analysis

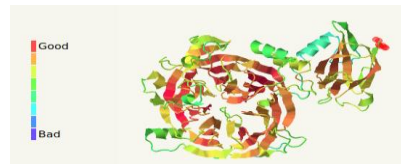


e) Alignment confidence

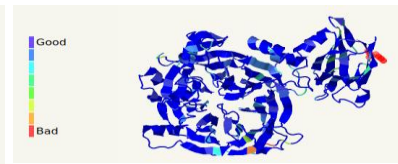


f) Disorder

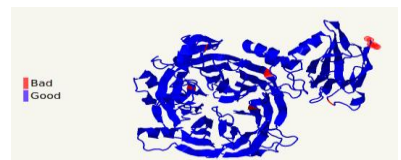
Algoriphagus lacus (Clado II)



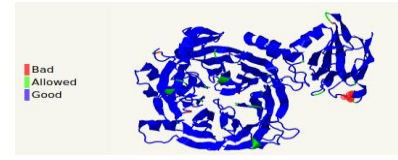
a) ProQ2 quality assesment



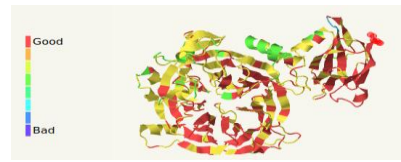
b) Clashes



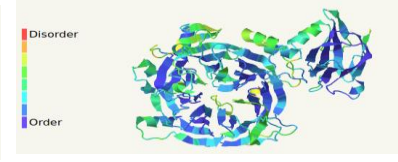
c) Rotamers



d) Ramachandran analysis



e) Alignment confidence



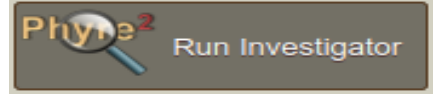
f) Disorder

Perfil de la proteína (función)

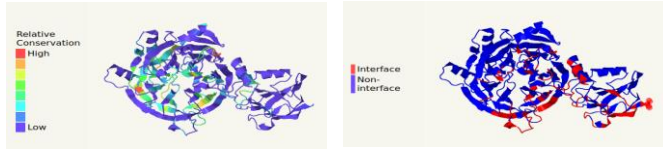
Phyre²



Nitro oxido reductasa
Resolución : 1,7 Å

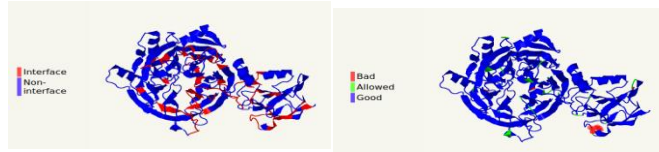


Pelagimona varians (Clado I)



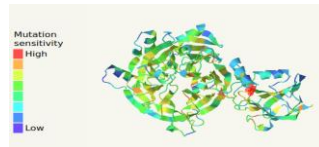
a) Conservation

b) ProtinDB Interface



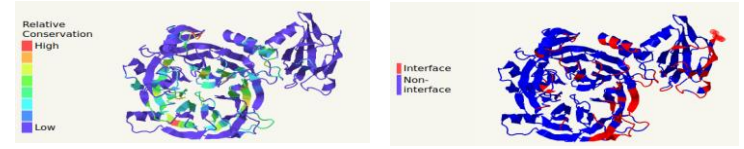
c) PI-Site Interface

d) Pocket detection



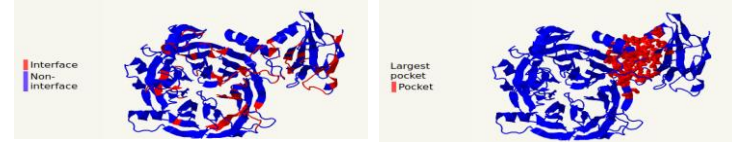
e) Mutational sensitivity

Algoriphagus lacus (Clado II)



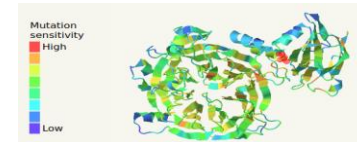
a) Conservation

b) ProtinDB Interface



c) PI-Site Interface




d) Pocket detection











e) Mutational sensitivity

Conclusiones

Objetivos generales

-  1) Realizar un **análisis filogenético** de las secuencias *nosZ* mediante **máxima verosimilitud** (ML) para secuencias de aminoácidos e **inferencia bayesiana** (Bayesian) en secuencias de nucleótidos.
-  2) Establecer **el perfil de la enzima** óxido nítrico reductasa.
-  3) Programación de **scripts** que automaticen partes de proceso.

Objetivos específicos

-  1) Descarga y filtrado de las secuencias correspondientes al gen *nosZ*.
-  2) Identificación de secuencias y filtrado de información redundante, así como información que suponga contaminación o ruido.
-  3) Alineamiento de las secuencias mediante algoritmos específicos para nucleótidos o aminoácidos.
-  4) Evaluación de la existencia de señal filogenética.
-  5) Realización de la filogenia mediante métodos de ML y Bayesian.
-  6) Evaluación de las topologías obtenidas.
-  7) Establecimiento del perfil de la proteína.
-  8) Programación de script que automatice algunos de los pasos más significativos del proceso.

FIN

Wow (Kielokaz ID 359)

by KieLoKaz



This song is free to use under specific conditions.

If I Can't Dance It's Not My
Revolution

by Quantum Jazz



This song is free to use under specific conditions.

[Continue reading...](#)