ClustScan

Arhitektura programa

jdiminic@pbf.hr

Janko Diminić

Zahtjevi

- Više komandno-linijskih “*bio alata*” za linux
- Omotač oko tih alata
- Mogući problem: procesorska snaga i količina podataka

Arhitektura

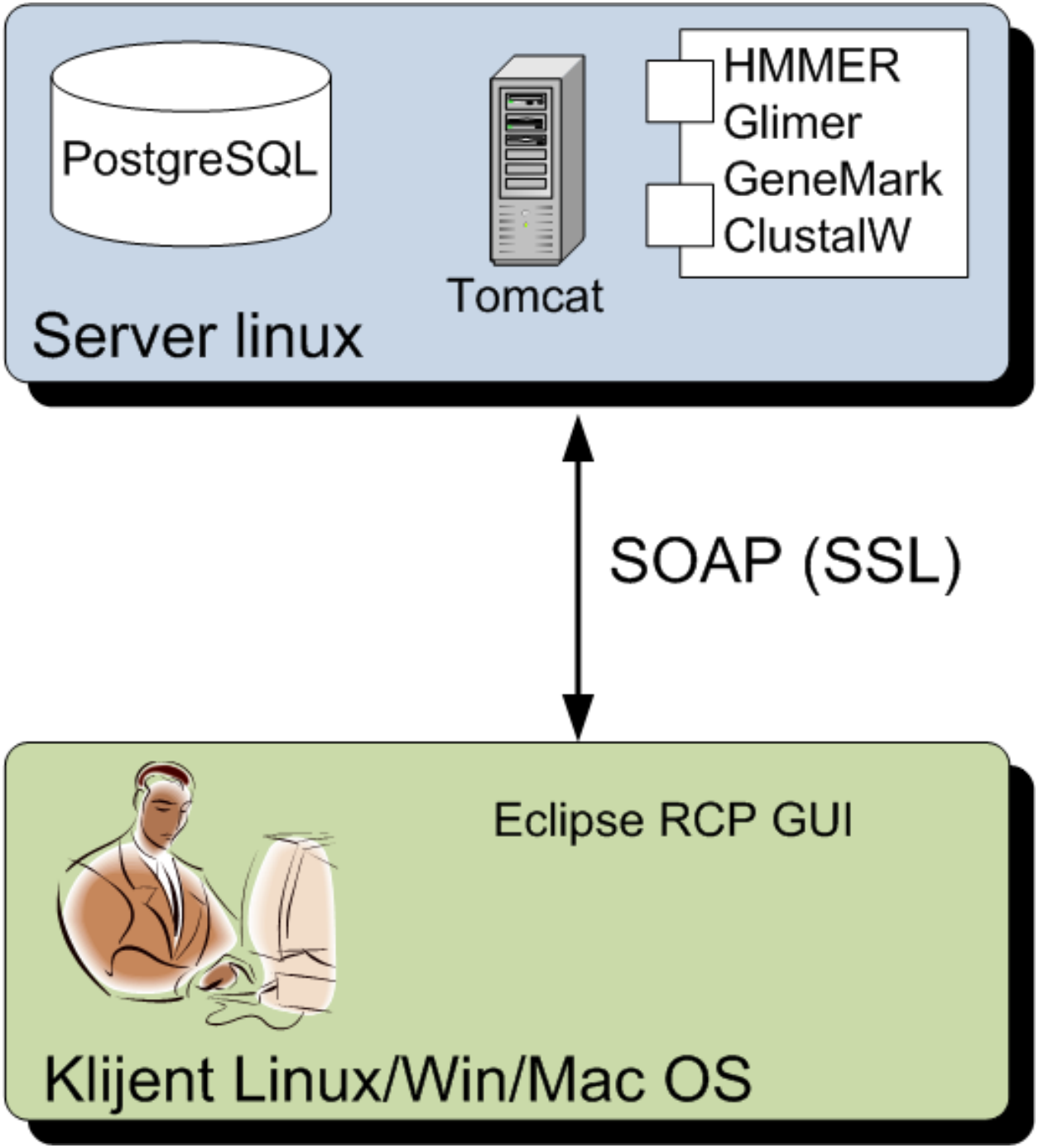
Klijent:

- Java GUI -> svi OS podržani

Server:

- Pokreće "*bio alate*"
- Čuva korisnikove podatke
- Web servis

Komunikacija: SOAP (XFire)


PostgreSQL
Tomcat
HMMER
Glimer
GeneMark
ClustalW
Server linux
SOAP (SSL)
Eclipse RCP GUI
Klijent Linux/Win/Mac OS

Što program radi?

- Unos sekvencije DNA (ReadSeq)
- Workspace
- Prikaz sekvencije DNA u anotacijskom editoru
- Osnovni alati za analizu sekvencije: HMMER, Glimmer, Genemark
- HMMER - izgradnja profila
- Pretraživanje Pfam baze profila

CompGen - from server <janko@bioserv.obf.hr:9445>

File Edit Tools Help

Workspace

- megalomycin
- niddamycin
  - niddamycin pks
  - Forward
  - Reverse
- erythromycin
  - erythromycin pks
  - Forward
    - forward 1 (0)
    - forward 2 (1)
    - forward 3 (0)
    - Gene 1
      - M0
      - M1
        - KS (522-944)
        - AT (1046-1368)
        - KR (1628-1793)
        - ACP (1905-1971)
      - M2
    - Gene 2
    - Gene 3
    - Gene 4
  - Reverse
- oleandomycin patent
  - oleandomycin pks
  - Forward
  - Reverse
- anabaena
  - Forward
  - Reverse
- Recombinants
  - erythromycin pks | niddamycin pks
  - oleandomycin pks | megalomycin pks
  - oleandomycin pks | erythromycin pks
  - niddamycin pks | megalomycin pks
  - oleandomycin pks | niddamycin pks

megalomycin

Annotation editor: megalomycin

```
 R  G  Q  S  R  A  D  D  D  V  G  E  L  G  L  T  G  H

T  E  V  S  P  A  P  T  T  T  S  V  S  S  G  *  L  A  T

 P  R  S  V  P  R  R  R  R  R  R  *  A  R  A  D  W  P  P

 10330     10340     10350     10360     10370     10380
CCGAGGTCAGTCCCGCGCCGACGACGACGTCGGTGAGCTCGGGCTGACTGGCCACCC

 37651     37641     37631     37621     37611     37601
GGCTCCAGTCAGGGCGCGGCTGCTGCTGCAGCCACTCGAGCCCGACTGACCGGTGGG

 S  T  L  G  A  G  V  V  V  D  T  L  E  P  Q  S  A  V

G  L  D  T  G  R  R  R  R  R  R  H  A  R  A  S  Q  G  G

 R  P  *  D  R  A  S  S  S  T  P  S  S  P  S  V  P  W  G
```

Editor | Script

Details | Log | Progress | Console

KR

Domain properties — Show in workspace

DNA coordinates:	4998..5496 (498 pb)
Protein frame:	Forward 1
Protein coordinates:	1628..1793 (165 aa)
Score:	298.075
E-value:	1.86419E-90
Activity:	☑ Active
Specificity:	Chirality of Me: S
	Chirality of OH: R
Description:	Ketoreductase
Note:	

Color

Alignment

```
Profile:   GTVLITGGTGGLGLAVARWLVEEHGARHLVLVSRRGLEREWDTSLGPDAPGAAELRPI
Alignment: GTVL+TGGTGG G+++ARWL+++ GA HL+LVSR+          GPDA+GA+EL
Hit:       GTVLVTGGTGGVGGQIARWLARR-GAPHLLLVSRS---------GPDADGAGEL---
```

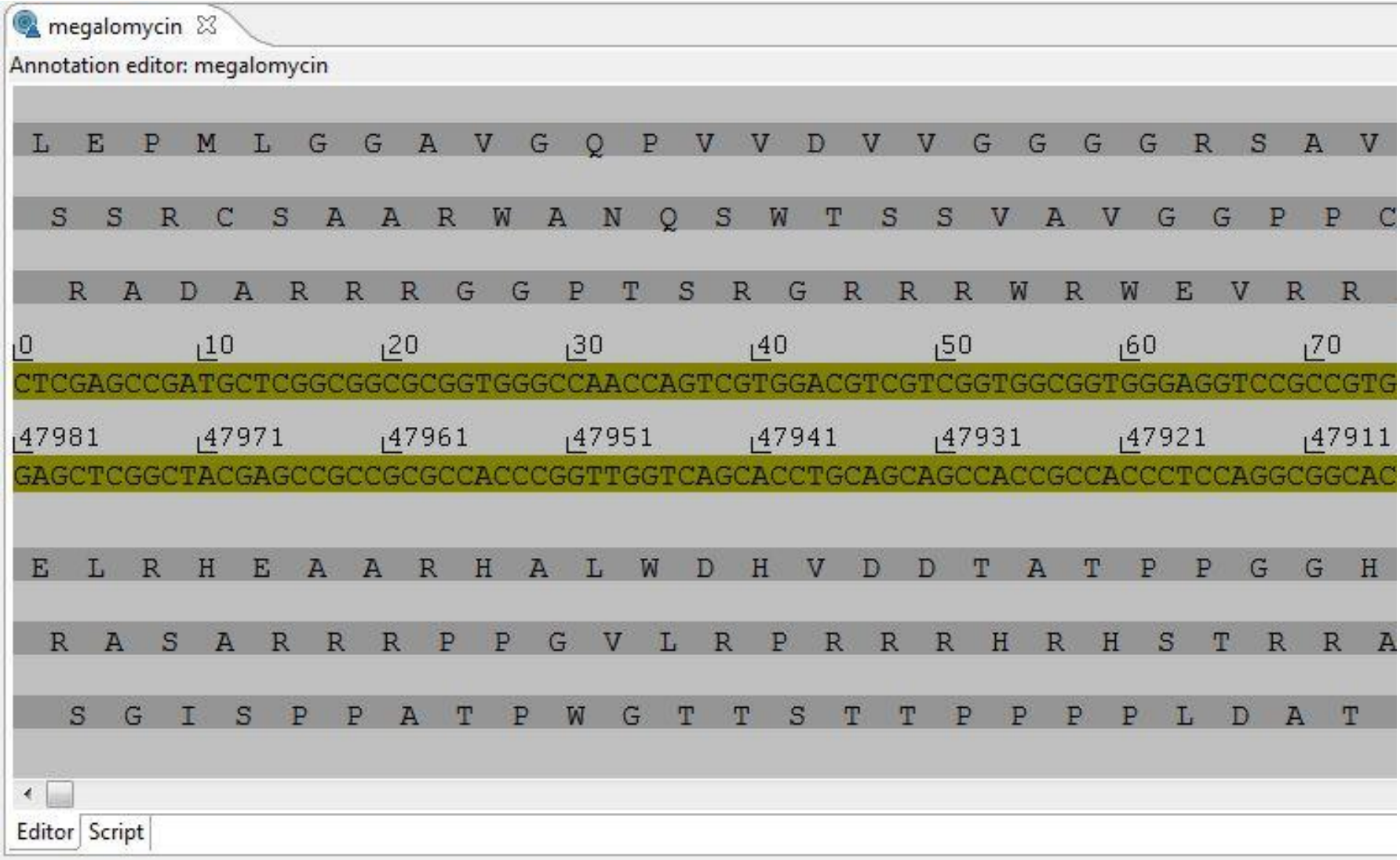
Serverski dio - tehnologije

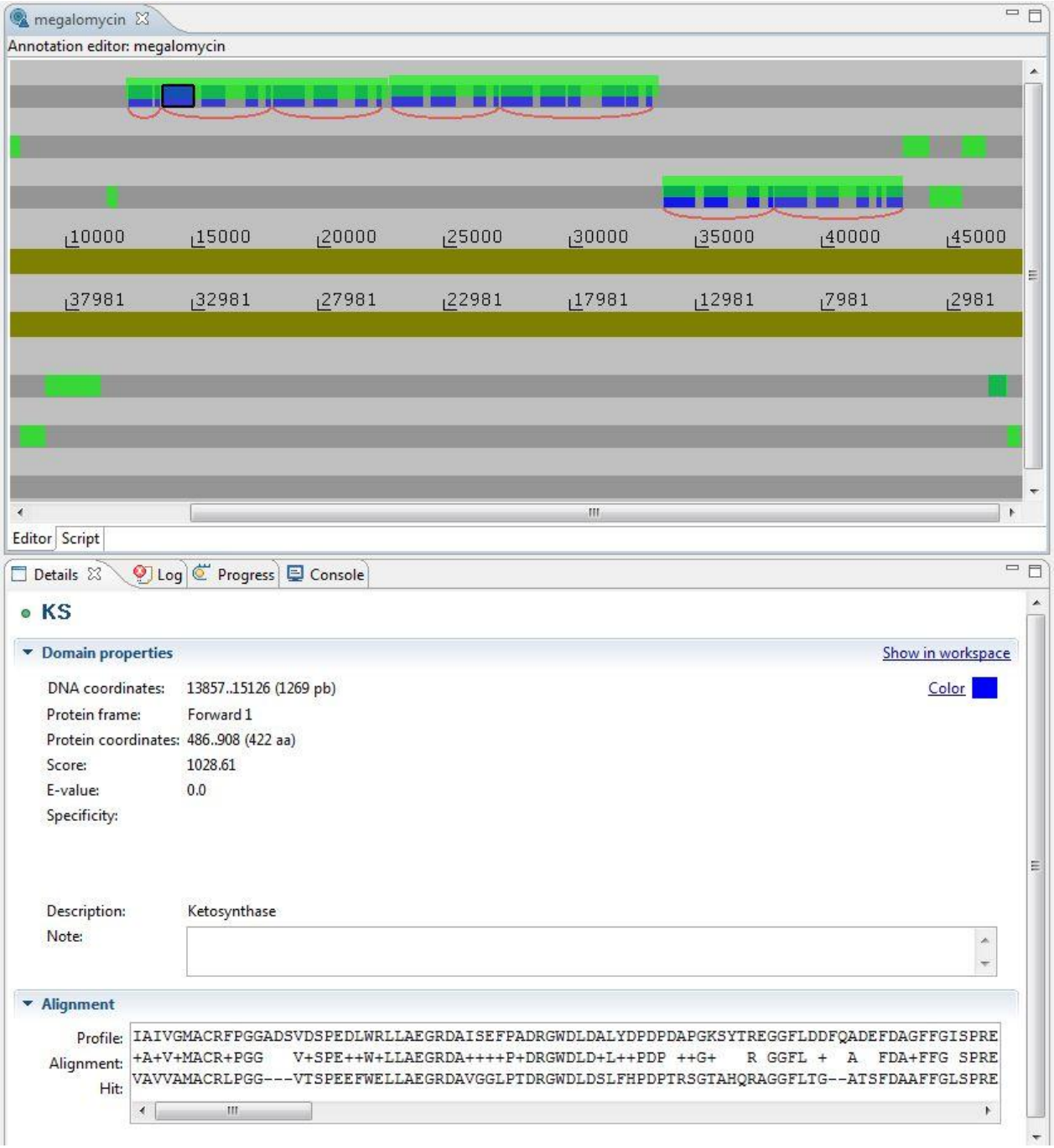
- Open source tehnologije
- PostgreSQL
- Spring
- Hibernate
- XFire (SOAP)
- Tomcat
- Java, Perl (Bioperl), Bash skripte, PHP

RCP - eclipse

- Opsežan framework
- SWT GUI
- Plugin
  - Update
  - Editor - views
  - Help
- Ekstenzije postojećih plugin-ova

Anotacijski editor




megalomycin
Annotation editor: megalomycin
10000
15000
20000
25000
30000
35000
40000
45000
37981
32981
27981
22981
17981
12981
7981
2981
Editor
Script
Details
Log
Progress
Console
KS
Domain properties
Show in workspace
DNA coordinates: 13857..15126 (1269 pb)
Color
Protein frame: Forward 1
Protein coordinates: 486..908 (422 aa)
Score: 1028.61
E-value: 0.0
Specificity:
Description: Ketosynthase
Note:
Alignment
Profile: IAIVGMACRFPGGADSVDSPEDLWRLLAEGRDAISEFPADRGWDLDALYDPDPDAPGKSYTREGGFLDDFQADEFDAGFFGISPRE
Alignment: +A+V+MACR+PGG V+SPE++W+LLAEGRDA++++P+DRGWDLD+L++PDP ++G+ R GGFL + A FDA+FFG SPRE
Hit: VAVVAMACRLPGG---VTSPEEFWELLAEGRDAVGGLPTDRGWDLDSLFHPDPTRSGTAHQRAGGFLTG--ATSFDAAFFGLSPRE

Script editor

- Izrada malih dodataka programu
- Podrška za java 6 skriptne jezike (preko 30 jezika)
- Editor za pisanje skripti: bojenje sintakse i code completion
- Mogućnost spremanja i ponovnog korištenja već napisanih skripti

Script editor

megalomycin

Annotation editor: megalomycin

Language: ECMAScript Script name: GC sastav

Run | New | Save | Load... | Settings | Find

proteins | String getSequence(int from, int to, int numFame)

```
1  //Program za računanje GC sastava DNA sekvencije
2
3  // uzmi forward DNA seqvenciju
4  var dnaSeq = dna.getSequence(1);
5
6  var gc = 0;
7
8  for(var pos = 0; pos < dnaSeq.length(); pos++) {
9      // uzimaj jedan po jedan znak iz sekvencije
10     symbol = dnaSeq.substring(pos, pos+1);
11     if(symbol == 'G' || symbol == 'C') {
12         ++gc;
13     }
14 }
15
16 "GC sastav = "+ ((gc * 100) / dnaSeq.length()) + " %"
```

Ln 15 Col 55 (391)

Editor | Script

Details | Log | Progress | Console

toolbar: dnaSeq

```
GC sastav = 72.77880827827681 %
```

Code completion

megalomycin

Annotation editor: megalomycin

Language: ECMAScript Script name: dnaSeq

Run | New | Save | Load... | Settings | Find

proteins | String getSequence(int from, int to, int numFame)

```
1  //Program za računanje GC sastava DNA sekvencije
2
3  // uzmi forward DNA seqvenciju
4  var dnaSeq = dna.
5
6  var gc = 0;
7
8  for(var pos = 0; p
9      // uzimaj jedan
10     symbol = dnaS
11     if(symbol == '
12         ++gc;
13     }
14 }
15
16
17 "GC sastav = "+ ((
```

- String getName()
- String getSequence(int strand)
- String getSequencePart(int from, int to, int strand)

objects and methods

Vraca java String objekat koji sadrzi DNA sekvenciju.
Parameters:
strand Prestavlja DNA strand, moze biti -1 or 1

description

Ln 3 Col 17 (102)

Editor | Script

Clustcan database CSDB

Back Forward Refresh Stop Compile/Browse

http://localhost:8888/hr.pbf.web.Index.ClustScanWeb/ClustScanWeb.html

ClustScan database (CSDB)

how to cite it

Clusters

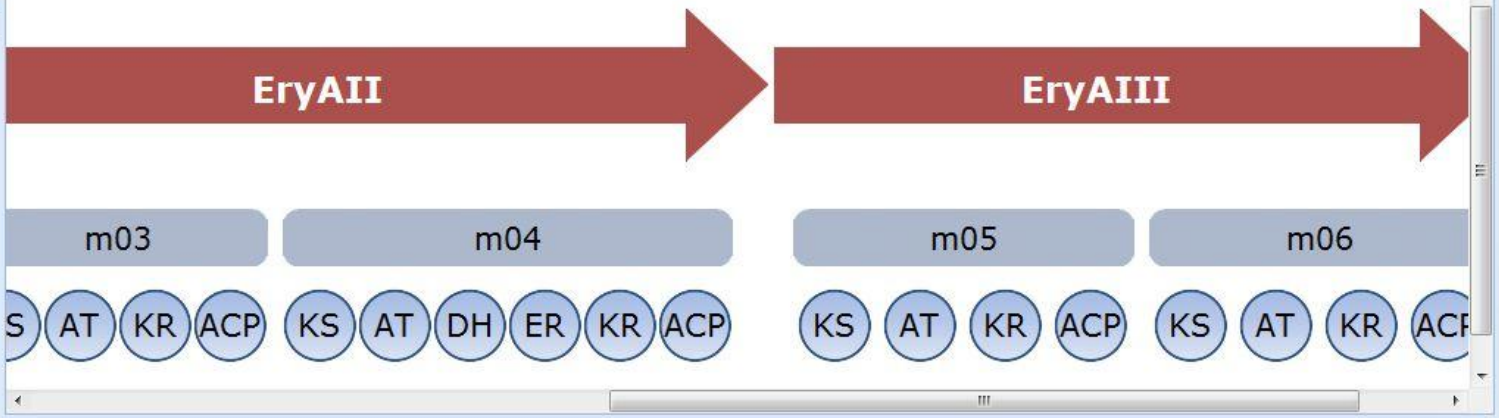
erythromycin

Image Show

- eryAI
  - m00
    - AT
    - ACP
  - m01
    - KS
    - AT
    - KR
    - ACP
  - m02
    - KS
    - AT
    - KR
    - ACP
  - Gene 2
- eryAII
  - m03
    - KS
    - AT
    - KR
    - ACP
  - m04
    - KS
    - AT
    - DH
    - ER
    - KR
    - ACP
- eryAIII
  - m05
    - KS
    - AT
    - KR
    - ACP

Home | erythro...

search



erythromycin

PKS type: Type I modular

Genomes: Saccharopolyspora erythraea NRRL 2338 chromosome

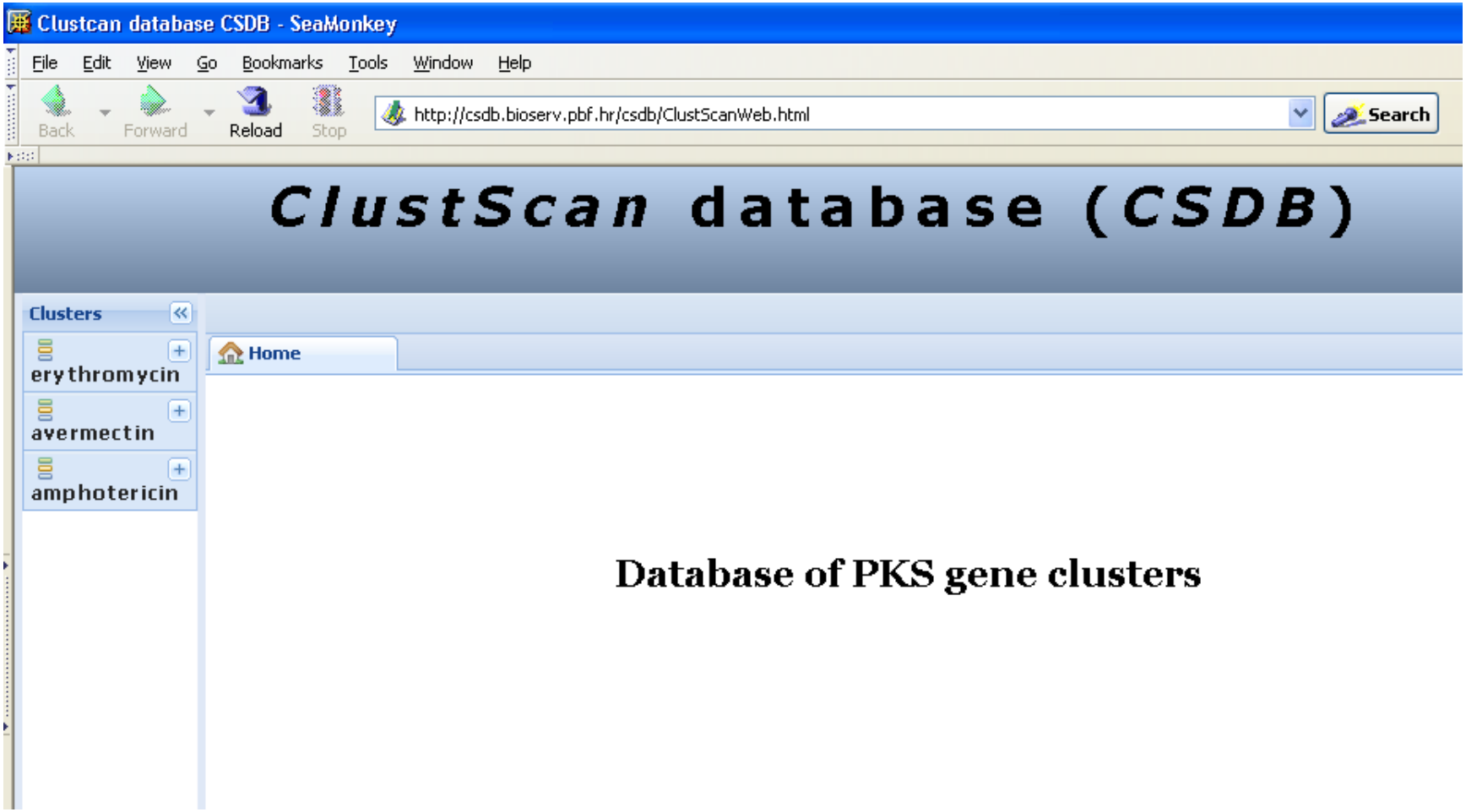
Starter unit: m00

Chain length: 7

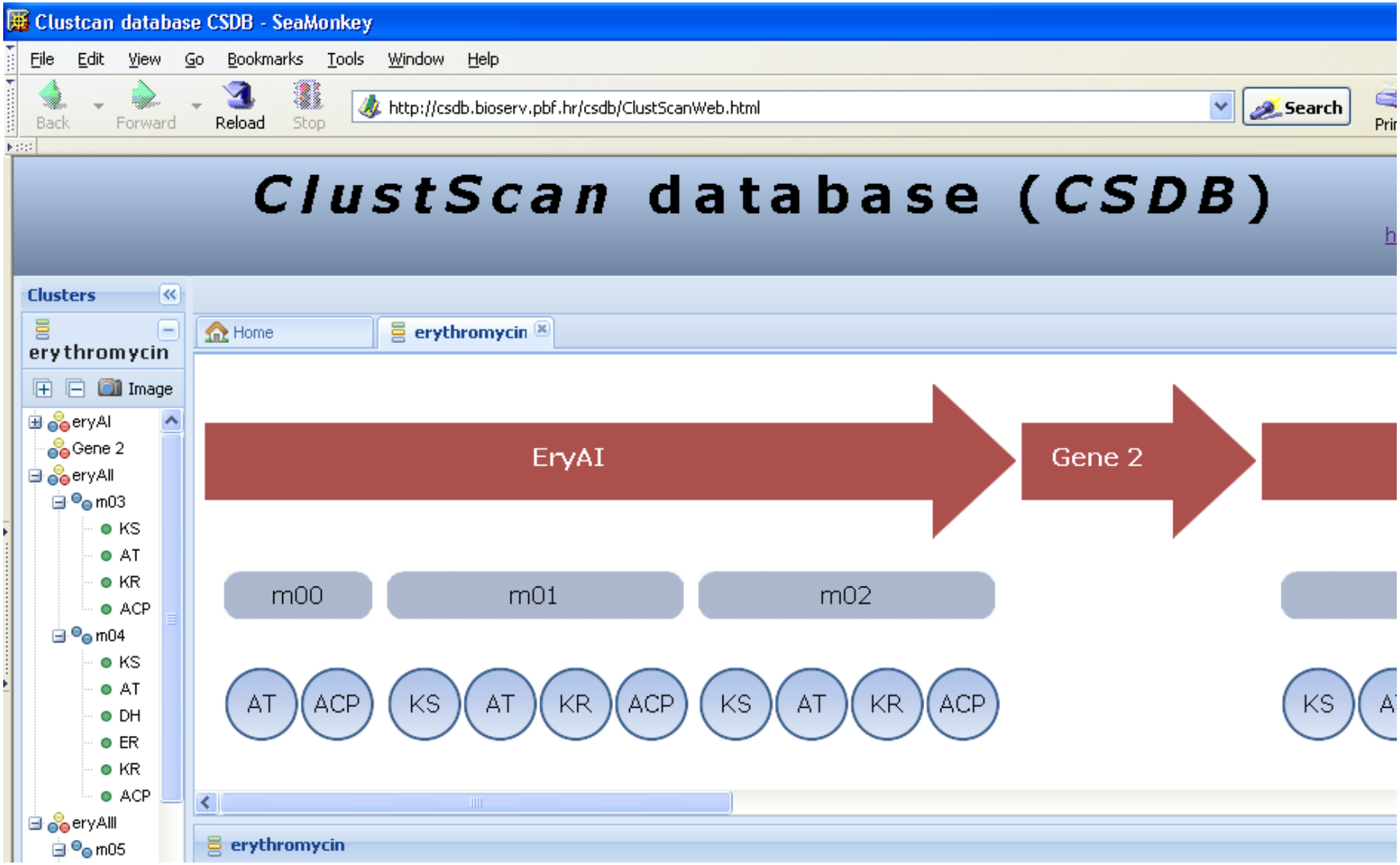
Description: Erythromycin is a macrolide antibiotic that has an antimicrobial spectrum similar to or slightly wider than that of penicillin, and is often used for people that have an allergy to penicillins. For respiratory tract infections, it has better coverage of atypical organisms, including mycoplasma and Legionellosis. It is also used to treat outbreaks of chlamydia, syphilis, acne, and gonorrhea. In structure, this macrocyclic compound contains a 14-membered lactone ring with ten asymmetric centers and two sugars (L-cladinose and D-desoamine), making it a compound very difficult to produce via synthetic methods. Erythromycin is produced from a strain of the actinomycete Saccharopolyspora erythraea, formerly known as Streptomyces erythraeus.

Reference: Haydock SF, Dawson JA, Dhillon N, Roberts GA, Cortes J, Leadlay PF (1991) loning and sequence analysis of genes involved in erythromycin biosynthesis in Saccharopolyspora erythraea. Molecular and General Genetics MGG 230: 120-128.

Linear	Aglycone
Smiles: [C@H](C)[C@@H](O)[C@@H](C)[C@H](O)[C@H](C)C(=O)C(C)C [C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C(=O)S	Smiles: [C@H](C)[C@@H](O1)[C@@H](C)[C@H](O)[C@H](C)C(=O)C(C)C [C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C1(=O)



Database of PKS gene clusters


Clustcan database CSDB - SeaMonkey
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html
Search
ClustScan database (CSDB)
Clusters
erythromycin
Image
eryAI
Gene 2
eryAII
m03
KS
AT
KR
ACP
m04
KS
AT
DH
ER
KR
ACP
eryAIII
m05
Home
erythromycin
EryAI
Gene 2
m00
m01
m02
AT
ACP
KS
AT
KR
ACP
KS
AT
KR
ACP
KS
erythromycin

Clustcan database CSDB - SeaMonkey

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html Search Prin

ClustScan database (CSDB)

Clusters

erythromycin

Image

- eryAI
  - Gene 2
- eryAII
  - m03
    - KS
    - AT
    - KR
    - ACP
  - m04
    - KS
    - AT
    - DH
    - ER
    - KR

Home | erythromycin

erythromycin

PKS type: Type I modular

Genomes: Saccharopolyspora erythraea NRRL 2338 chromosome

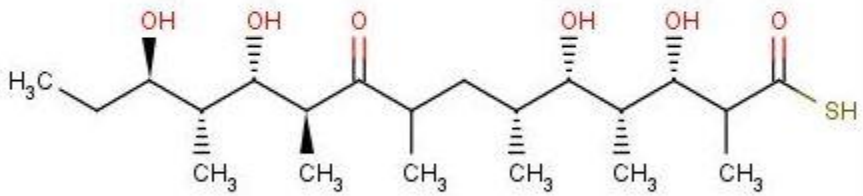
Starter unit: m00

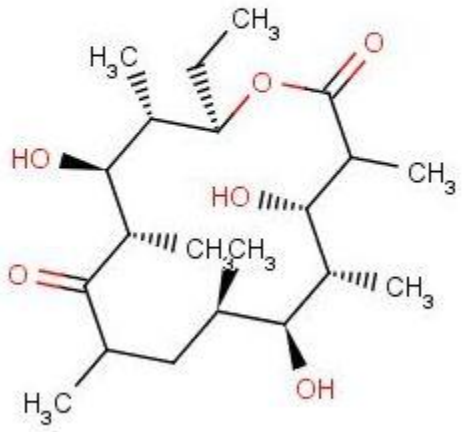
Chain length: 7

Description: Erythromycin is a macrolide antibiotic that has an antimicrobial spectrum similar to or slightly wider than that of penicillin, and is often u people that have an allergy to penicillins. For respiratory tract infections, it has better coverage of atypical organisms, including mycoplasma and Legi It is also used to treat outbreaks of chlamydia, syphilis, acne, and gonorrhea. In structure, this macrocyclic compound contains a 14-membered lacto with ten asymmetric centers and two sugars (L-cladinose and D-desoamine), making it a compound very difficult to produce via synthetic methods. Erythromycin is produced from a strain of the actinomycete Saccharopolyspora erythraea, formerly known as Streptomyces erythraeus.

Reference: Haydock SF, Dawson JA, Dhillon N, Roberts GA, Cortes J, Leadlay PF (1991) loning and sequence analysis of genes involved in erythrom biosynthesis in Saccharopolyspora erythraea. Molecular and General Genetics MGG 230: 120-128.

Linear	Aglycone
SMILES: [C@H](C)[C@@H](O)[C@@H](C)[C@H](O)[C@H](C)C(=O)C(C)C[C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C(=O)S	SMILES: [C@H](C)[C@@H](O1)[C@@H](C)[C@H](O)[C@H](C)C(=O)C(C)C[C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C1(=O)





Pitanja?