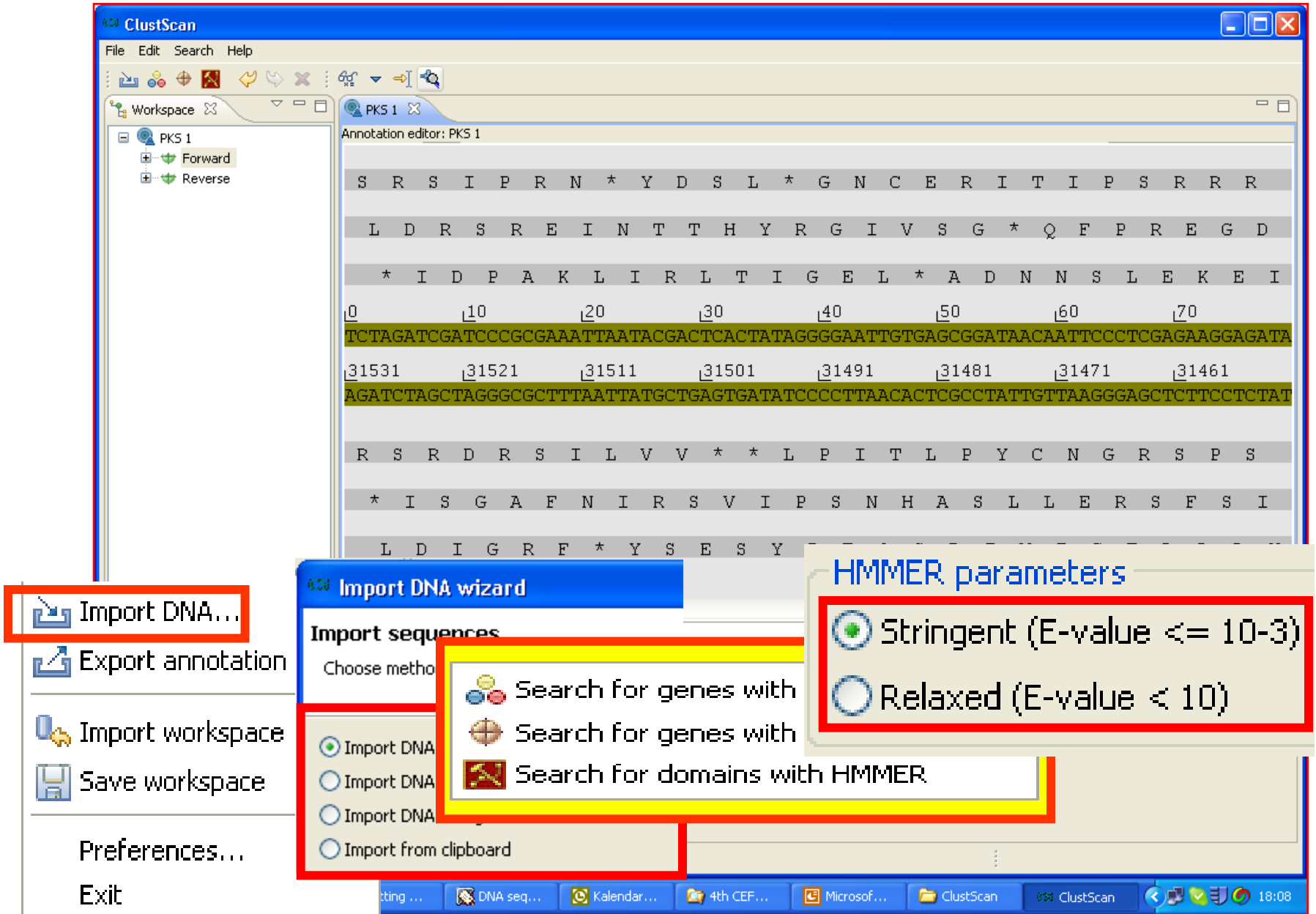
ClustScan and *CompGen* Program Packages

Novalis *and*

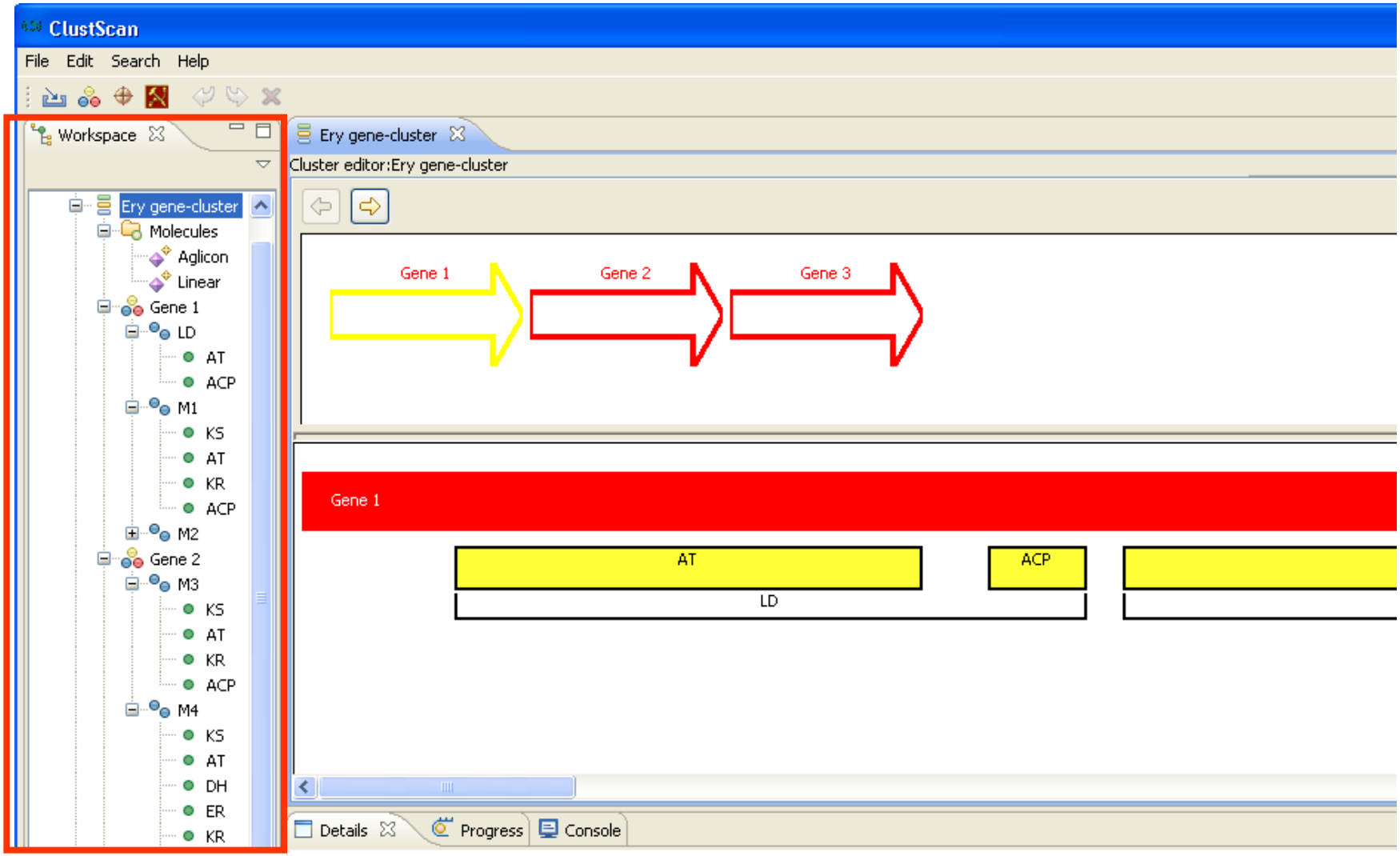
Antonio Starčević

astar@pbf.hr

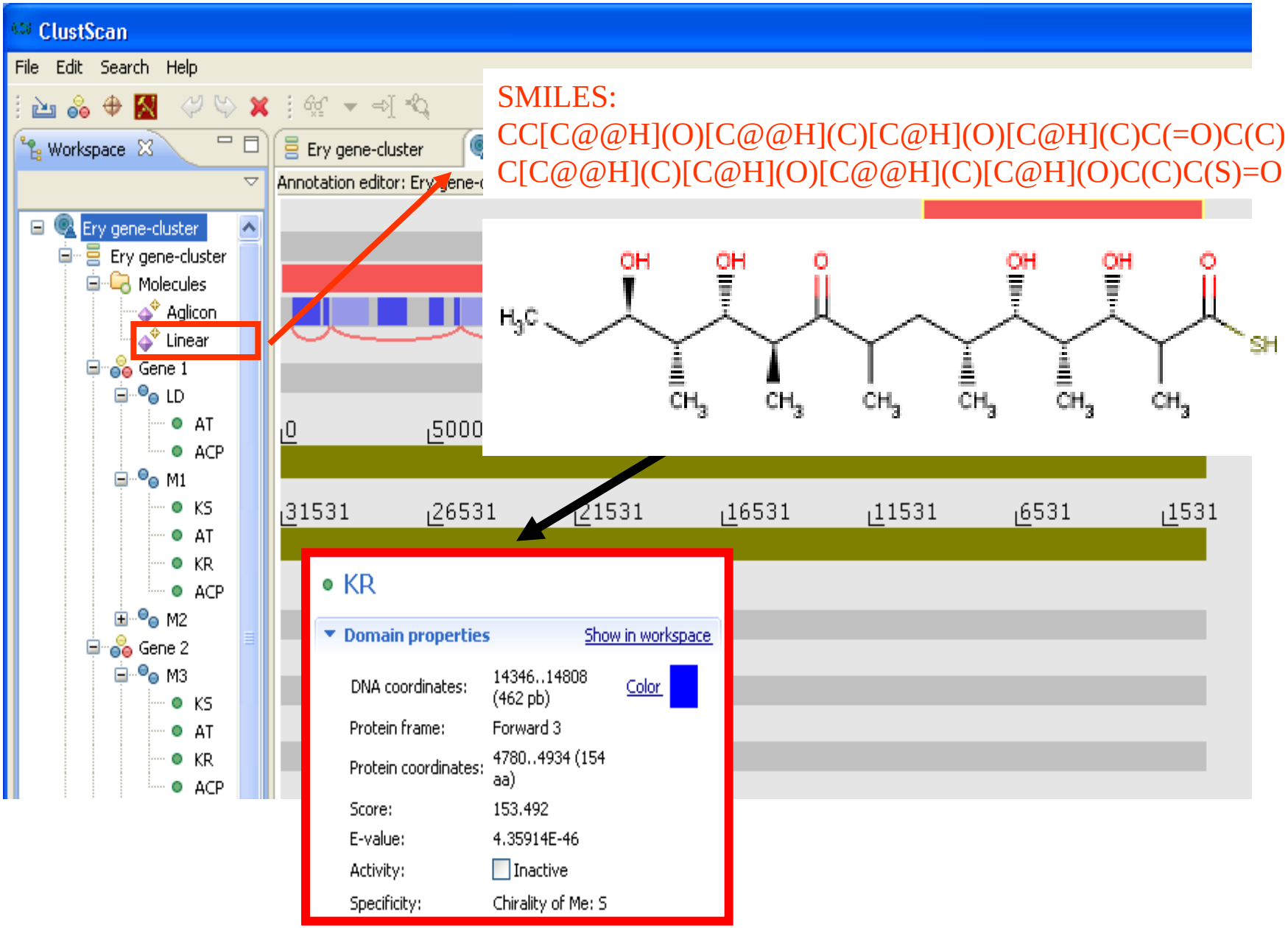
ClustScan program package:



Cluster editor and Workspace window:

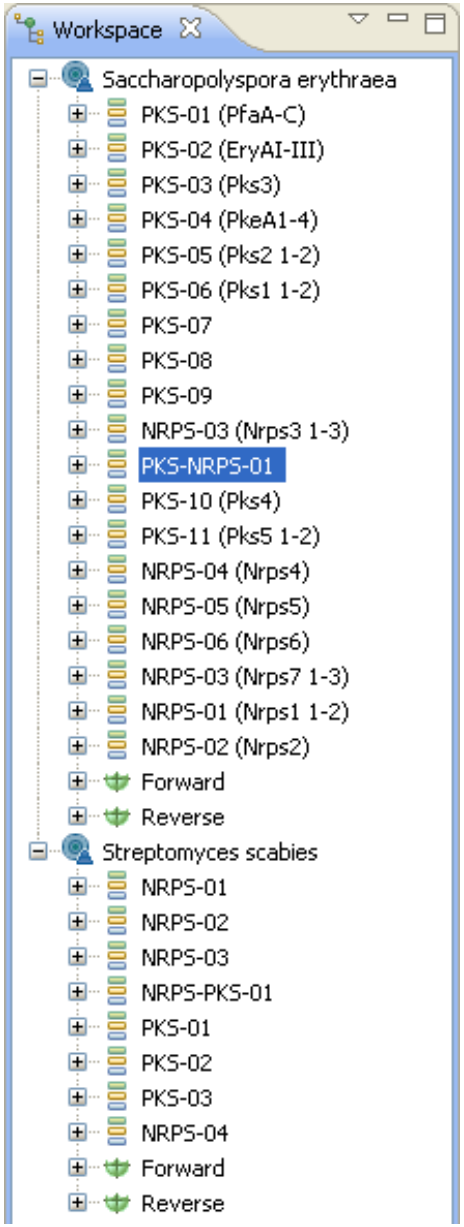


Annotation editor:



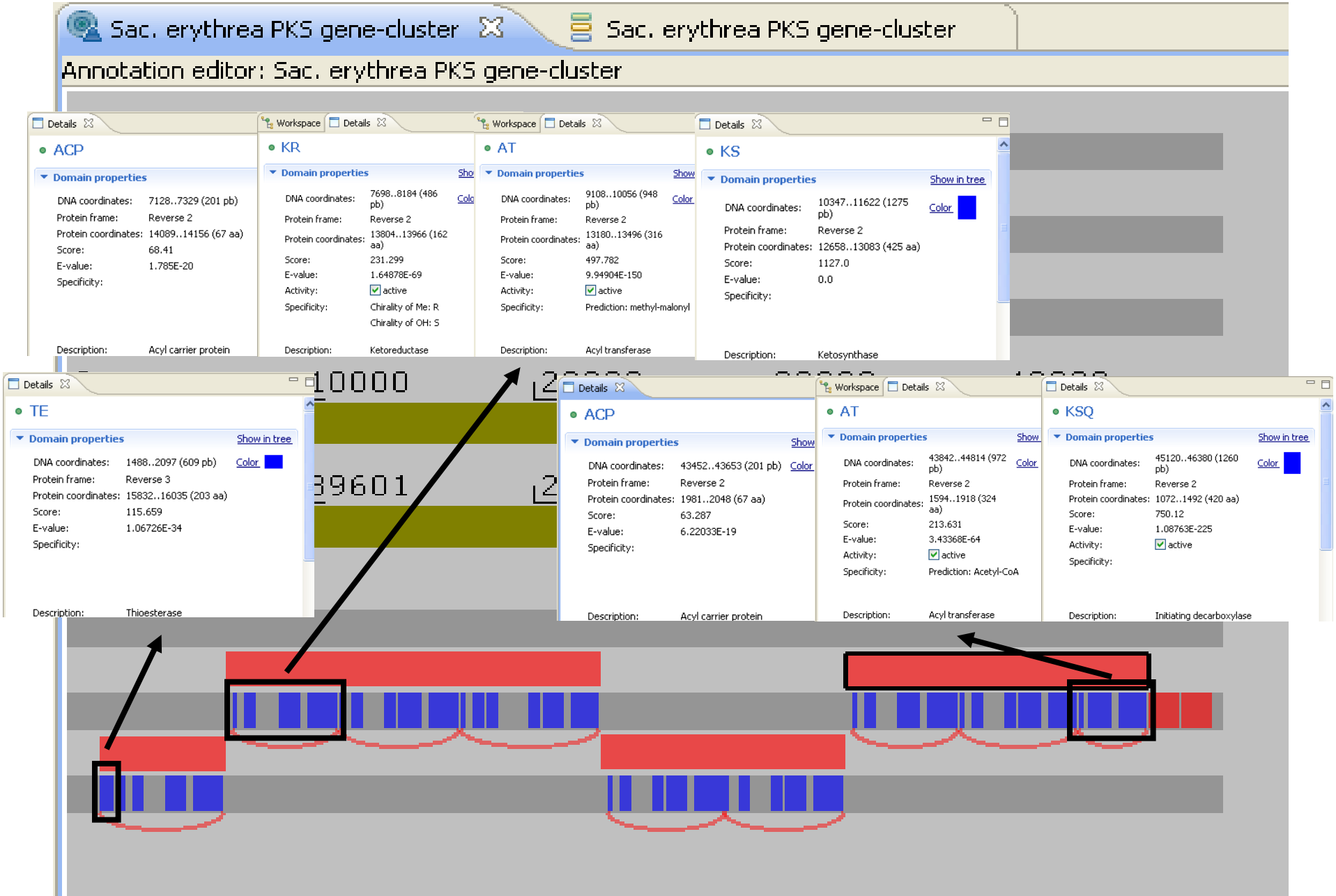
Sequencing of *Sac. erythrea* genome:

2007 - Peter Leadly's group published DNA sequence of the 8.2 Mb *Sac. erythraea* genome (11 PKS, 6 NRPS, 1 PKS-NRPS)



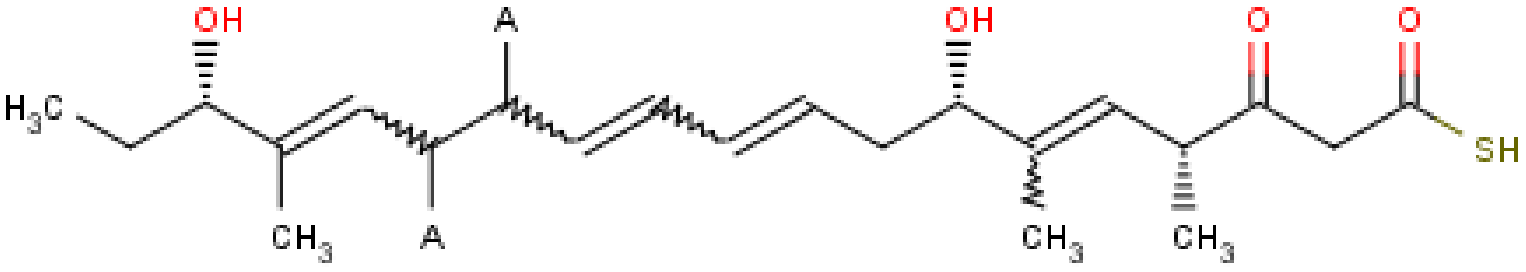
(Oliynyk *et al*, *Nat. Biotechnol.*, ***25***, 447, 2007)

ClustScan results using *Sac. erythrea* genome:



ClustScan results using Sac. erythrea genome:

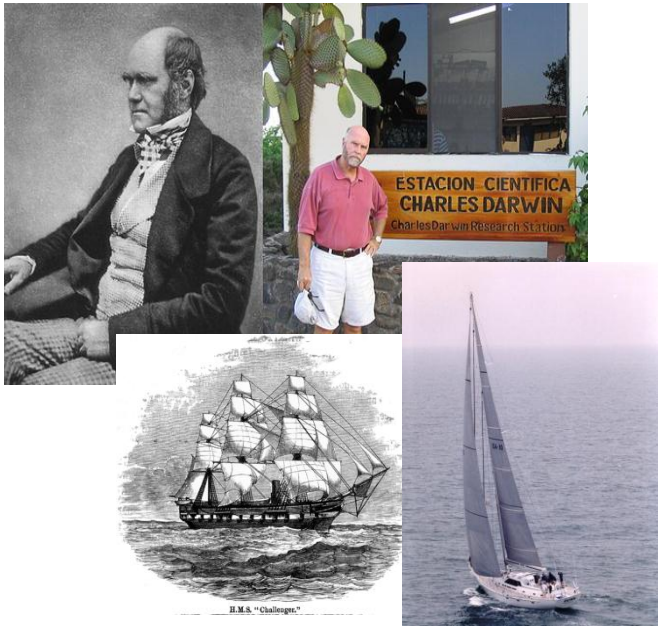
SMILES:
[C@@H](C)[C@H](O)C(C)=CC(*)C(*)C=CC=CC[C@H](O)C(C)=C
[C@@H](C)C(=O)CC(=O)S

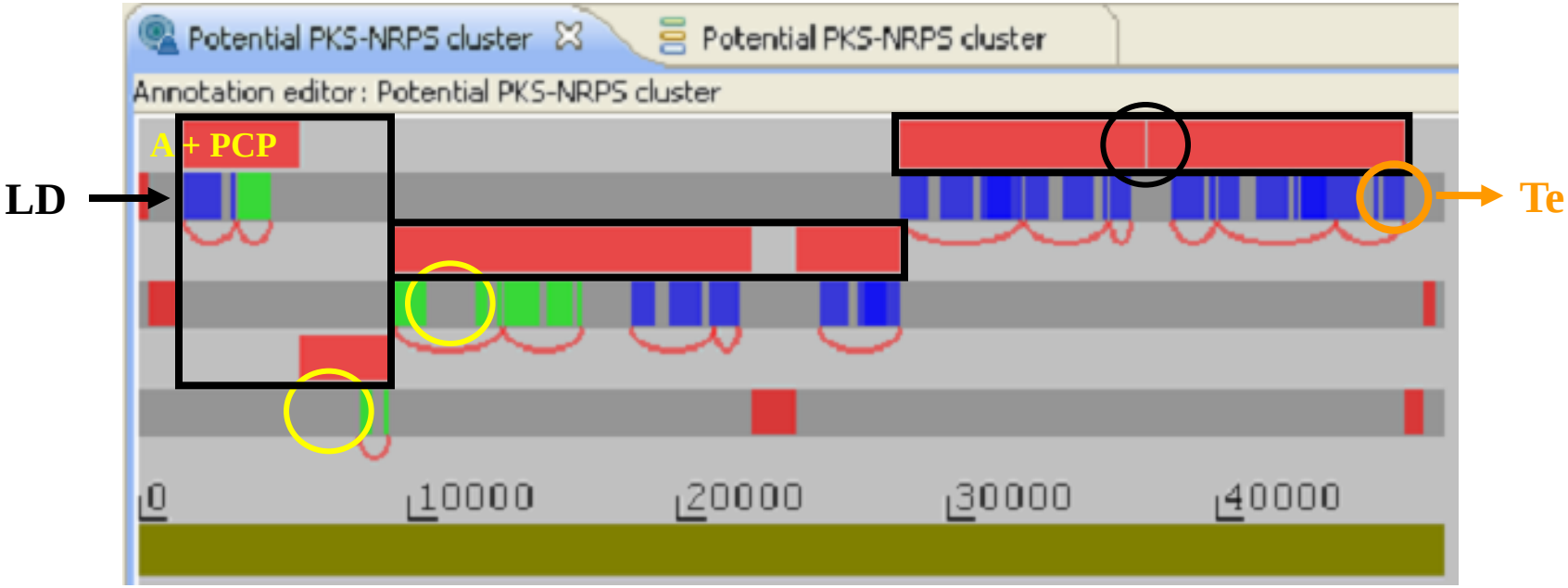


- **Ecopia BioSciences Inc.**
  ***Streptomyces aculeolatus* and *Streptomyces* sp. Eco86**
  Banskota *at al.*, *J. Antibiot.*, ***59***, 168–176, 2006
- **Rolf Müller's group**
  myxobacterial species *Stigmatella aurantiaca*
  Frank *et al.*, *J. Mol. Biol.*, ***374***, 24-38, 2007

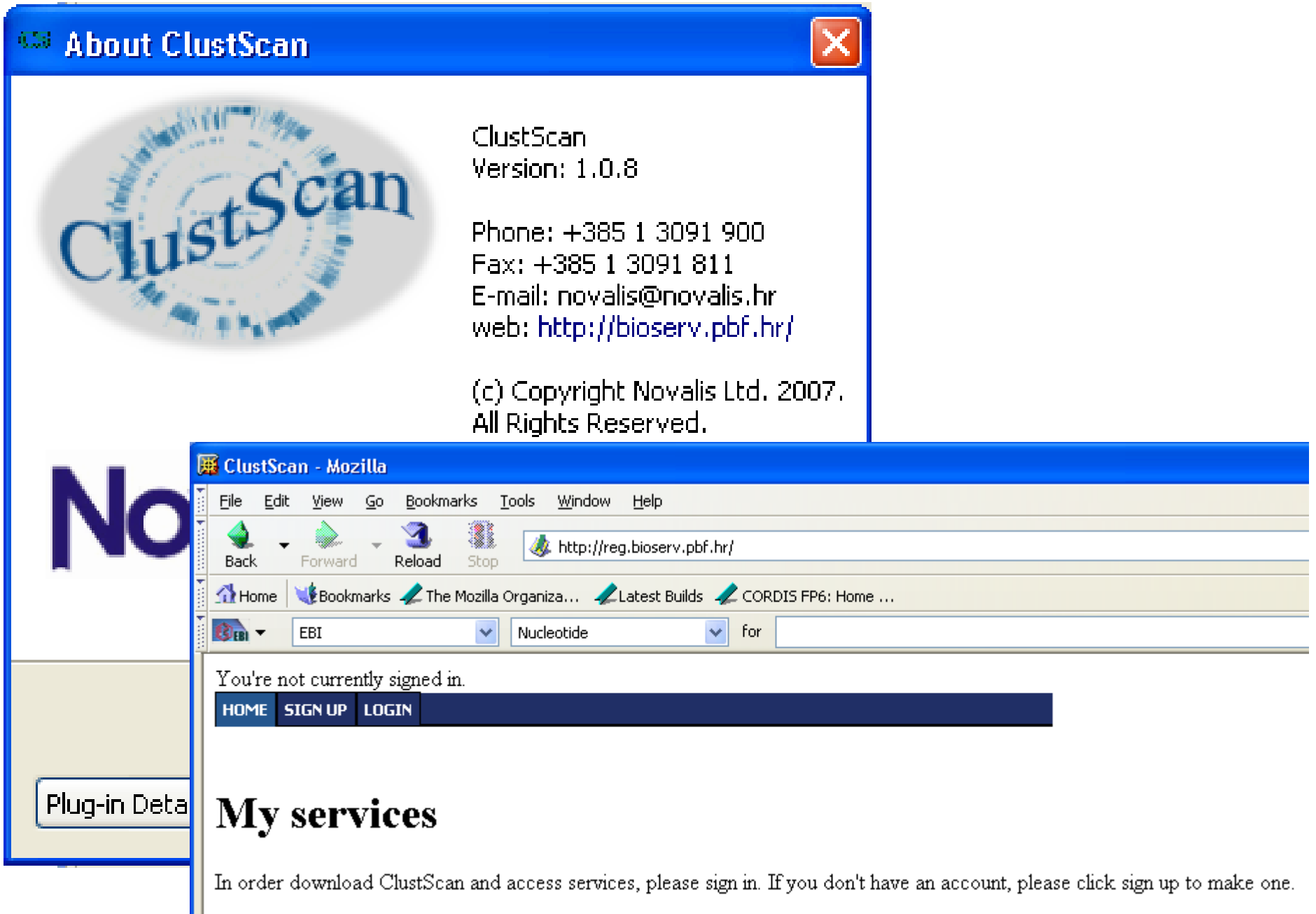
Sequencing of largest metagenomic dataset:

(**Rusch** *et al.*, *PLoS Biol.* **5**, e77, 2007)





ClustScan downloads:



CompGen program package:

- **major goal of *CompGen* is structuring and maintenance of custom database of entirely novel natural products developed by *in silico* modelling of homologous recombination**

- **future of *CompGen* will be prediction of biological activities of *in silico* generated novel chemical entities using CADD technology**

Why are Streptomyces good for recombination?

- **high G+C-content - longer stretches of identity**
- **no *mut*SL genes - these prevent recombination between sequences with mismatches**
- **linear DNA allows single crossovers**

✓ **have *rec*A gene - homologous pairing**

✓ **no *rec*BCD or *add*AB – DNA helicase/endonuclease**

✓ **4 conserved potential DNA helicases in genomes**

S. coelicolor* A3(2), *S. avermitilis* and *S. scabies

Recombination between PKS clusters:

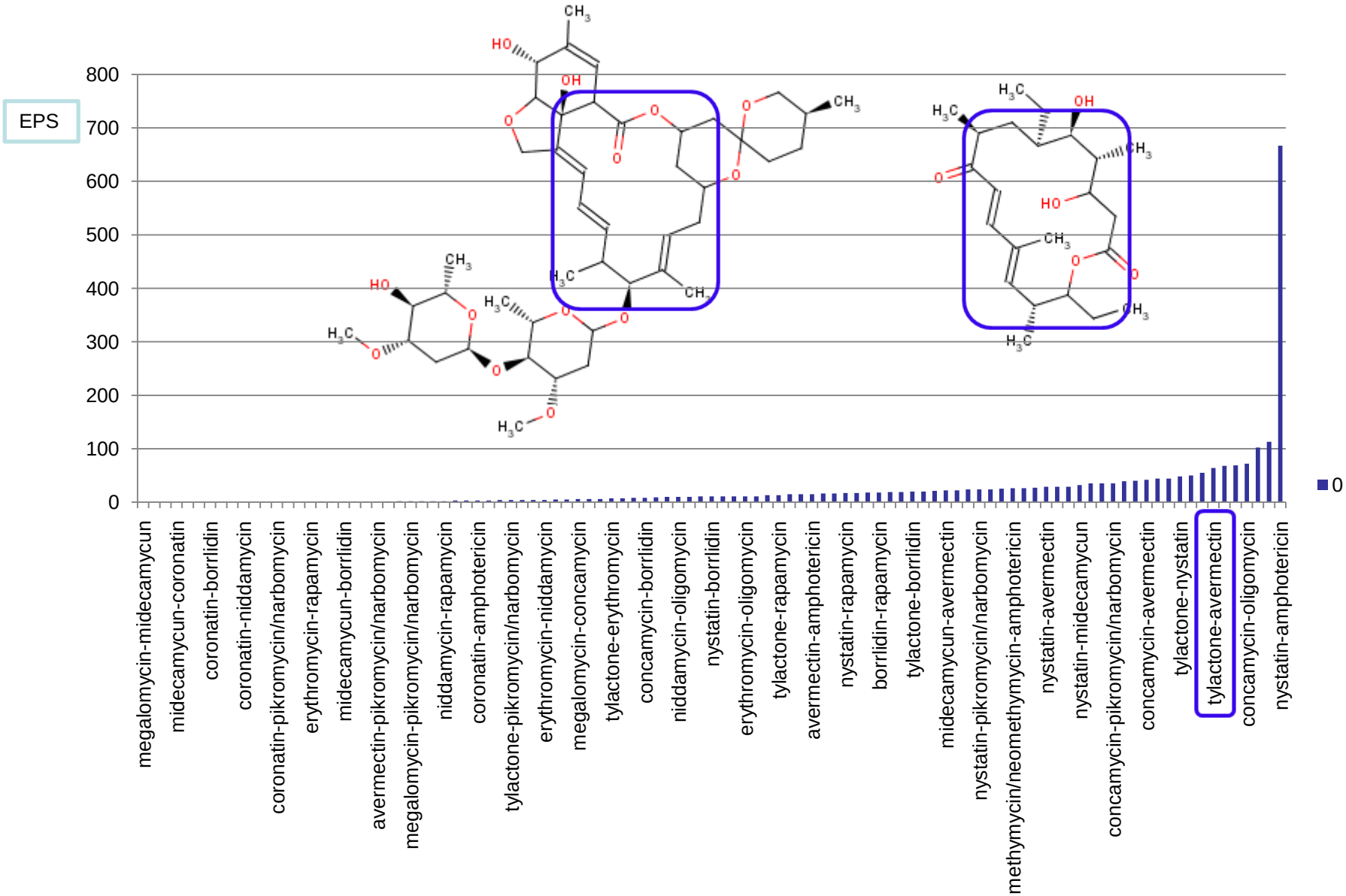
- **needs Minimal Essential Pairing Sequence (MEPS; ~ 30 bp) in region of reasonable matching (> 75% in ~ 200 bp)**

- **recombination program finds regions of identities and similarities**

- **coordinates are matched to custom database to find which domains are involved to allow product prediction**

Number of recombinants:

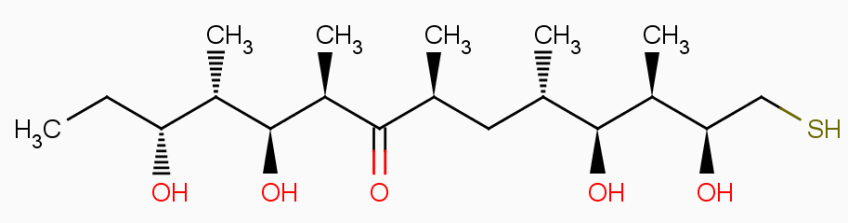
- **use parameters of 27 bp MEPS identity and 75% similarity in 200 bp**
- **2445 recombination points (15 clusters)**
- **cca. 47 recombination points/cluster pair**
- **number of viable recombinants much lower, depends on additional fitness criteria imposed (2-4 rec point/cluster pair)**

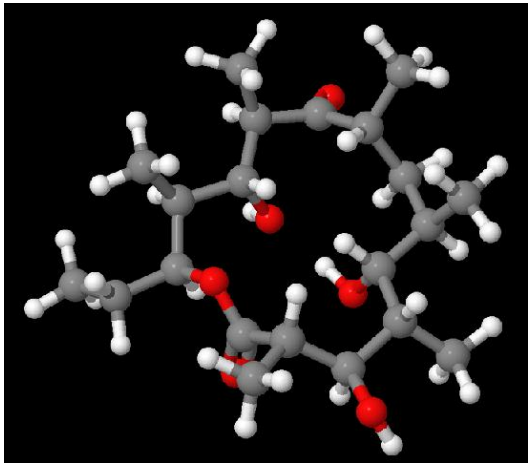


Number of EPS (**E**ffective **P**airing **S**equences) in case of PKSs, correlates with overall sequence similarity, sequence length and chemical structure.

Erythromycin and niddamycin gene-clusters:

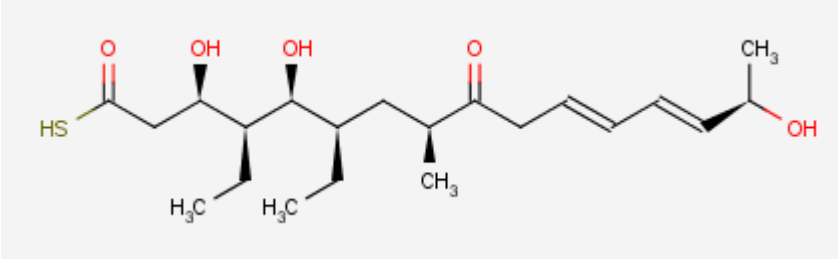
ery 14-membered macrolide

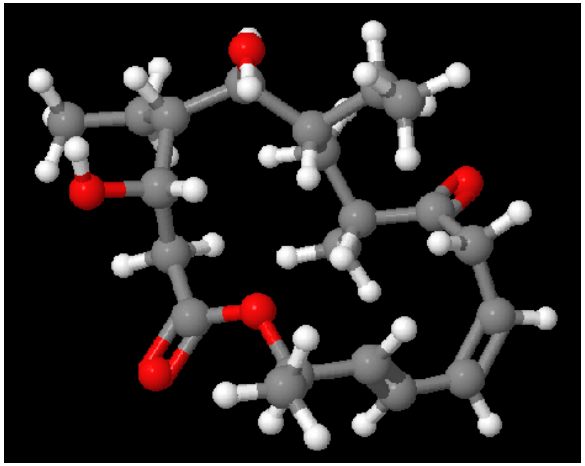




3 genes and 6 modules – 7 building blocks

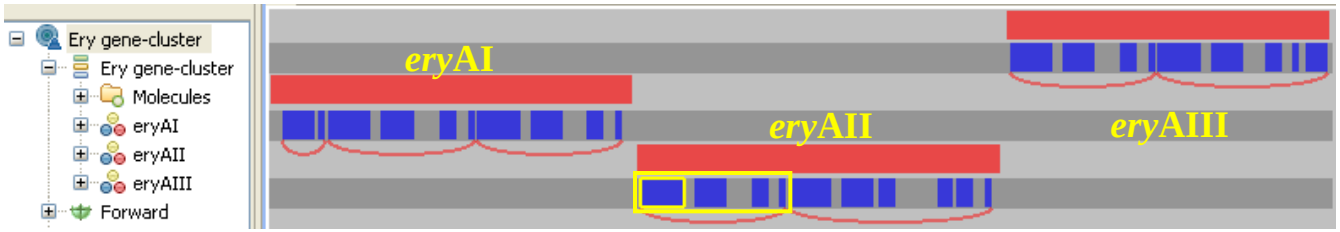
nid 16-membered macrolide



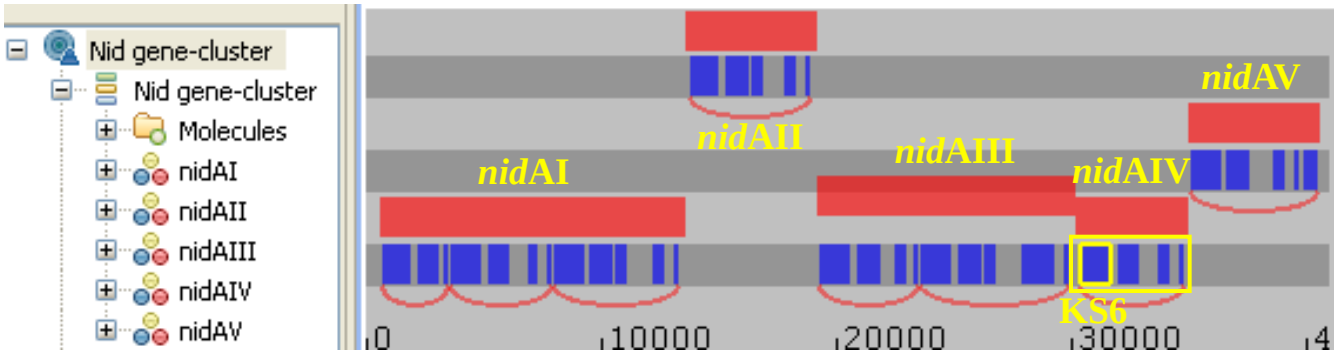


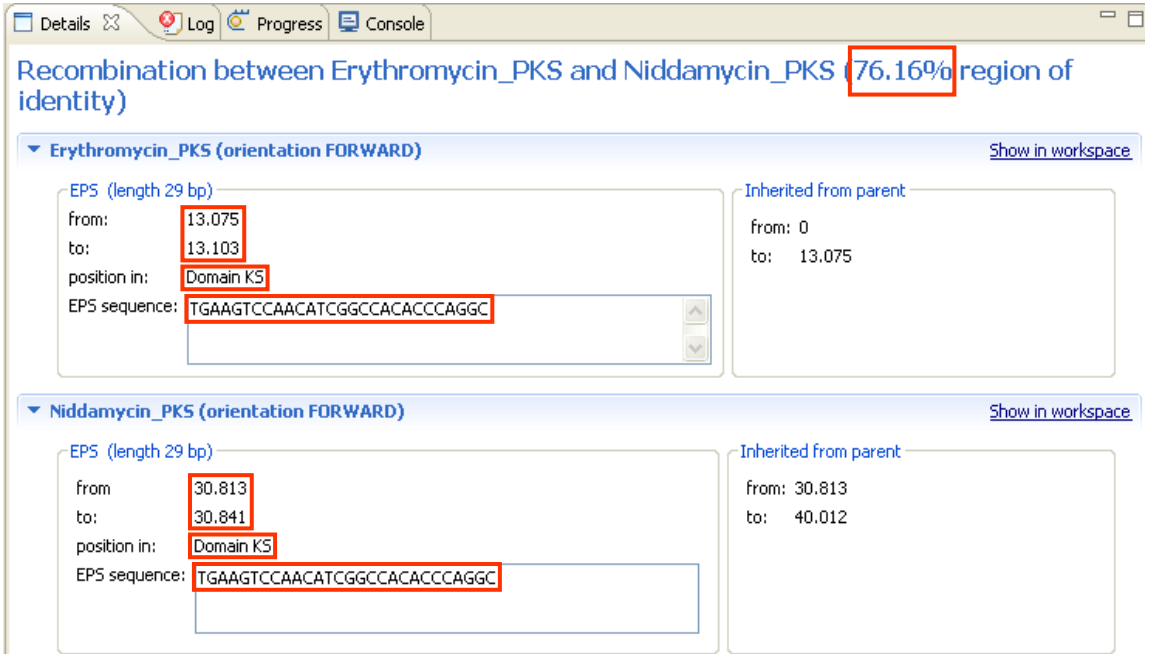
5 genes and 7 modules – 8 building blocks

Recombination between PKS clusters cont':



KS3



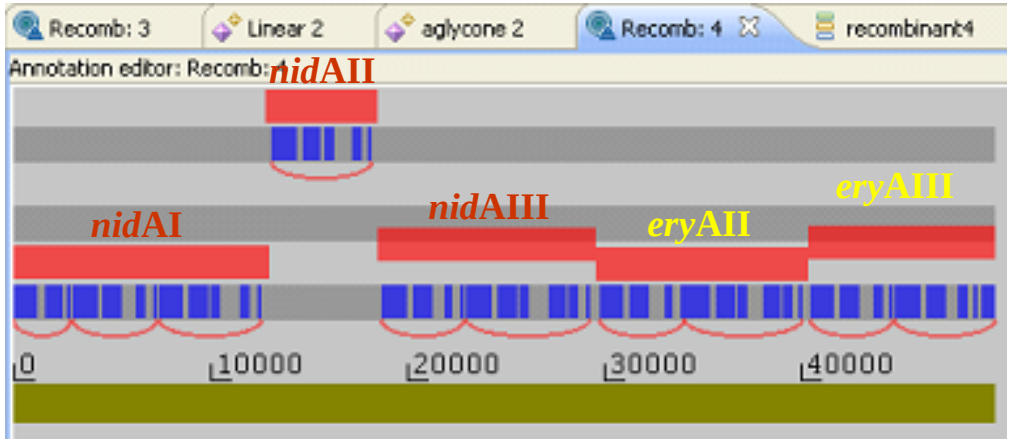


CompGen program package cont':

Rec-2

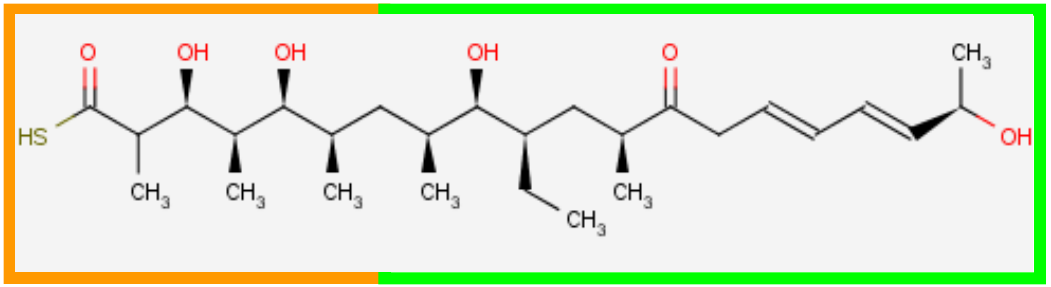
SMILES:

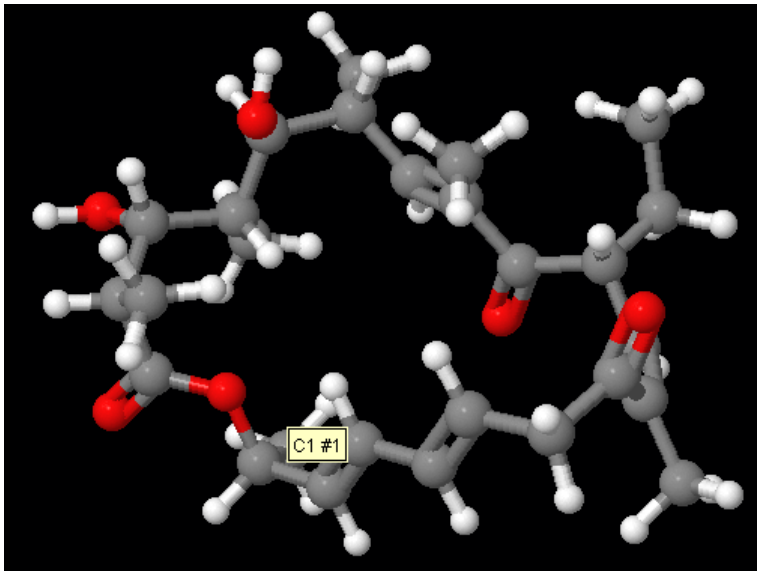
C[C@@H](O)C=CC=CCC(=O)[C@@H](C)C[C@@H](CC)[C@H](O)[C@@H](C)C[C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C(=O)S



5 genes, 9 modules and 10 building blocks

18-membered macrolide



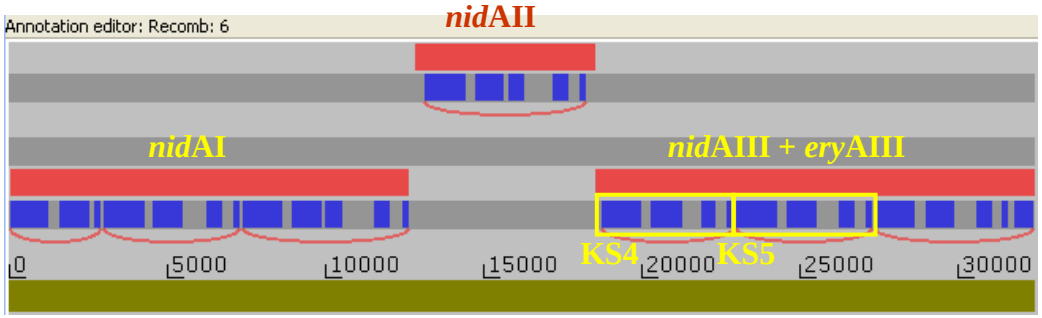


CompGen program package cont':

Rec-3

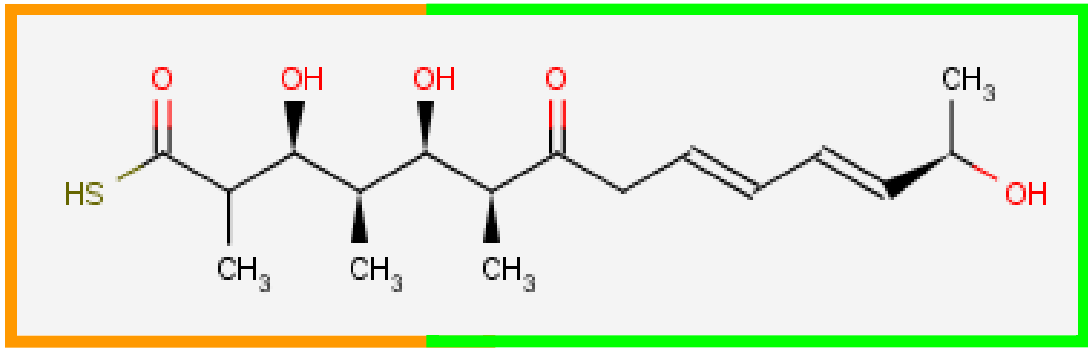
SMILES:

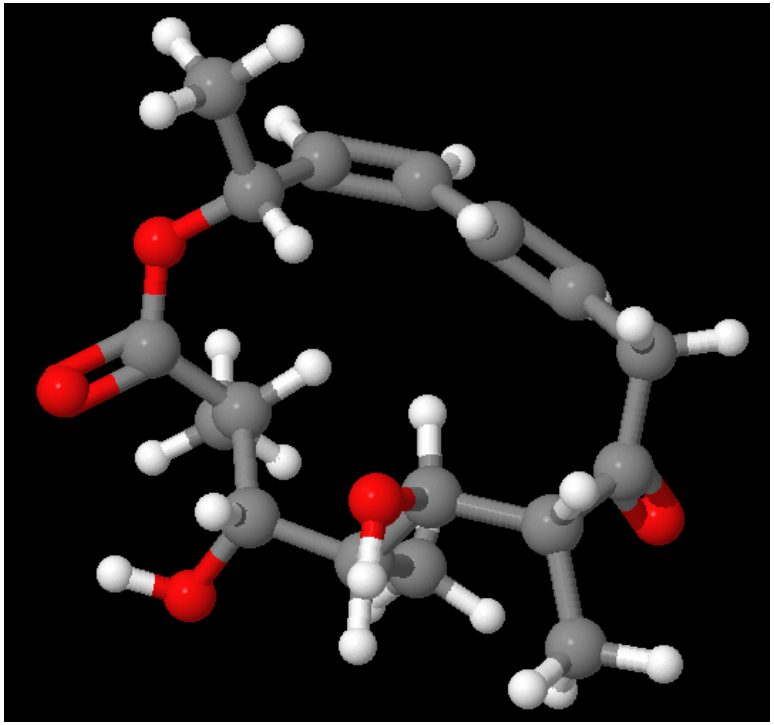
C[C@@H](O)C=CC=CCC(=O)[C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C(=O)S



3 genes, 6 modules and 7 building blocks

14-membered macrolide





Summary:

- ***ClustScan*** **prototype completed**
- ***CompGen*** **prototype completed**
- **reverse** ***CompGen*** **established**
- **when product looks promising** ***in silico*****, "designer bug" can be created**