**Comparative Genomics 2018**

**Practical 6: Orthology Prediction**

Group 11

Shuhan Xu, Fuqi Xu, Milda Valiukonyte

**Summary**

<Text>

**Key questions**

2.

database: InParanoid, PhylomeDB, HieranoiDB

organism: Escherichia coli

protein identifiers for three selected genes:

P30143 (./09.fa.txt\_orf00010\_rev)

P0AF03 (./09.fa.txt\_orf00012)

P0A6Y8 (./09.fa.txt\_orf00018)

Algorithms of the databases:

Motivation for choice of databases:

**blastp -outfmt 7 -query 09.fa.txt.pfa -db 83333.fasta -max\_target\_seqs 1 | head -40**

Escherichia coli: 09.fa.txt vs 83333.fasta

./09.fa.txt\_orf00002 P00561

./09.fa.txt\_orf00005 P00934

./09.fa.txt\_orf00009\_rev P0A8I3

./09.fa.txt\_orf00010\_rev P30143

./09.fa.txt\_orf00011 P0A870

./09.fa.txt\_orf00012 P0AF03

./09.fa.txt\_orf00018 P0A6Y8

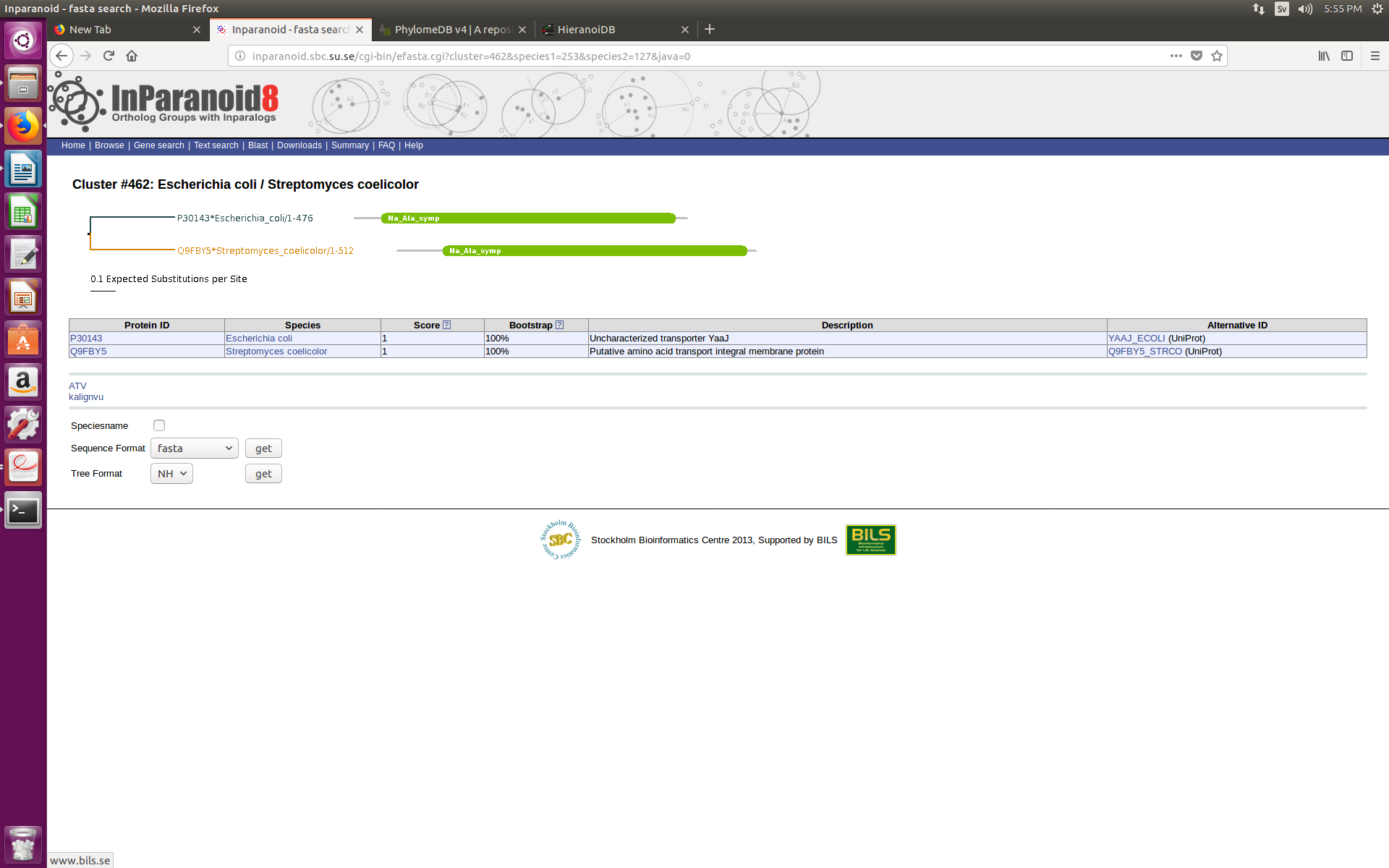
Streptomyces coelicolor: 17.fa.txt vs 100226.fasta

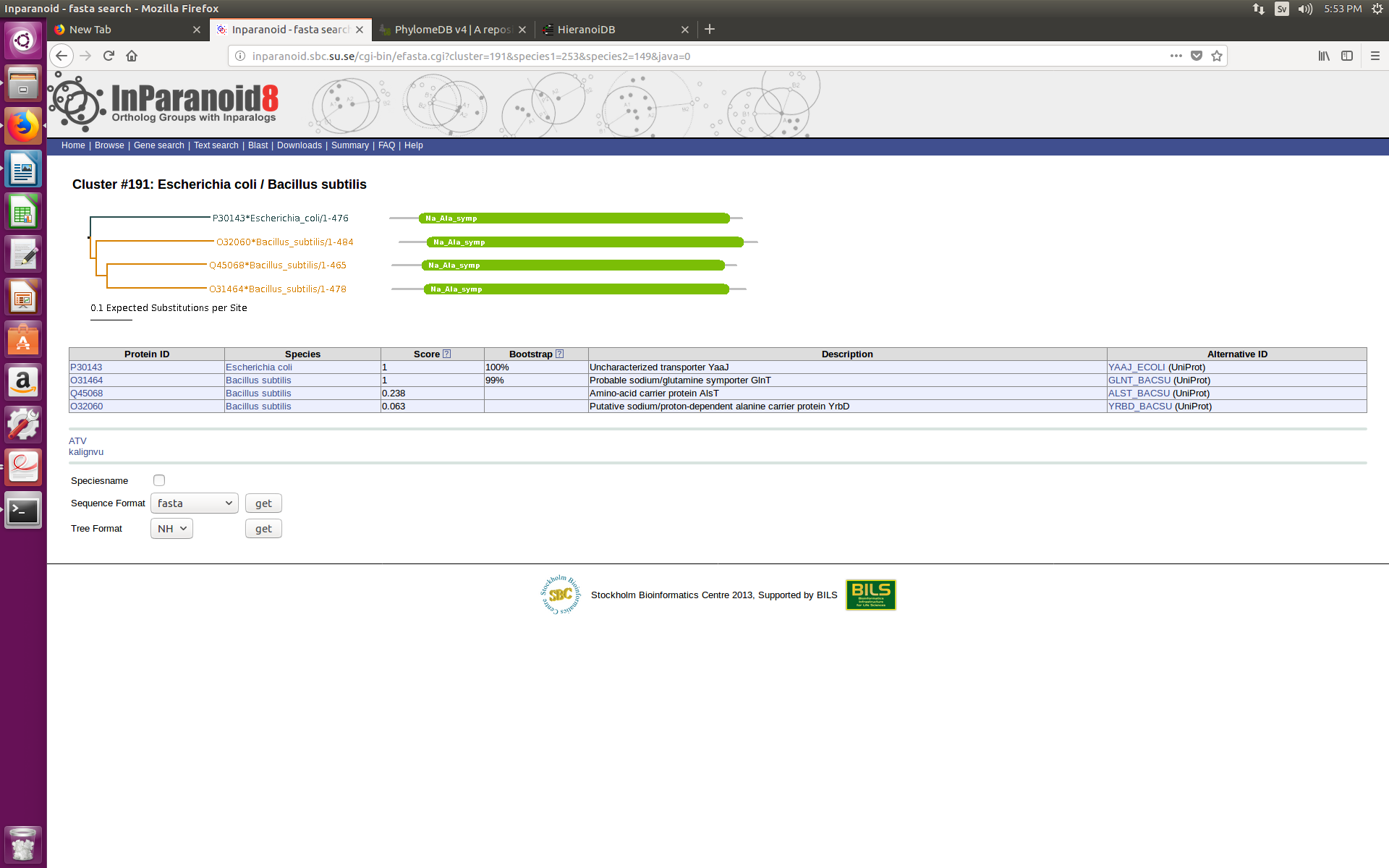
3.

a.

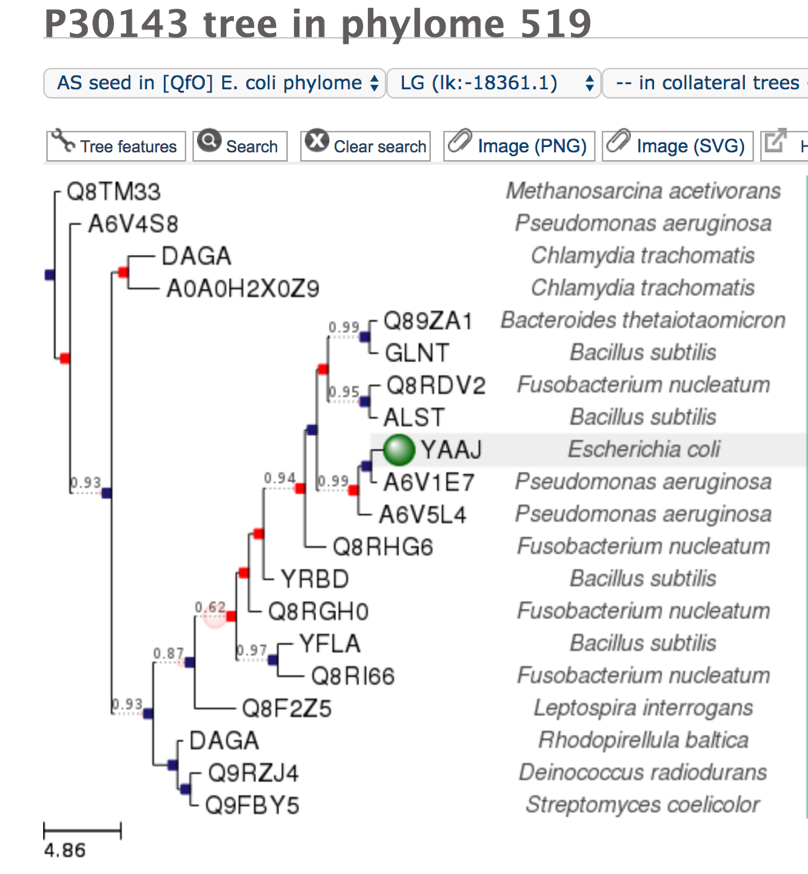
For Escherichia coli P30143, we searched for orthologs in Streptomyces coelicolor and Bacillus subtilis.

InParanoid:

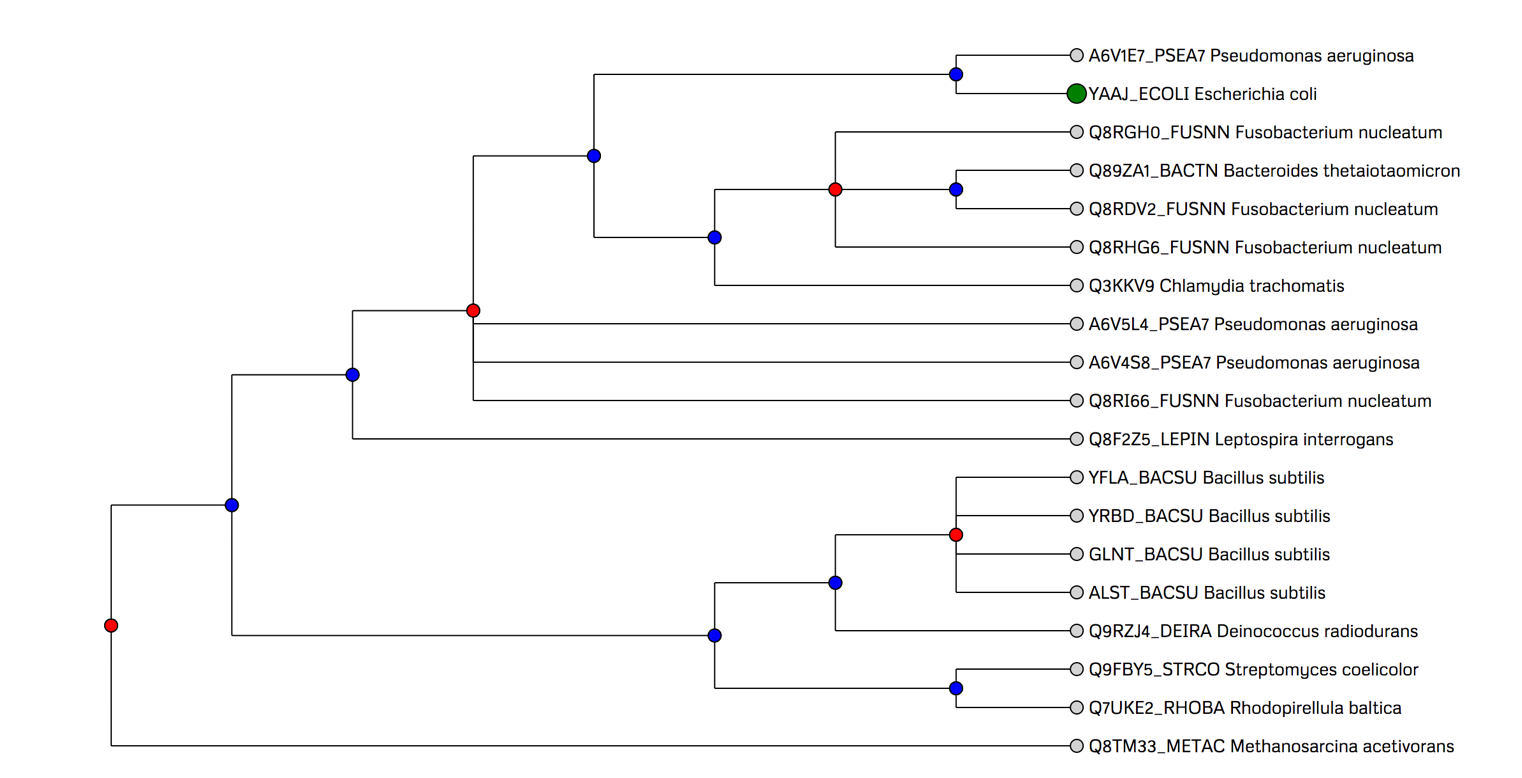




PhylomeDB:



HieranoidDB:

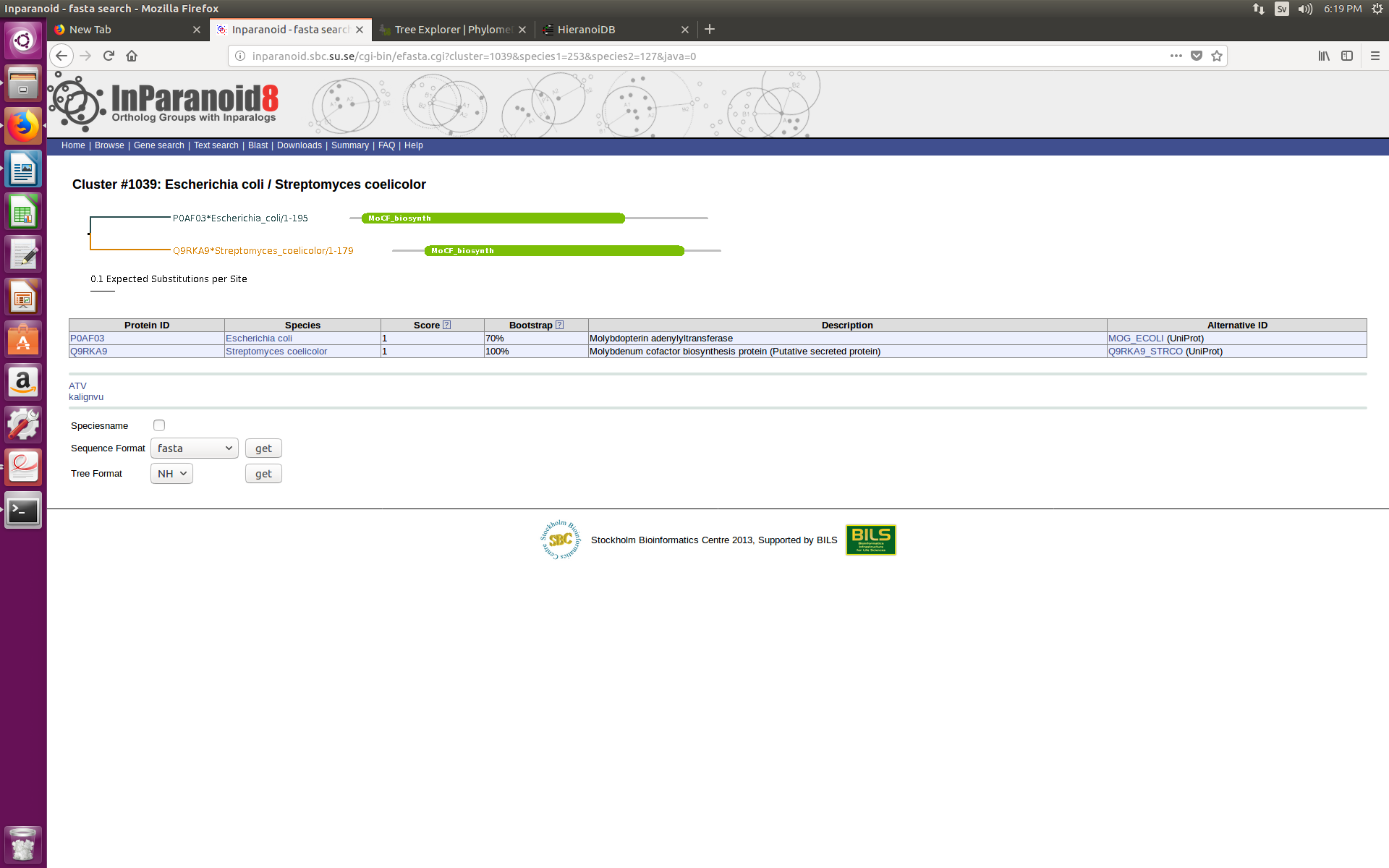


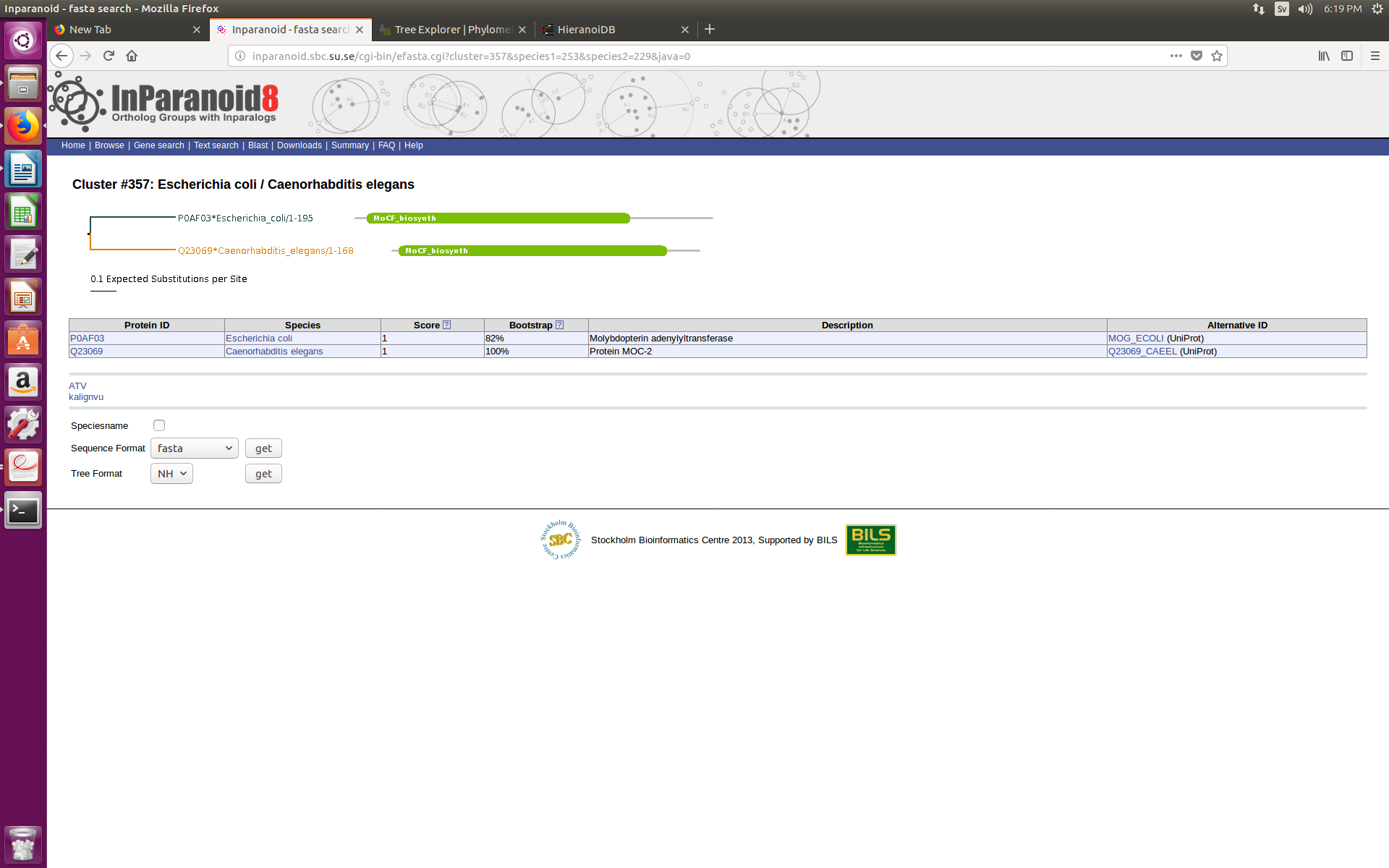
For Streptomyces coelicolor, all three databases identified Q9FBY5 as the ortholog of Escherichia coli P30143 (YAAJ).

For Bacillus subtilis, InParanoid identified O31464 (GLNT), Q45068 (ALST), O32060 (YRBD) as co-orthologs of Escherichia coli P30143(YAAJ). However, PhylomeDB identified O31464 (GLNT) and Q45068 (ALST) as co-orthologs and O32060 (YRBD) and O34708 (YFLA) as out-paralogs. HieranoidDb identified O31464 (GLNT), Q45068 (ALST), O32060 (YRBD) and O34708 (YFLA) as co-orthologs.

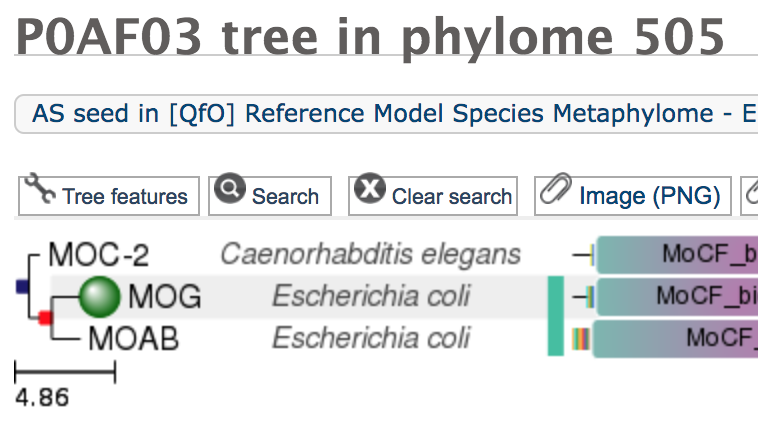
For Escherichia coli P0AF03, we searched for orthologs in Streptomyces coelicolor and Caenorhabditis elegans.

InParanoid:

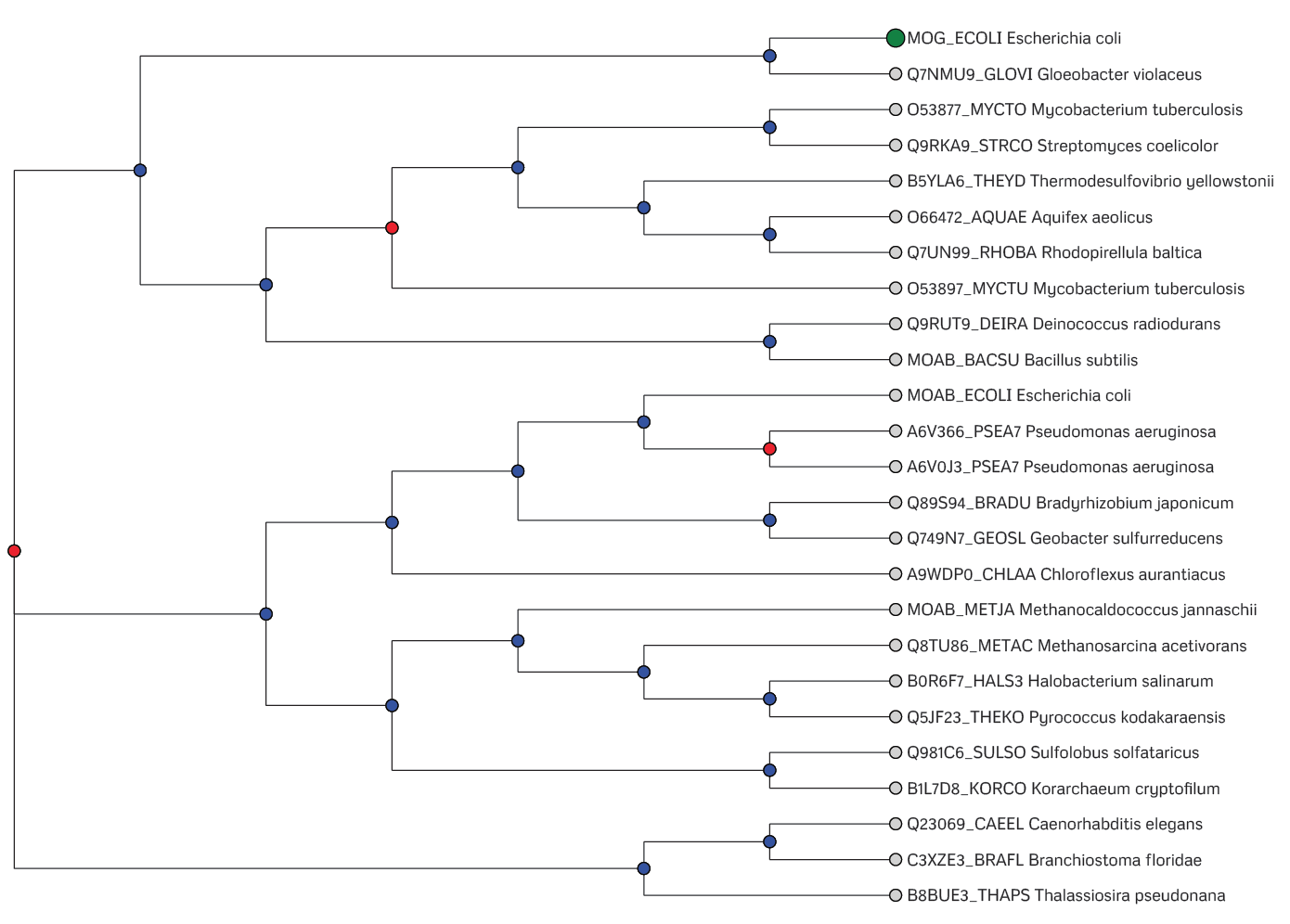




PhylomeDB:



HieranoidDB:

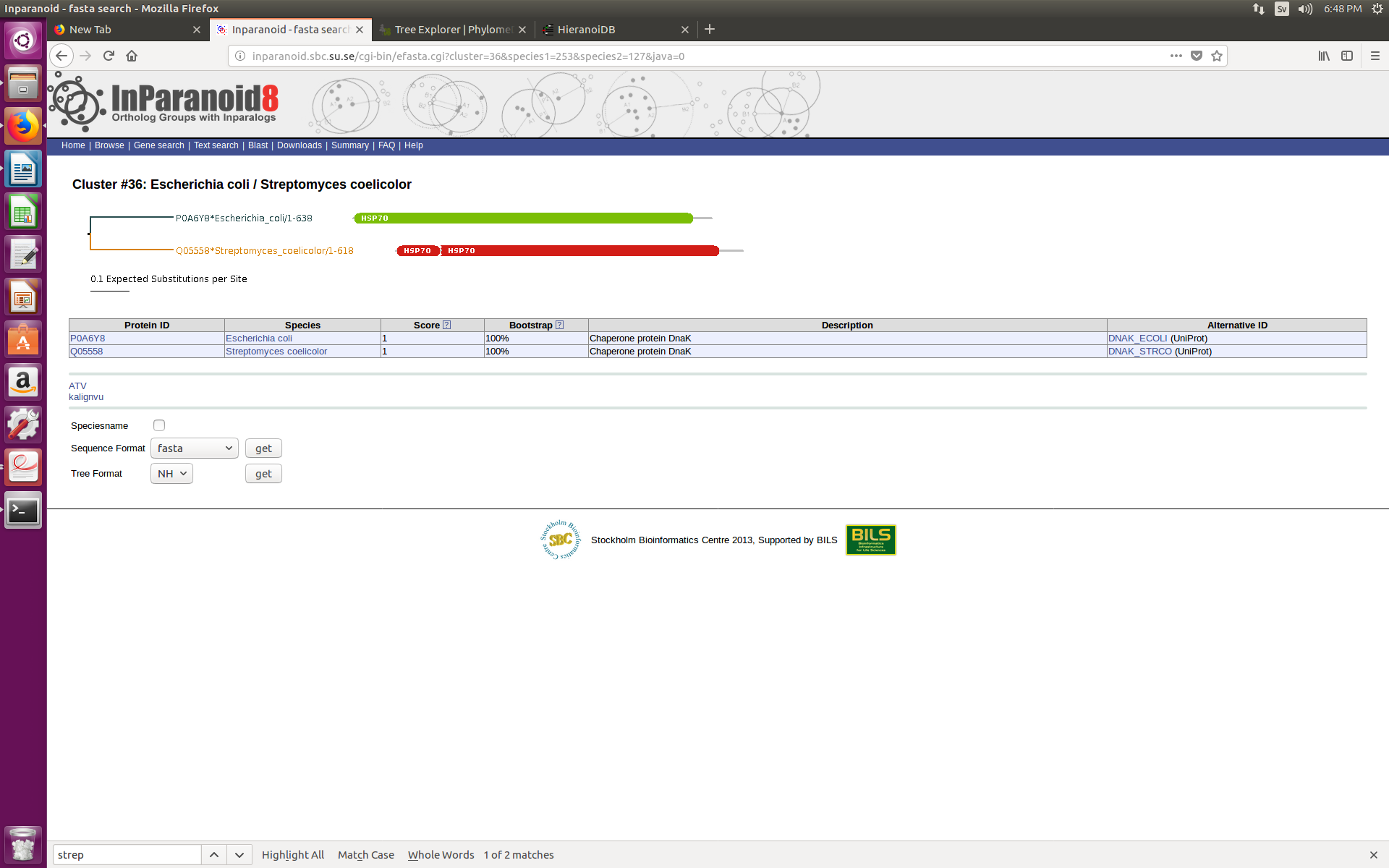


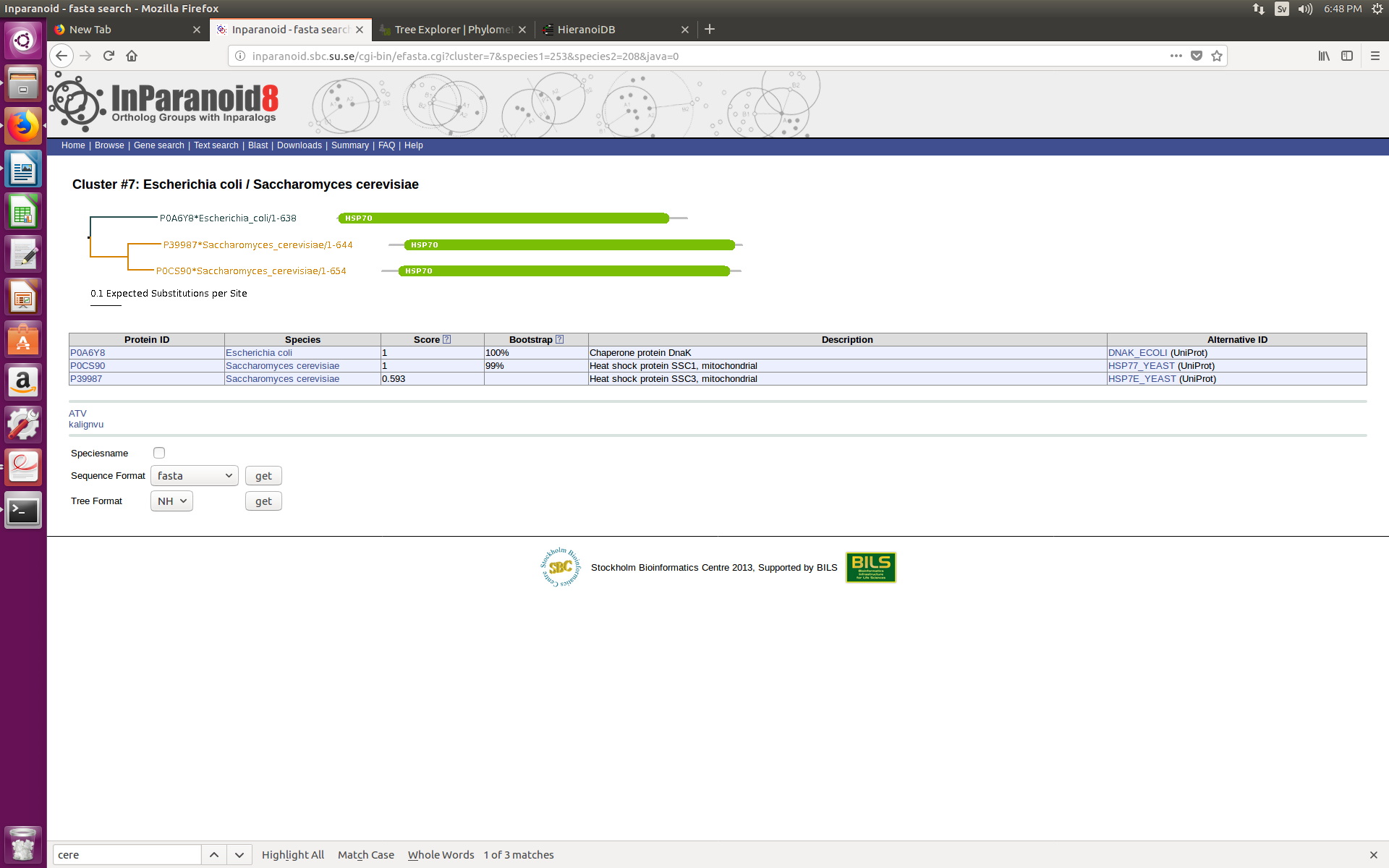
For Streptomyces coelicolor, only InParanoid and HieranoiDB identified Q9RKA9 as the ortholog of Escherichia coli P0AF03 (MOG). PhylomeDB did not identify any ortholog in Streptomyces coelicolor.

For Caenorhabditis elegans, all 3 databases identified Q23069 (MOC-2) as ortholog of Escherichia coli P0AF03 (MOG). In addition, PhylomeDB also identified Escherichia coli P0AEZ9 (MOAB) as in-paralog of Escherichia coli P0AF03 (MOG).

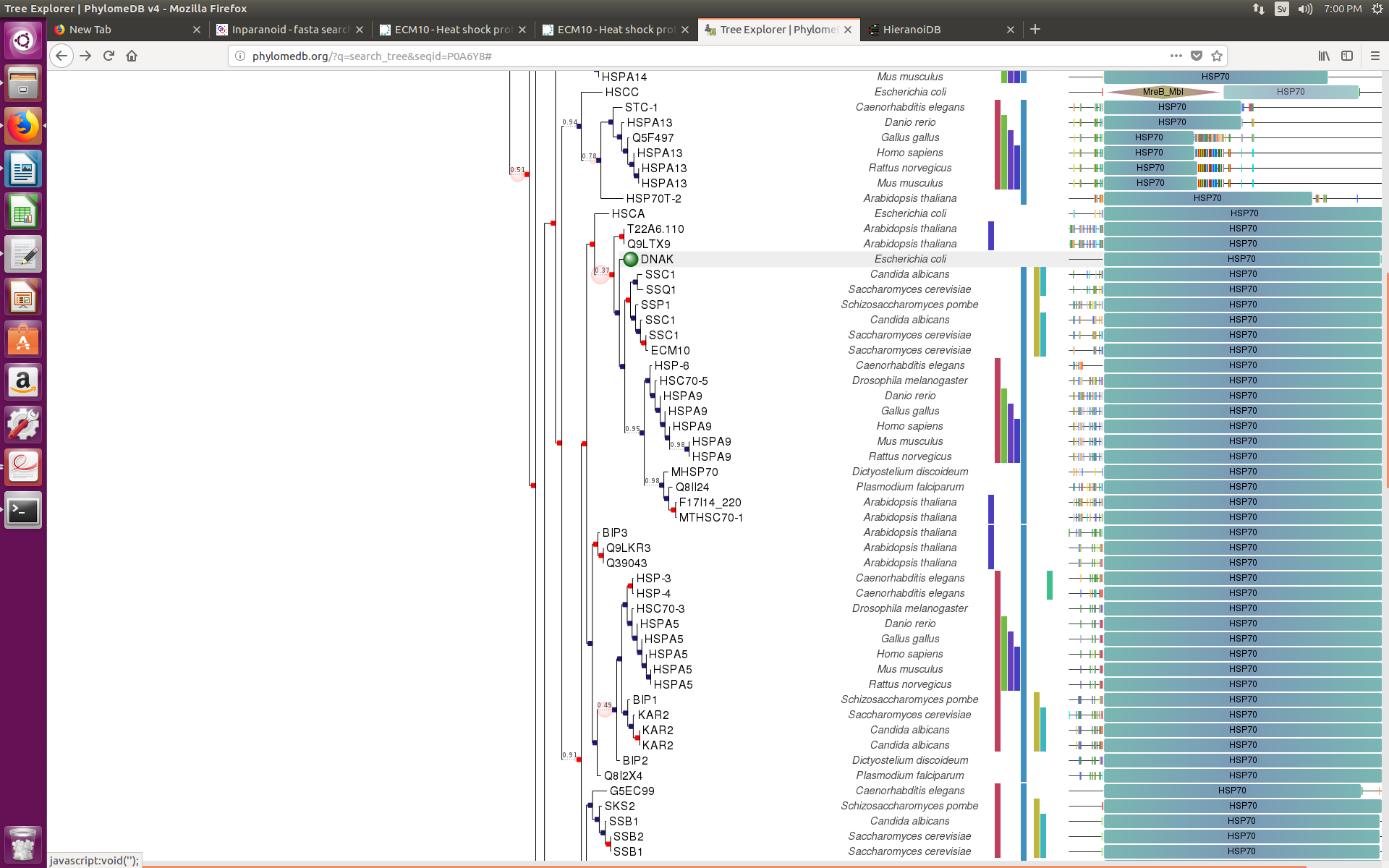
For Escherichia coli P0A6Y8, we searched for orthologs in Streptomyces coelicolor and Saccharomyces cerevisiae.

InParanoid:

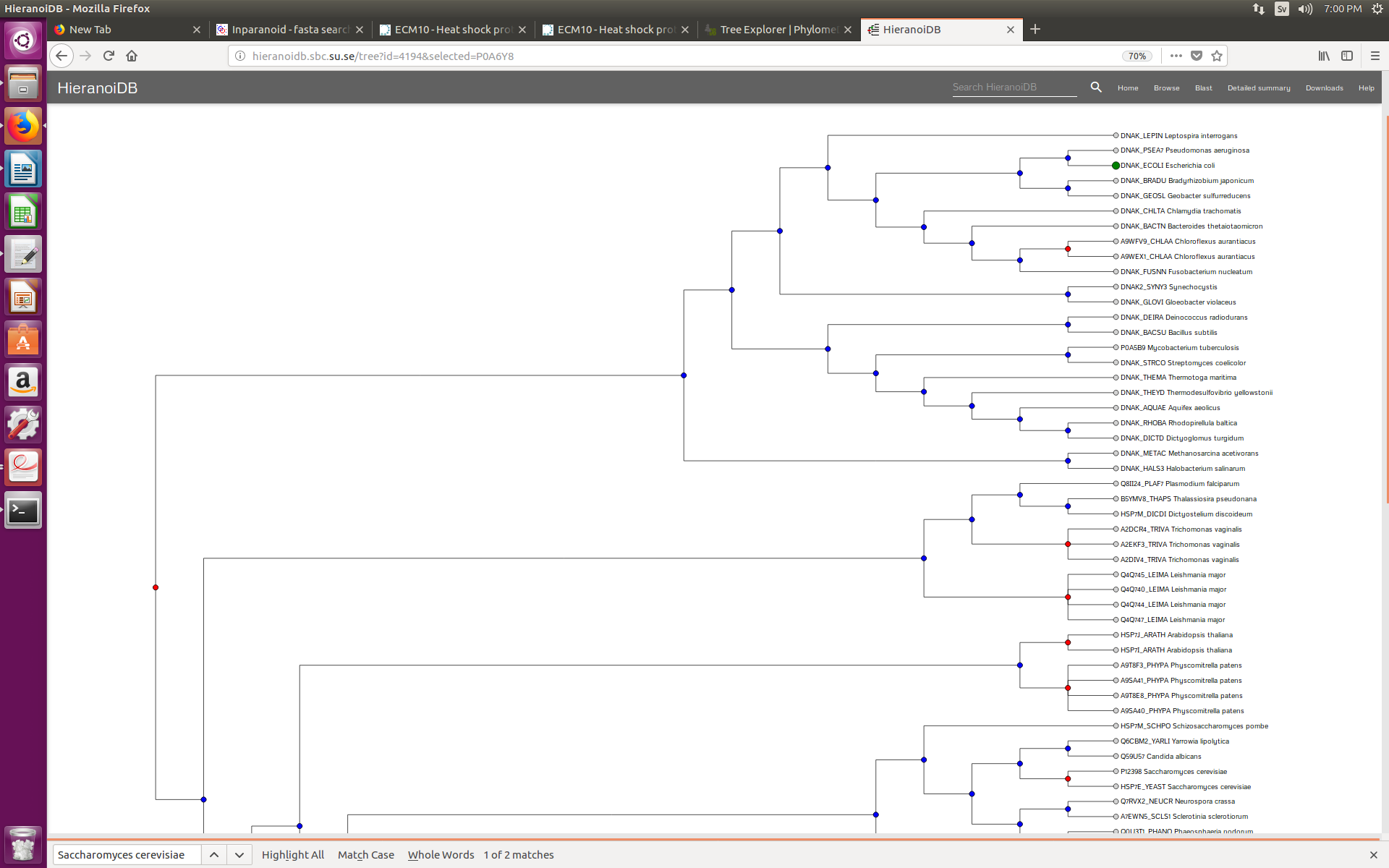




PhylomeDB:



HieranoidDB:



For Streptomyces coelicolor, only InParanoid and HieranoiDB identified Q05558 (DNAK\_STRCO) as the ortholog of Escherichia coli P0A6Y8 (DNAK\_ECOLI). PhylomeDB did not identify any ortholog in Streptomyces coelicolor.

For Saccharomyces cerevisiae, while InParanoid and PhylomeDB identified P0CS90 (P12398/SSC1) and P39987 (ECM10) as co-orthologs of Escherichia coli P0A6Y8 (DNAK\_ECOLI), HieranoiDB identified P0CS90 (P12398/SSC1) and P39987 (ECM10) as out-paralogs instead.

c.

Size of ortholog groups for selected genes in databases

|  |  |  |  |
| --- | --- | --- | --- |
|  | InParanoid | PhylomeDB | HieranoiDB |
| P30143 | 11 clusters | 20 genes in tree | 19 genes in tree |
| P0AF03 | 36 clusters | 3 genes in tree | 25 genes in tree |
| P0A6Y8 | 266 clusters | 150 genes in tree | 75 genes in tree |

d.