**Configuration for a single sequence search.**

To search for a single sequence (e.g. a protein kinase from *Bacillius*) in all sequenced genomes, set parameters as in the example:

Job title: PrkC from Bacillus

Sequence File: (Paste sequence INCLUDING HEADER in the text box)

Subject Data Base: All bacteria and archaea

BlastType: Protein

Evalue: 1e-08

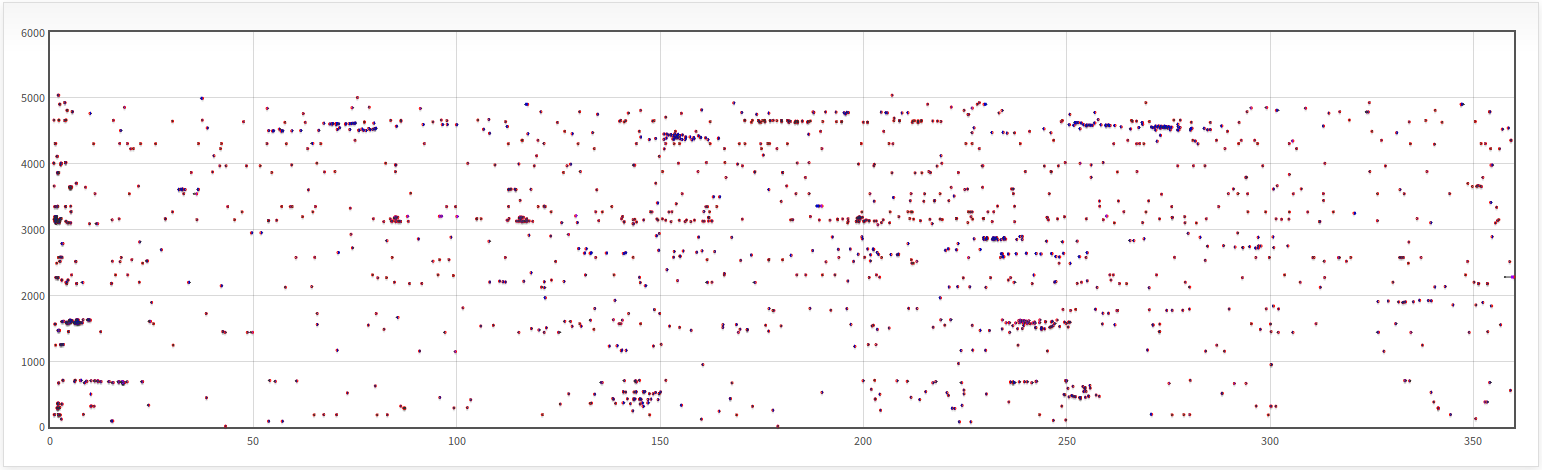
MaxResults: 30000

Because there is only one sequence, it can be directly pasted in the text box. It is necessary to include the header.

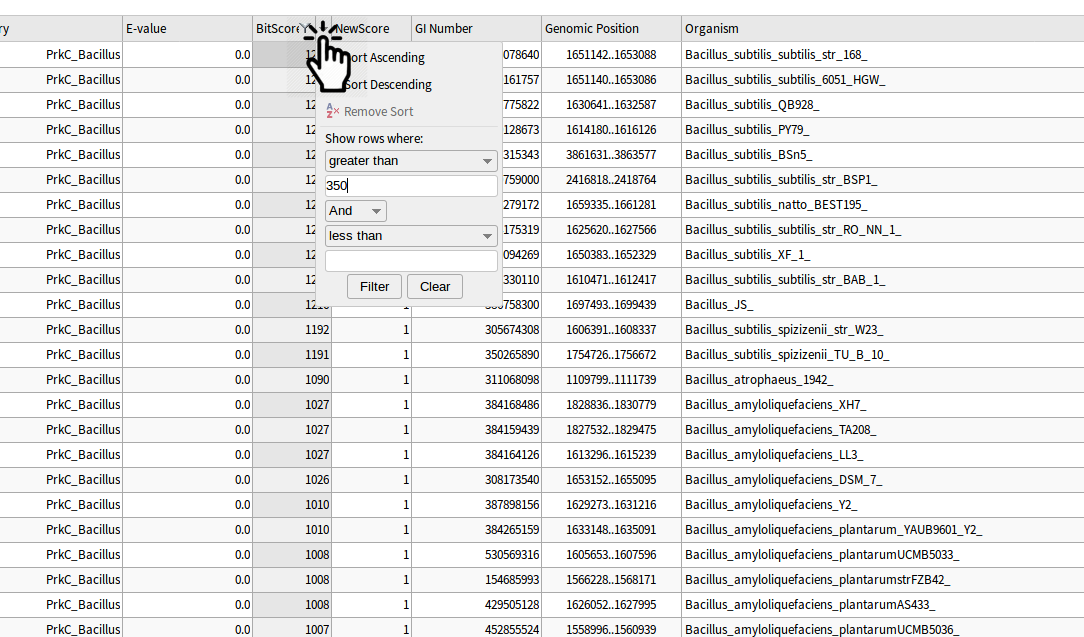
>PrkC\_Bacillus

MLIGKRISGRYQILRVIGGGGMA...

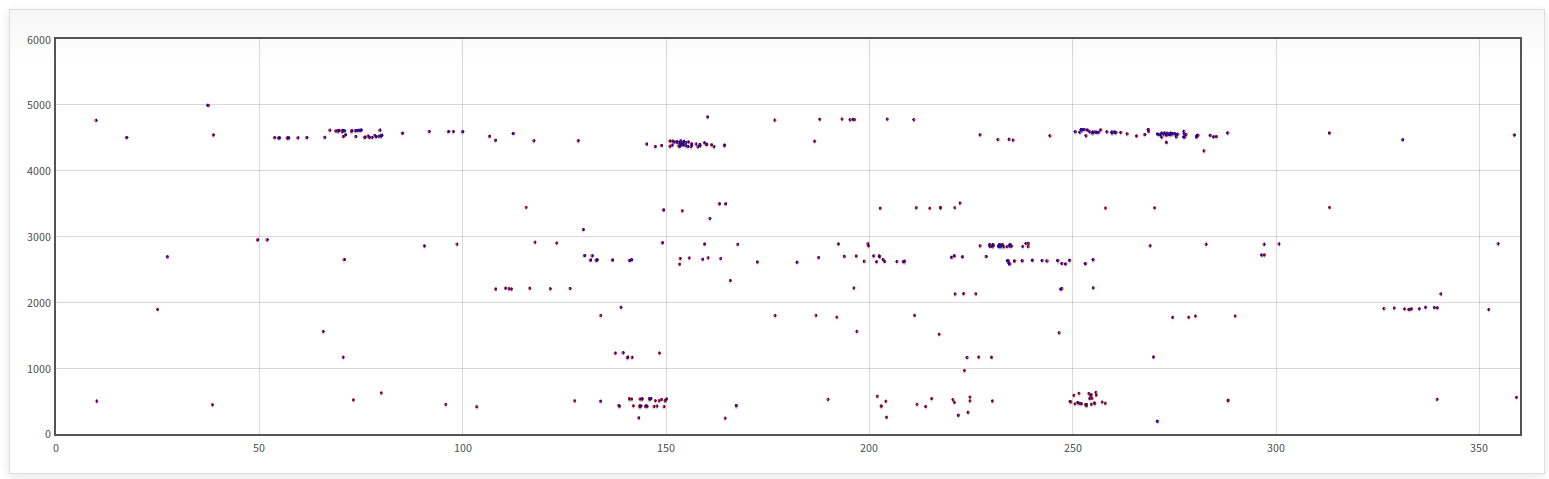
This kind of search typically takes around five minutes, so it is not mandatory to include an e-mail address to get the results. The web page updates automatically and will show links for *Graph* an *Table* when there are ready. Main plot shows the overall distribution of BLAST-results.



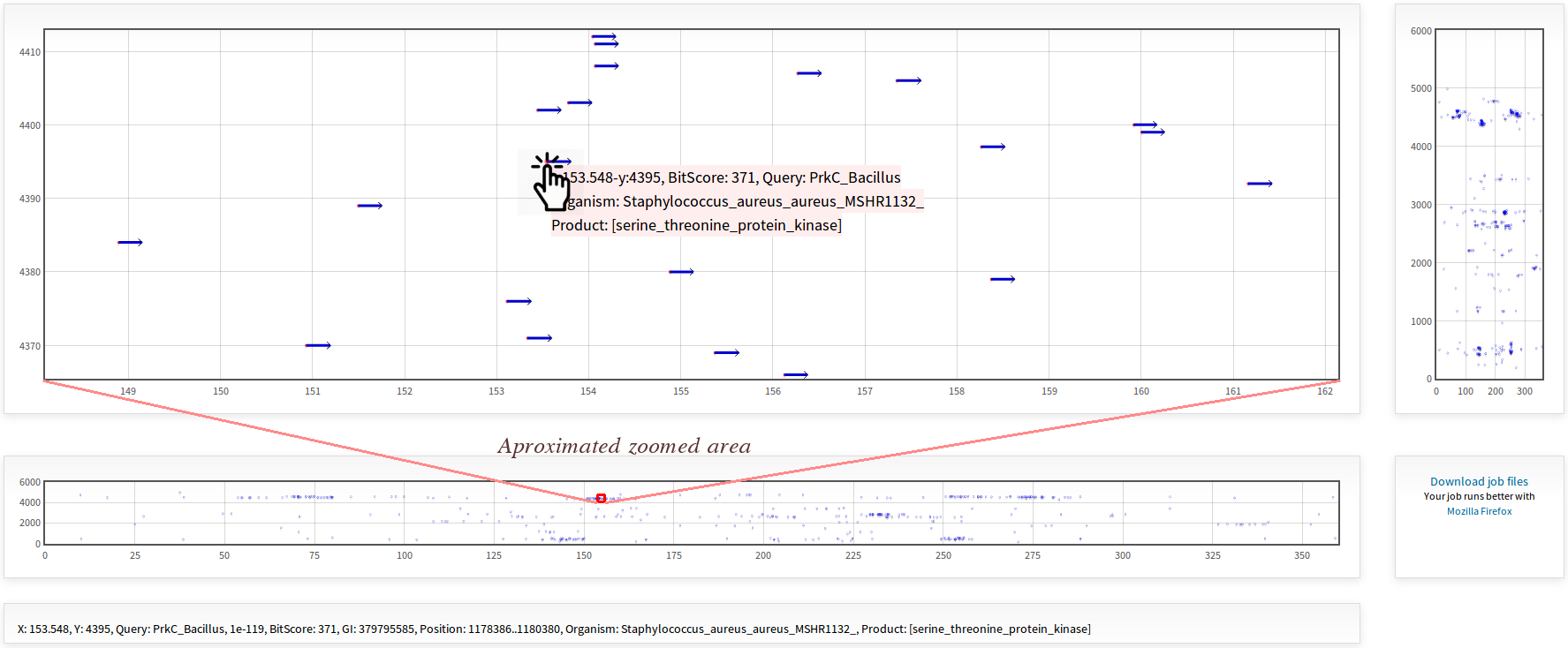
Because the E-value was relative high, the results include non-significant hits. Using the table of data, it is possible to filter results using a BitScore cut of value. For this example, a value “greater than” 350 works well. Click on the “Actualize data” button and then refresh the *Graph* web page.



The plot will be updated with the filtered results and show only the most significant.



By zooming in with the mouse scroll wheel, it is possible to see individual results. When mouseover of one particular result, a pop up text display information about relative position, BitScore value, query name, organisms and product of the subject sequence. When click on it, the text box below shows additional information, like E-value, GI number and real (nucleotide) position.



**Configuration for Capistruin gene cluster search in *Burkholderia*.**

Capistruin is a bacteriocin produced by *Burkholderia thailandensis* E264. Genes involved in the production of capistruin are organized in a locus named *capABCD* that code for bacteriocin precursor, two proteins involved in maturation and an exporter respectively. To search for more locus of this bacteriocin in the *Burkholderia* genus (including 125 replicons at the moment of performing this search) using the XY-plot viewer, parameters could be set as follows:

Job title: Search for Capistruin cluster in Burkholderia genomes

Sequence File: Capistruin.sec

Subject Data Base: Genus

Genus: Burkholderia

BlastType: Genes

Evalue: 1e-08

MaxResults: 30000

A file with sequences must include the headers for each gene:

>capA

ATGGTTCGACTTTTGGCGAAGC...

>capB

ATGCAACGGTCGCGCTATTTTC...

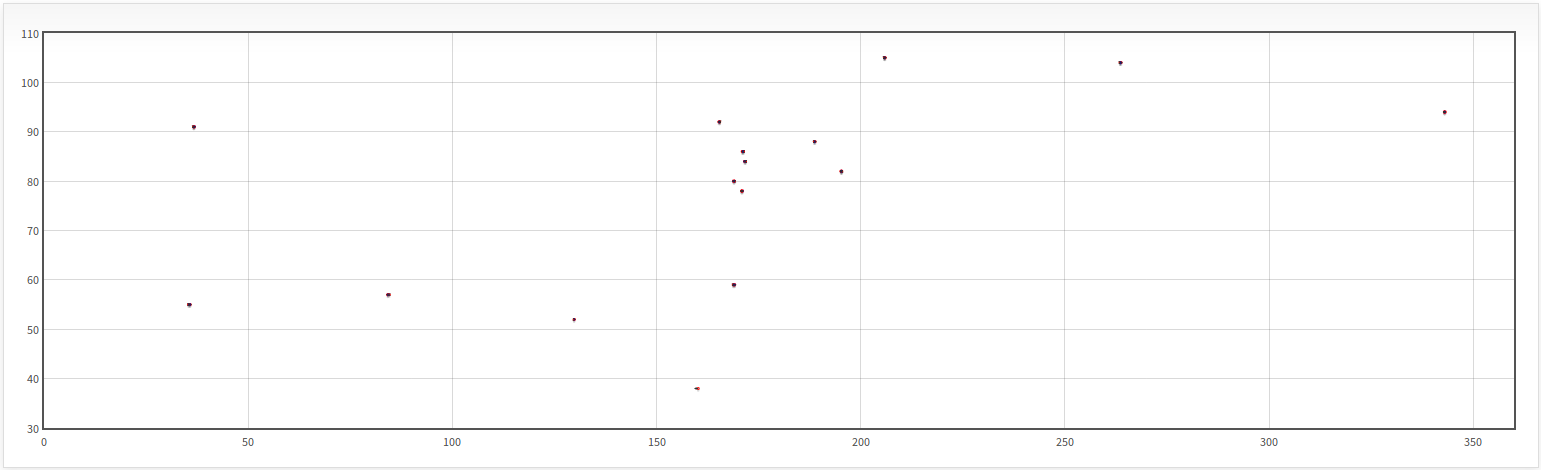
>capC

ATGGCGAAATCTATCGAACGCC...

>capD

ATGGCCCTTCCCATCCGAAACG...

Only one sequence for each gene is necessary, but it is also possible to include multiple sequences for the same gene (orthologous genes) in order to increase the coverage of sequence for those genes with low nucleotide conservation. The running time for this example would take around one minute. In this example, the plot for results showed fifteen capistruin gene clusters (and one isolated gene) where only eight contained an annotated *capA* homologue precursor peptide.



**Configuration for the search of secretion systems.**

Secretion systems are composed by different quantity and types of proteins. To search for secretion systems in a specific genus (e.g. *Burkholderia*), parameters can be defined as follows:

Job title: Search for Secretion Systems in Burkholderia genomes

Sequence File: All-secretion-systems.sec

Subject Data Base: Genus

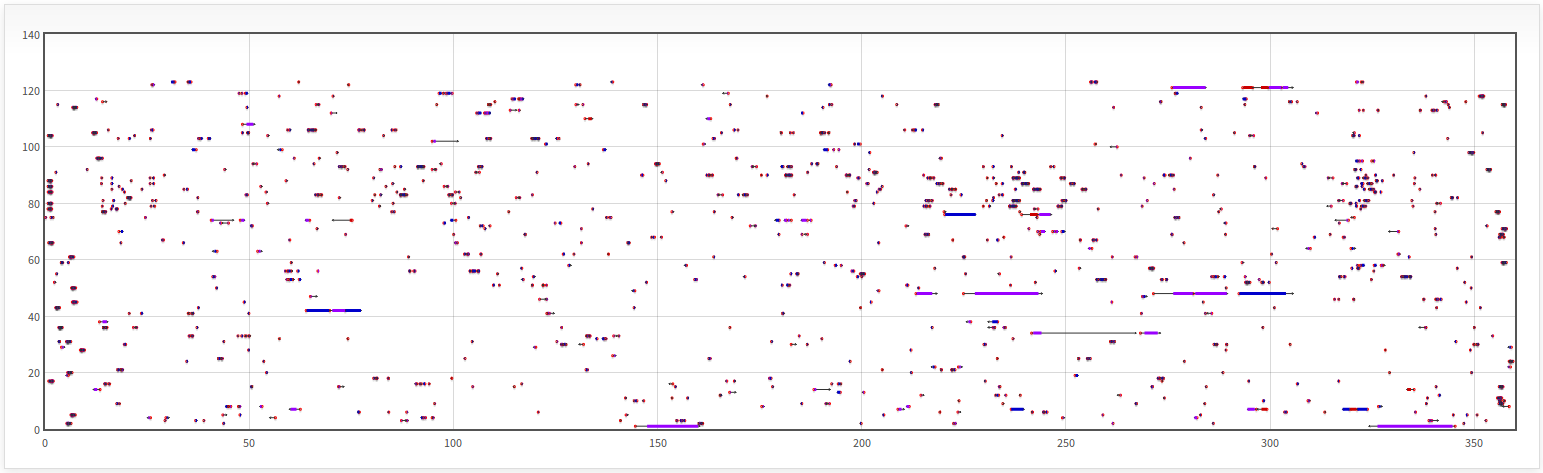
Genus: Burkholderia

BlastType: Proteins

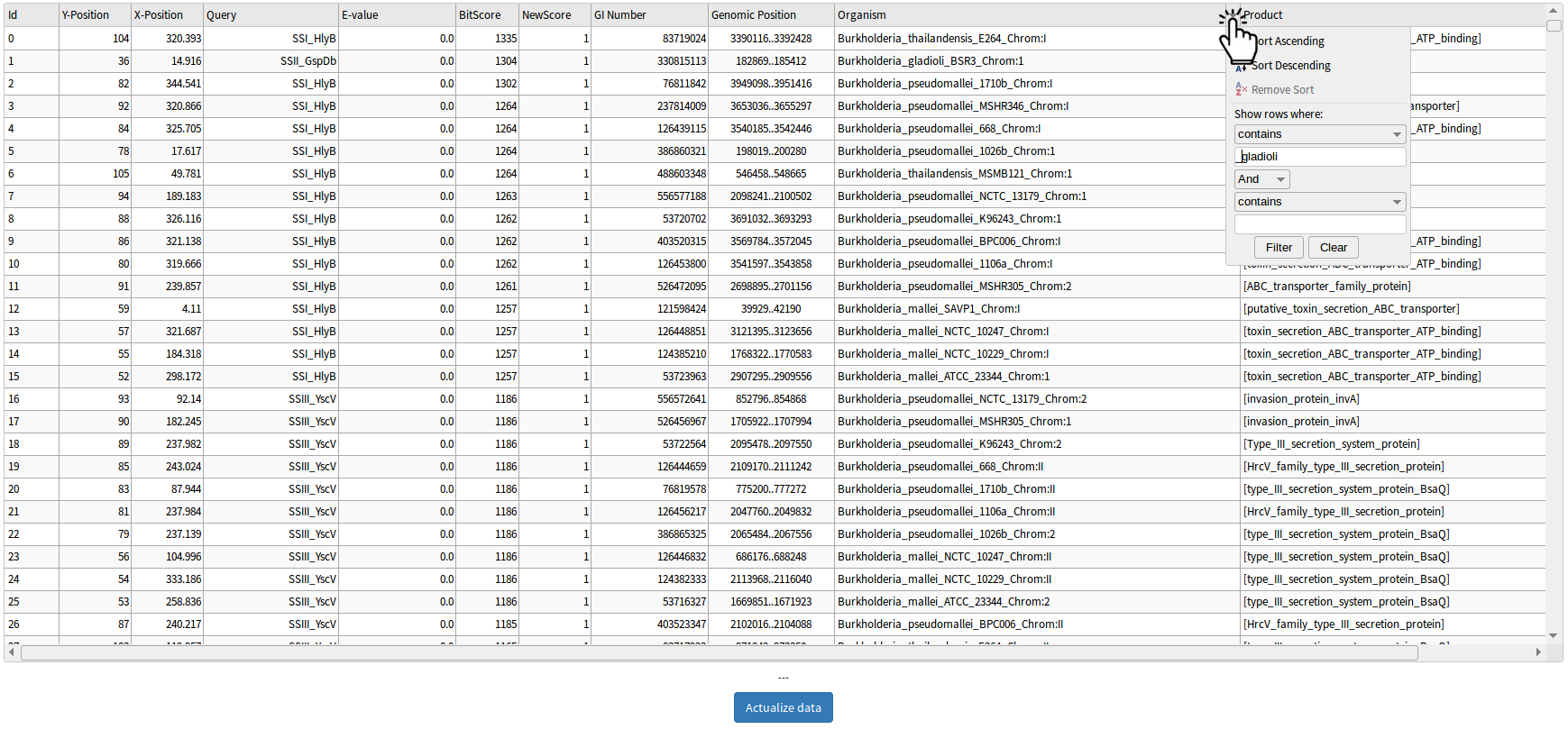
Evalue: 1e-14

MaxResults: 50000

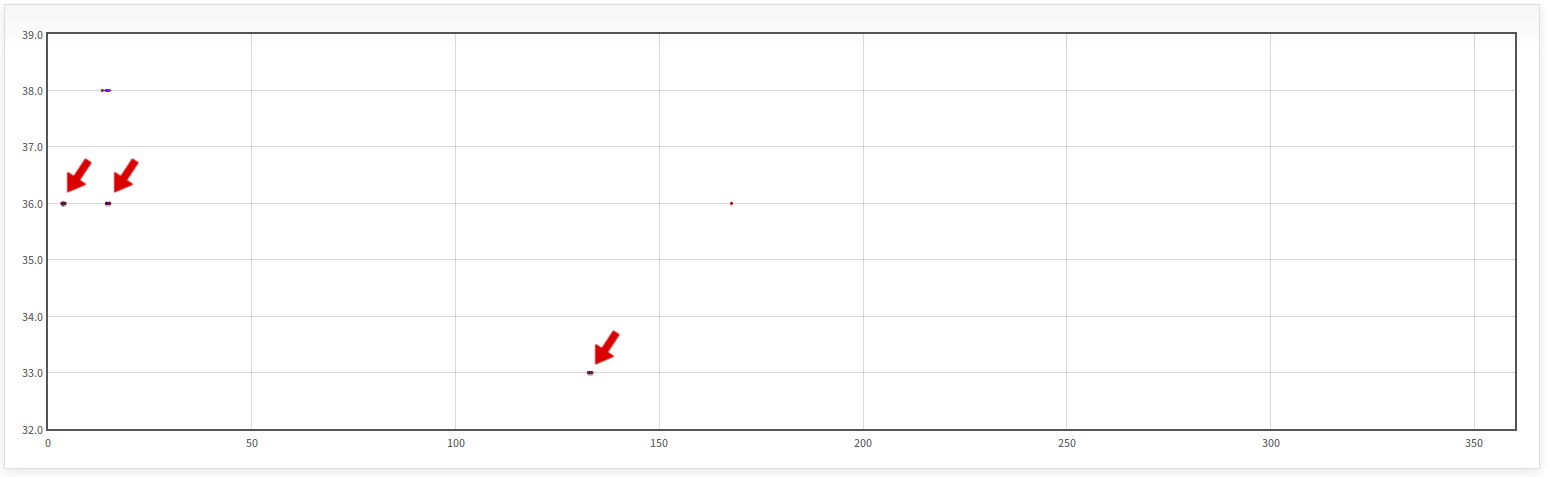
The query sequence file included one representative sequence for each protein involved in the formation of each secretion system (with a total of 84 sequences). To facilitate the analysis of results, the headers of all sequences included a meaningful tag (e.g. >SSI\_HlyD, >SSI\_TolC..., >SSII\_GspC, >SSII\_GspE...,) to filter them in the *Table* of data. The running time for this example was around 5 minutes. The plot for results shows clusters for different secretion systems. In this example, the large arrows correspond to BLAST-results of proteins located in plasmids.



To browse through massive results generated by large jobs like in this example, it might be more convenient to filter them by specific criteria using the *Table*. For example, to see the distribution of secretion systems in a particular organism, e.g. *Burkholderia gladioli*, data can be filtered by typing “Burkholderia\_gladioli” (or simply “gladioli”) in the drop down menu of the *Organism* column, and click on “Actualize data” button to update the plot. When browsing the results it would possible to detect different secretion systems present in *B. gladioli*. Similarly, by typing “SSII\_” in the filter menu for the *Query* column of the *Table*, and “gladioli” in the *Organism* column we can see the distribution only of the Secretion System type II in the genome of *B. gladioli*.



Once data have been filtered in the *Table*, the *Graph* will display the filtered results (indicated by red arrows) by clicking on the “refresh” button of the web browser.



A more extensive search of secretion systems could be performed in all bacterial genomes, and data can be filtered easily in the *Table* following the same criteria described above.

**Configuration for the search of bacteriocins.**

Bacteriocins synthesis usually involves different kind of proteins needed as precursor, like peptide maturation/modification enzymes, exportation machinery and self-immunity. As a consequence, the search for them in sequenced genomes turns laborious. To search for bacteriocins in a genus of interest (e.g. *Burkholderia*) parameters could be set as follows:

Job title: Search for Bacteriocins in Burkholderia genomes

Sequence File: All-bacteriocins-and-related.sec

Subject Data Base: Genus

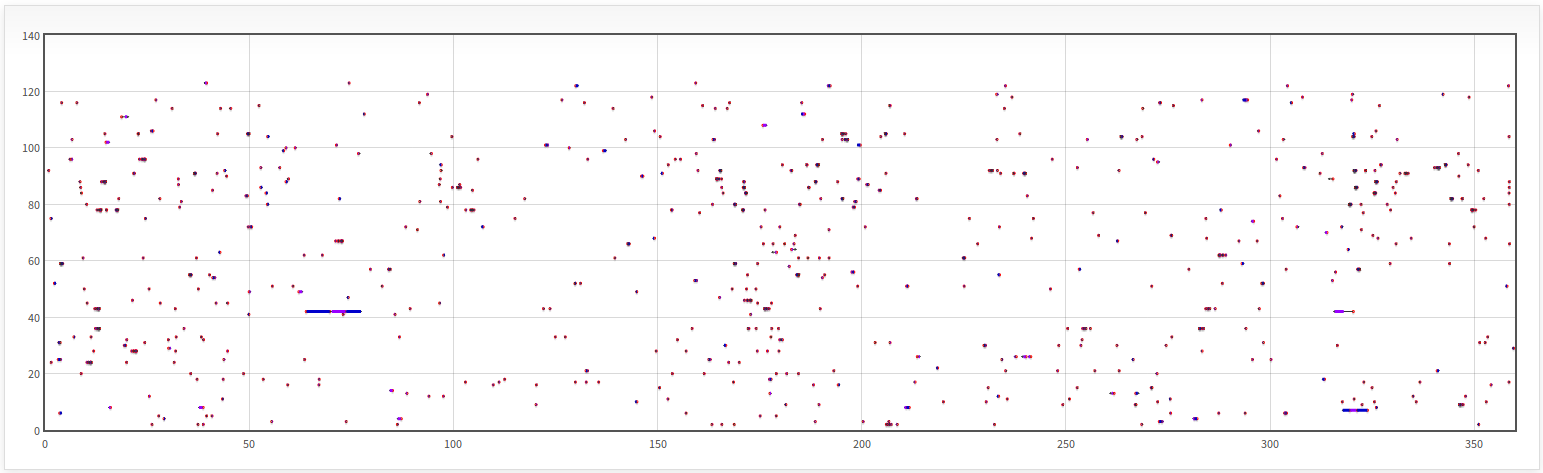
Genus: Burkholderia

BlastType: Proteins

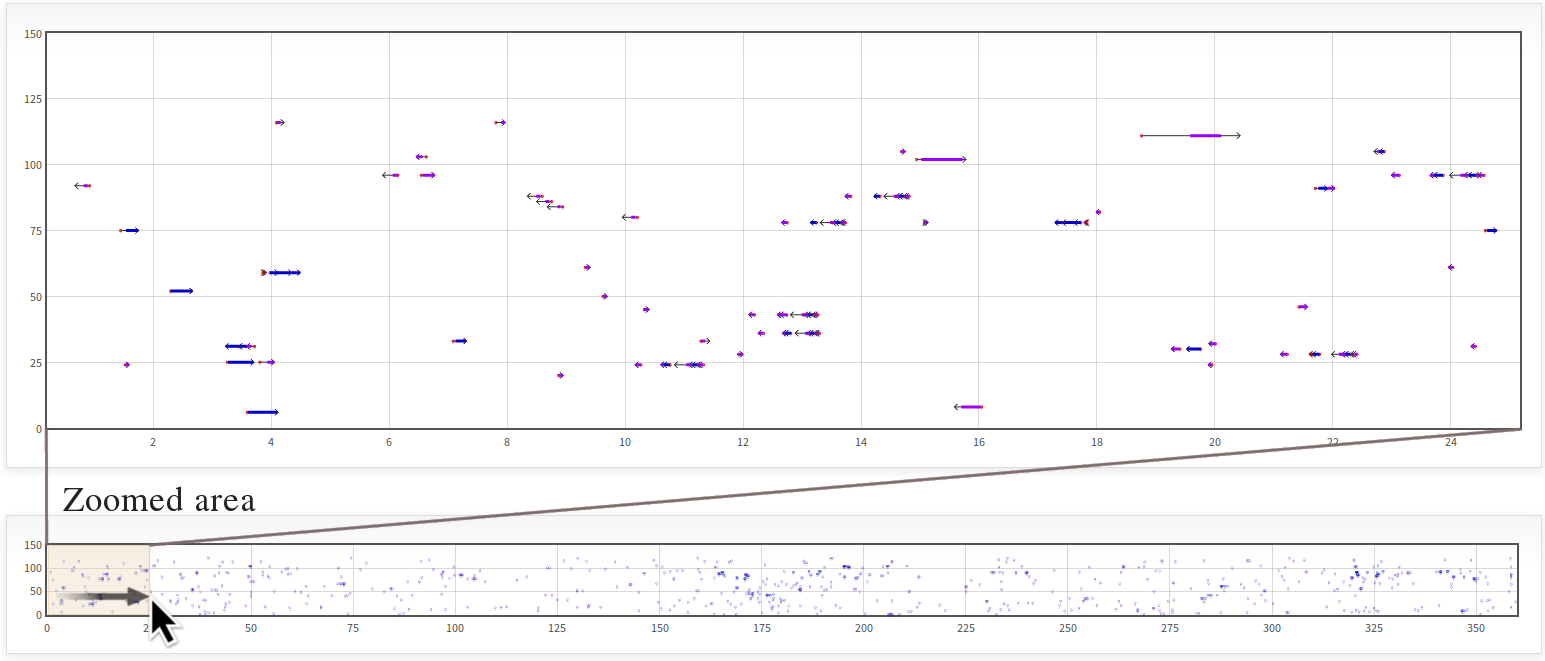
Evalue: 1e-18

MaxResults: 50000

In this example, the *sequence file* includes more than 2700 protein sequences downloaded from different databases available on the web [15-21]. The running time for this example was around twenty minutes.



To browse through the results, a zoom can be applied by dragging the mouse in the small plot below.

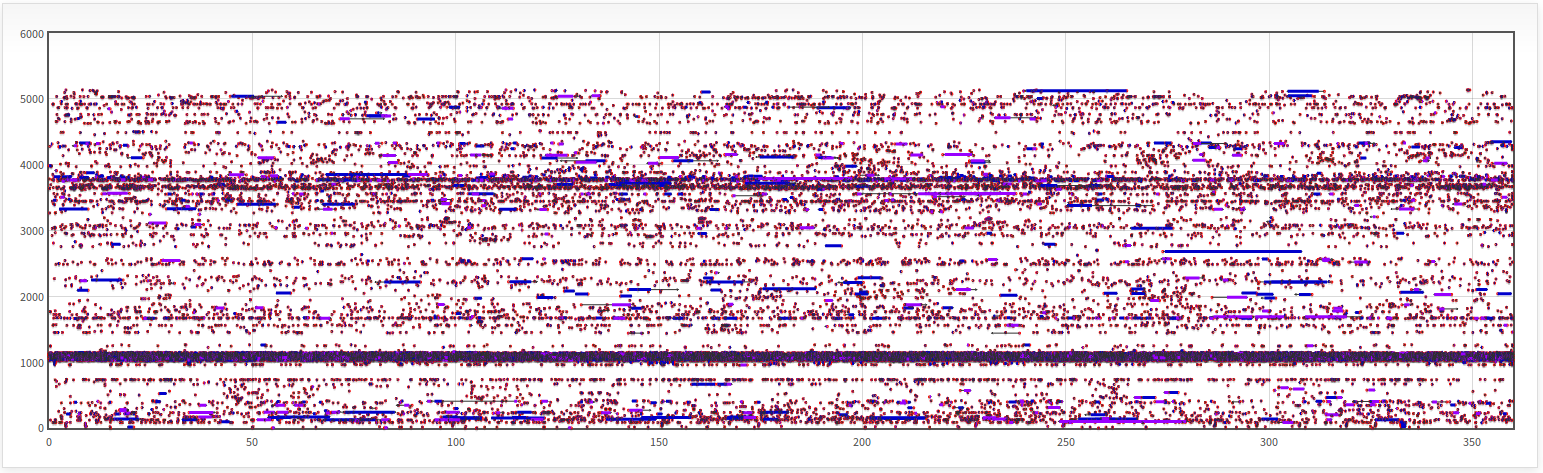


**Configuration of search for many sequences in multiple genomes**

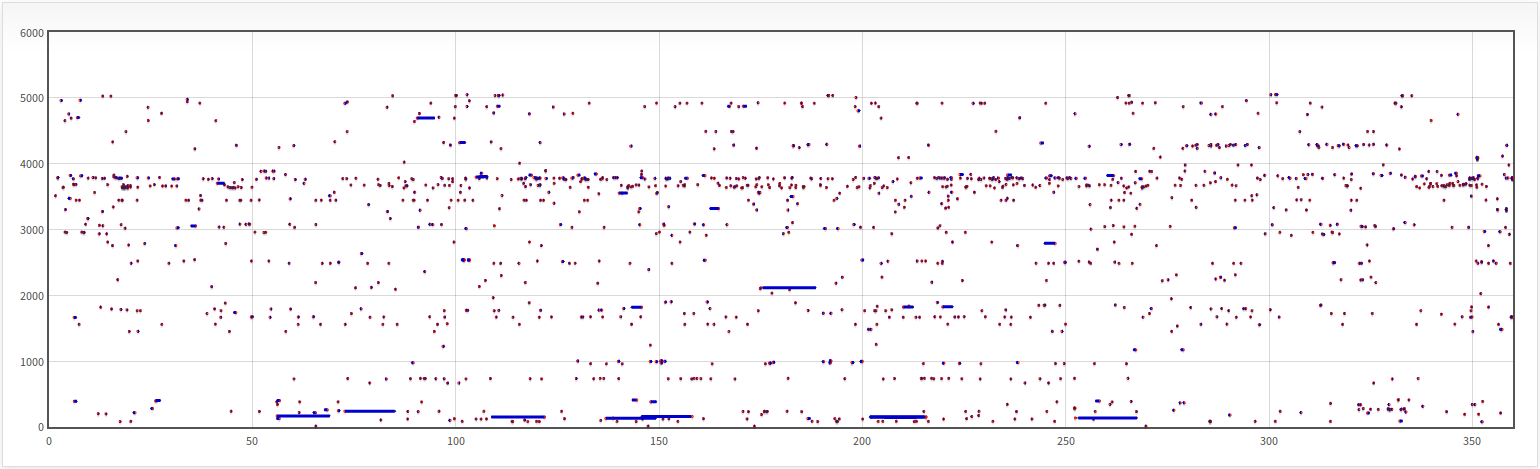
For this example we used all protein sequences annotated as “hypothetical” proteins in *Burkholderia gladioli* strain BSR3 (Total of 1967 sequences) and performed a comparison with all sequenced genomes in the database in order to find possible homologous sequences with functional annotations. An e-mail address was provided because it is expected that this job will take several hours (actual run time was 18 hours). Parameters were set as:

Job Title: B gladioli hypothetical proteins in all genomes   
Sequences file: B\_gladioli\_hypothetical\_proteins.sec  
Subject DB: All bacterial and archaea  
Blast Type: Proteins  
Cut off E-value: 1e-30  
Results Limit: 50000

The main plot shows the distribution of 43917 results.



Results must be filtered to remove hits with the self-sequence of *B. gladioli* and close homologous in others *Burkholderia*, those that match with other “hypothetical” proteins, and non-relevant hits. To filter data on the table with these tree criteria, type “Burkholderia” on the “Organism” column drop down filter menu and select the “does not contain” option. Type “hypothetical” on the “Product” column’s menu and select the “does not contain” option. Finally, type “0.0” on the “E-value” option and select the “equal” option. Click on the “Actualize data” button and refresh the plot web page.



The *Graph* and the *Table* now include BLAST-results only for those proteins with a functional annotation that match with high scores with hypothetical proteins of *B. gladioli*.