9th CCIC
8.18-8.22

Prokaryotic resistance gene search system database

Team:SYSU-Software

1.Background

> Prokaryotes have been fighting against viruses for billions of years with resistance genes like CRISPR-Cas system as their weapons.
> Interestingly, defense island, which means resistance genes that encode different defense systems are distributed in clusters like islands in oceans, hidden in their genomes.
> The flanking region of defense island often are house-keeping genes, and at least one known resistance gene will be included.

Fig.1 Defense Island

2.Model

Select loci of interest
↓
Select neighbourhood around seeds
↓
Cluster proteins with permissive parameters
↓
Search for proteins relative to clusters
↓
Calculate relevence metrics for each cluster
↓
Search for proteins instances relative to seeds
↓
Manual curation of the candidates

DS3P

3.Workflow

All steps are based on protein sequences.

>Stage 1: Select Baits

Baits are genes that you have interest in(like bacterial's resistance genes).

The content should include locus ID, ORFStart:End, etc

>Stage 2: Select the neighbourhood around Baits

Different systems have different gene retrieval lengths (like CRISPR-Cas 10~20 kbp,TA 1~2 kbp)

>Stage 3: Search for clusters

With MMSeq2 and PSIBLAST running for 3~5 times, we can get high quality clusters.

>Stage 4: Identify Hits

Hits are protein relative to clusters. Using PSIBLAST we can get a data set with both sensitivity and specificity

>Stage 5: Rank by relevance

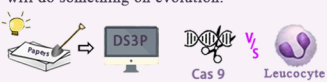
Using metrics like 'DS3P' to evaluate the possibility of Hits being members of defense system.

>Stage 6: Final Screen

With pre-set parameters and manual curation, we can finally dig out which genes can be a new member of defense system.

4.Meaning

> We dig the data hidden in papers and build a platform to give existing databases a update and extension.
> Our platform can enhance the understanding of defense system in prokaryotes and through prediction we may find new gene editing tools.
> Forward: We want to find out whether relationships between defense system in prokaryotes and immune system in eukaryotes exist and address this question will do something on evolution.



Reference:

Shmakov, S.A., Faure, G., Makarova, K.S. et al. Systematic prediction of functionally linked genes in bacterial and archaeal genomes. Nat Protoc 14, 3013–3031 (2019).

Contact us
sysu_software2022@163.com