

CRISPRCasFinder



Funcionamiento básico

INPUT

Recibe archivos en formato FASTA (incluso multi-FASTA)

1

Detección y validación de sistemas CRISPR

Una estructura CRISPR es una sucesión de repeticiones máximas separadas por espaciadores. El programa utiliza esta propiedad para encontrar posibles localizaciones.

- Usa un paquete basado en Vmatch para detectar las repeticiones máximas bajo parámetros por defecto.
- Se emplean filtros para validar el sistema CRISPR:
 - Tamaño del espaciador comparado con el tamaño de la repetición (usa parámetros por defecto).
 - Los espaciadores no tienen que ser idénticos (elimina repeticiones en tandem).

Clasificación

2

Nivel de clasificación: analiza tanto los espaciadores como las repeticiones.

EBcon	Identidad porcentual de espaciadores	NIVEL
<70		2
>=70	>8%	3
>=70	<=8%	4

• EBcon (Entropy- based conservation) es un algoritmo que mide la conservación de las **repeticiones**.

• La identidad porcentual de los **espaciadores** habla de qué tan parecidos son unos con otros (no se espera que los espaciadores tengan similitud).

3

Identificación de la orientación

Hay dos indicadores para ello:

- Predicción con CRISPRDirection (base de datos con repeticiones consenso).
- Uso del %AT en 100pb de cada extremo, generalmente el extremo 5' es rico en AT (la ventaja de este indicador es que no requiere sistemas homólogos para comparar).

Búsqueda de genes Cas

4

Se buscan 'genes putativos' o candidatos:

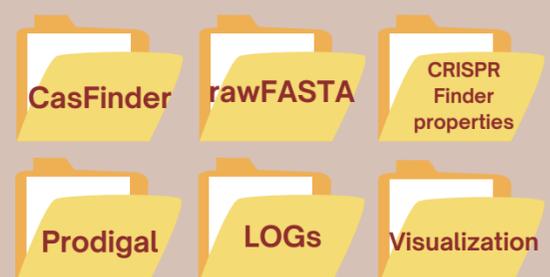
- El paquete Prodigal extrae CDSs (regiones codificantes) y CasFinder (basado en MacSyFinder) analiza CDSs traducidos, los compara con una base de datos empleando HMM (*Hidden Markov Model*).
- Hay tres niveles de precisión: general, typing, subtyping.

Archivos y carpetas principales



OUTPUT

Carpetas secundarias



Referencias

- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46(W1), W246–W251. <https://doi.org/10.1093/nar/gky425>
- CRISPRCas++ en <https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>