



UNIVERSIDAD DE
COSTA RICA

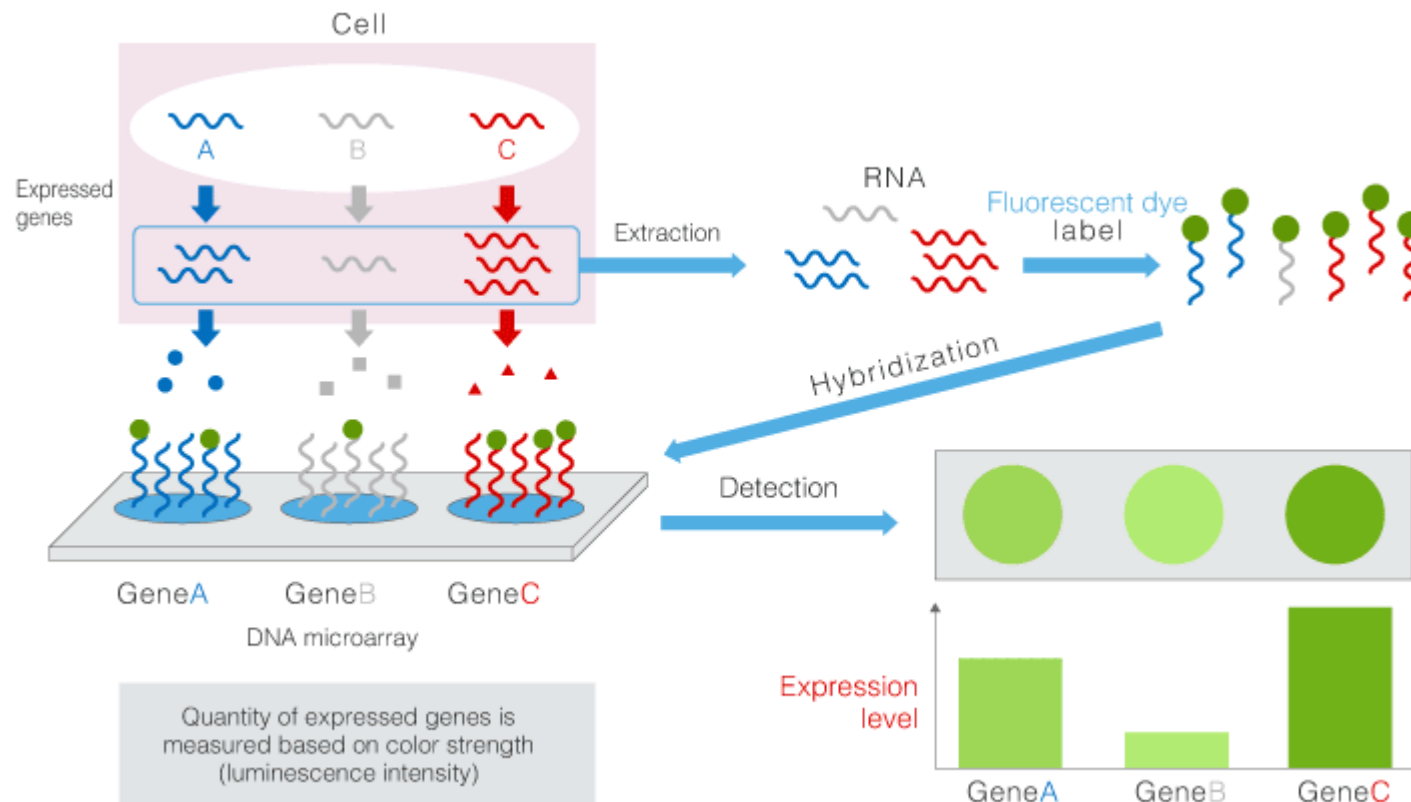
Genómica comparativa

Curso Genómica de procariontes

Bradd Mendoza

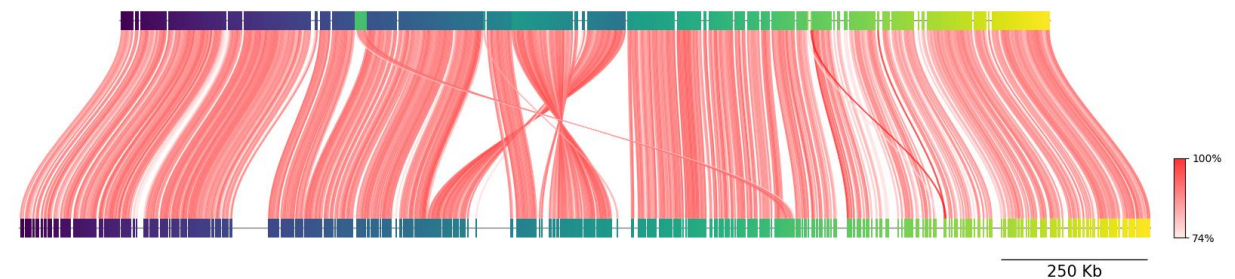
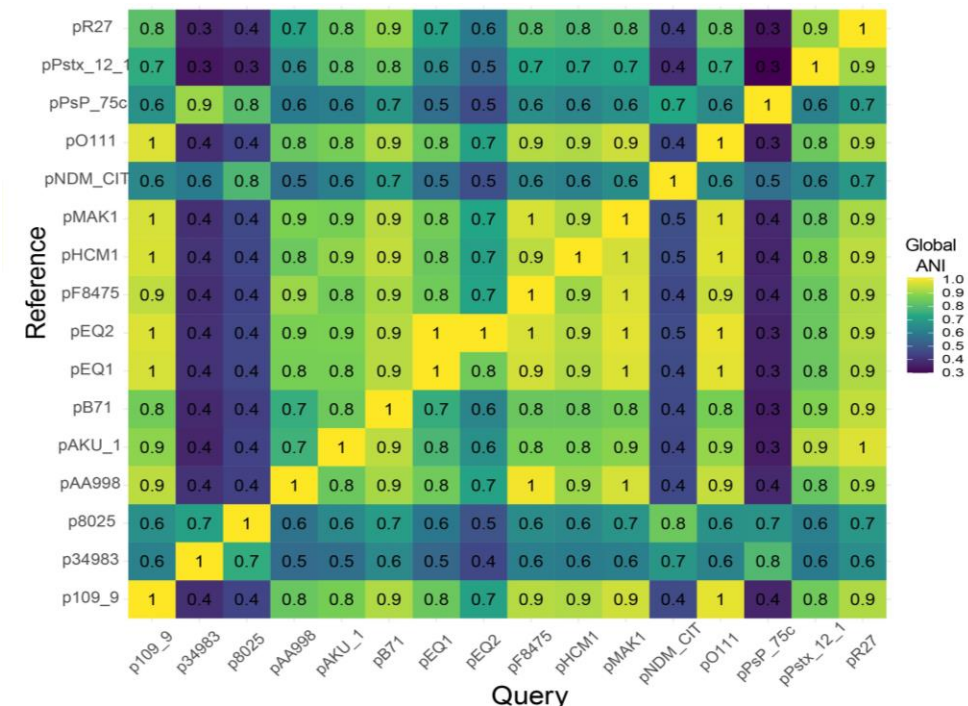
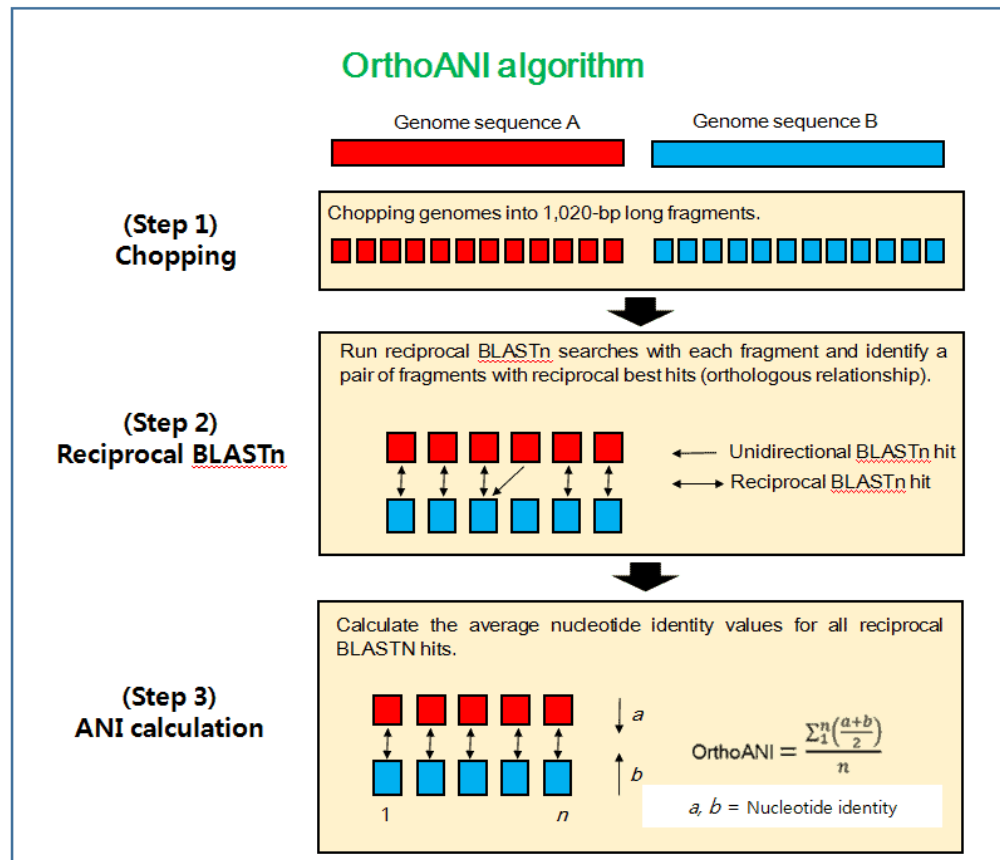
Algunos conceptos importantes

- DNA-DNA hybridization





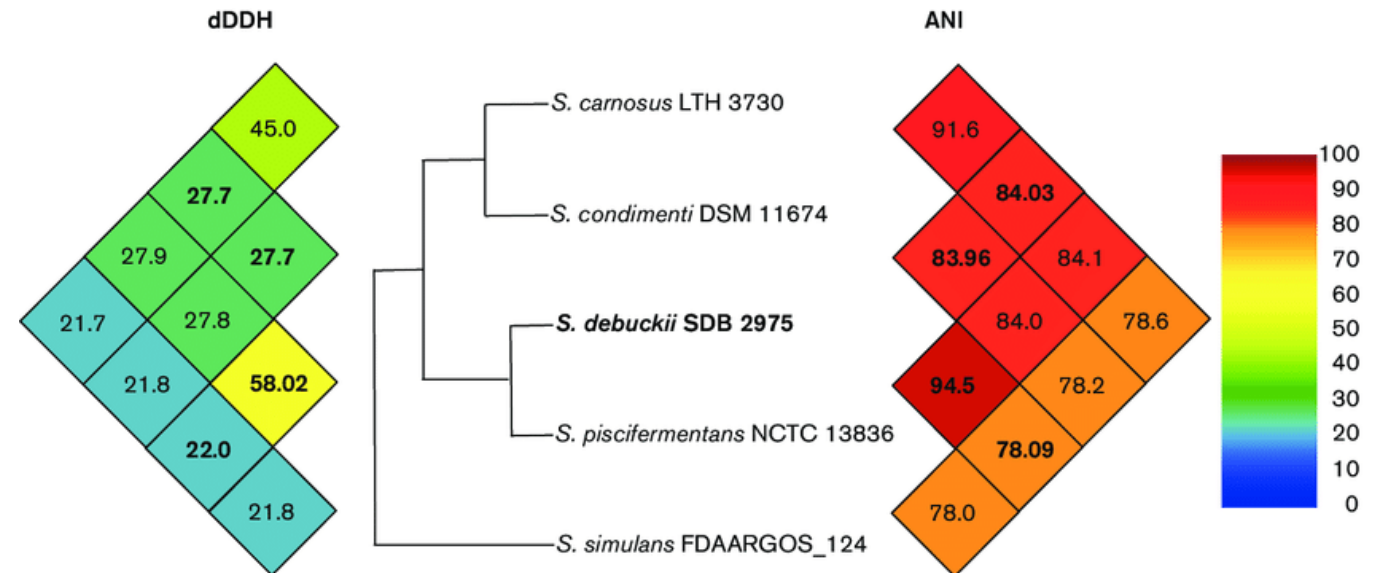
ANI



Delimitadores de especies

dDDH menor o igual a 70% es el delimitador para especie y 80% para subespecies.

El ANI varía dependiendo de los grupos pero ronda entre 95-96% para especies, no hay uno para subespecies.





Conceptos

- Gene cluster: Son grupos de genes asociados por su homogeneidad de secuencias. Se asocian con funciones similares. En un pangenoma se crean clúster con genes de todos los genomas individuales.
- Homogeneidad: Qué tan similares son esos genes dentro de un gene clúster



- Homogeneidad geométrica: Se basa principalmente en la calidad del alineamiento. Se ve afectada por la presencia de gaps y residuos.

MIT9202	449	-----MYLLARDC-----LRGFTKIGNMITTVKRHIGATTINIDSRDIKIFEFINK
MIT9215	1585	-----MYLLARDC-----LRGFTKIGKTITTVKRHIGATTINIDSRDIKNFEFINK
MIT9302		
GP2		
MIT9201	321	MYLVFSVPLSFFLFEAYLIQRIYLF EISVLLRGLTKIGRTITTVKRQIGATTINIDSNM-----IKS
MIT9314		
MIT9301		
AS9601	1544	-----MYLLGKDC-----LRGFTKIGKTITTVKRQIGATTINIDSRDINNFEFINK



- Homogeneidad funcional: Este se basa en los residuos de aminoácidos asignados a la secuencia, ignora totalmente los gaps. Lo importante es qué tan conservada sea la función de los aminoácidos.

Amino acids				
Hydrophobic amino acids non-polar		Hydrophilic amino acids polar		
R = alkyl	R = aromatic group	Neutral	Acidic	Basic
$\text{H}_3\text{N}^+-\text{CH}(\text{H})-\text{COO}^-$ Gly	$\text{H}_3\text{N}^+-\text{CH}(\text{H}_2\text{C}-\text{C}_6\text{H}_5)-\text{COO}^-$ Phe	$\text{H}_3\text{N}^+-\text{CH}(\text{H}_2\text{C}-\text{C}_6\text{H}_4-\text{OH})-\text{COO}^-$ Phe	$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{COO}^-)-\text{COO}^-$ Glu	$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{NH}_3^+)-\text{COO}^-$ Lys
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_3)-\text{COO}^-$ Ala	$\text{H}_3\text{N}^+-\text{CH}(\text{H}_2\text{C}-\text{C}_6\text{H}_5)-\text{COO}^-$ Phe	$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2\text{OH})-\text{COO}^-$ Ser	$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{COO}^-)-\text{COO}^-$ Asp	$\text{H}_3\text{N}^+-\text{CH}(\text{H}_2\text{C}-\text{C}_6\text{H}_4-\text{NH}_2)-\text{COO}^-$ His
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_3)-\text{COO}^-$ Val	$\text{H}_3\text{N}^+-\text{CH}(\text{H}_2\text{C}-\text{C}_6\text{H}_4-\text{NH}_2)-\text{COO}^-$ Trp	$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{OH})-\text{COO}^-$ Thr		$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{NH}_2)-\text{COO}^-$ Arg
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CH}_3)-\text{COO}^-$ Leu		$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{SH})-\text{COO}^-$ Cys		
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3)-\text{COO}^-$ Ile		$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CONH}_2)-\text{COO}^-$ Gln		
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3)-\text{COO}^-$ Met		$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CONH}_2)-\text{COO}^-$ Asn		
$\text{H}_3\text{N}^+-\text{CH}(\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3)-\text{COO}^-$ Pro				

PC_00001039 with 21 genes detailed

<<< prev | 7314 of 7385 | next >>>

Wrap: 140

Font Size: 12

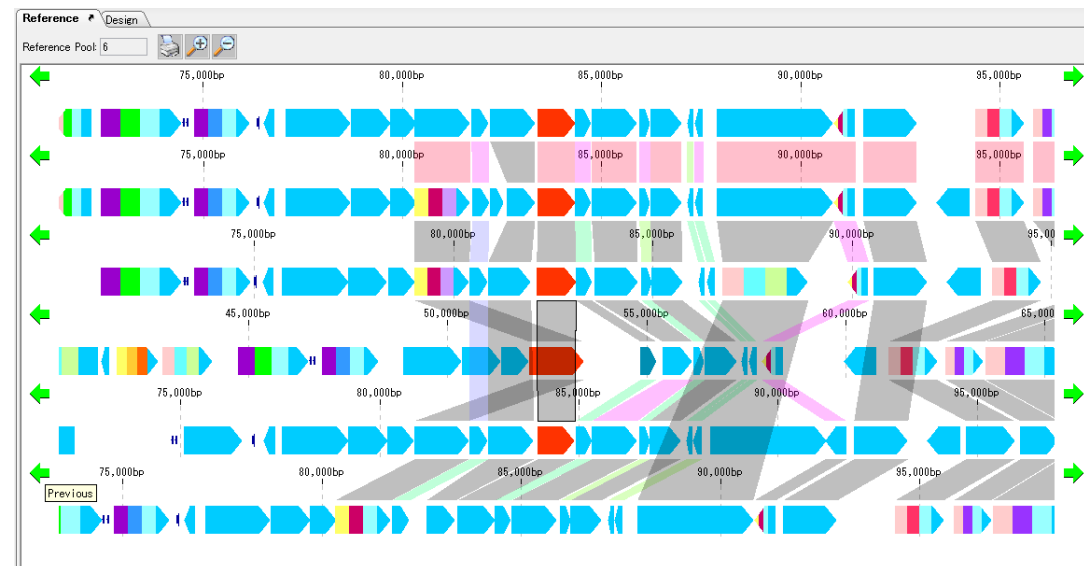
Color Settings

MIT9515
MED4
EQPAC1
MIT9107
MIT9123
MIT9116
MIT9321
MIT9322
MIT9401
MIT9312
MIT9311
MIT9215
MIT9202
MIT9302
GP2

45716 MSKYLLEIGTEEL ANFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
10072 MSKYLLEIGTEELSKFSHSVINOINSLEIEFDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
4988 MSKYLLEIGTEELSKFSHSVINOINSLEIEFDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
12567 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
16638 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
14441 -----ISLSLIIKVIKSLQ
13732 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
39587 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
42580 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
44442 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
33654 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
32601 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
23308 MSKYLLEIGTEELANFSHYVLNOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
20460 MSKYLLEIGTEELANFSHYVLNOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
26665 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ
6325 MSKYLLEIGTEELAKFSHSVINOIYSNIKFEDKKLIKFNIICTSTRRIVLFLEGLVYADDKRVTRKQKASAAFLNQTQAALGFANSLGINIEQLIRKTEKGFVFGTKIEKGESTKSALSLIIKLIKNLQ

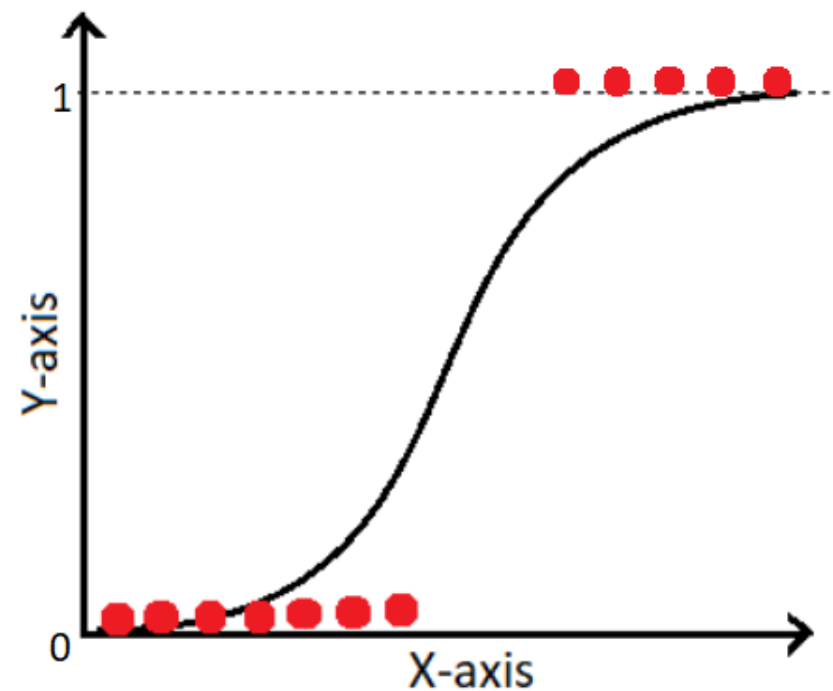
Análisis de enriquecimiento

- No todos los genes de un mismo gene clúster se asociarán con la misma función.
- Anvio asigna la función de un gene clúster basado en la función asignada a la mayoría de los genes de ese clúster.



Análisis de enriquecimiento

- Asigna una función “asociada” a cada gene clúster y determina si esta función está enriquecida en un grupo de genomas.
- No se recomienda utilizar si se tiene menos de 8 genomas de un mismo grupo.
- La asignación se basa en un GLM (distribución logística)





Genómica comparativa

