

# **Estudio del resistoma del compost: identificación de genes bacterianos resistentes a antibióticos**

**Aritz Durana**

**Master Universitario en Bioinformática y Bioestadística UOC-UB**  
**Área 4, Subárea 6: Microbiología, biotecnología y biología molecular**  
**Consultora: Paloma Pizarro Tobias**

Enero 2021

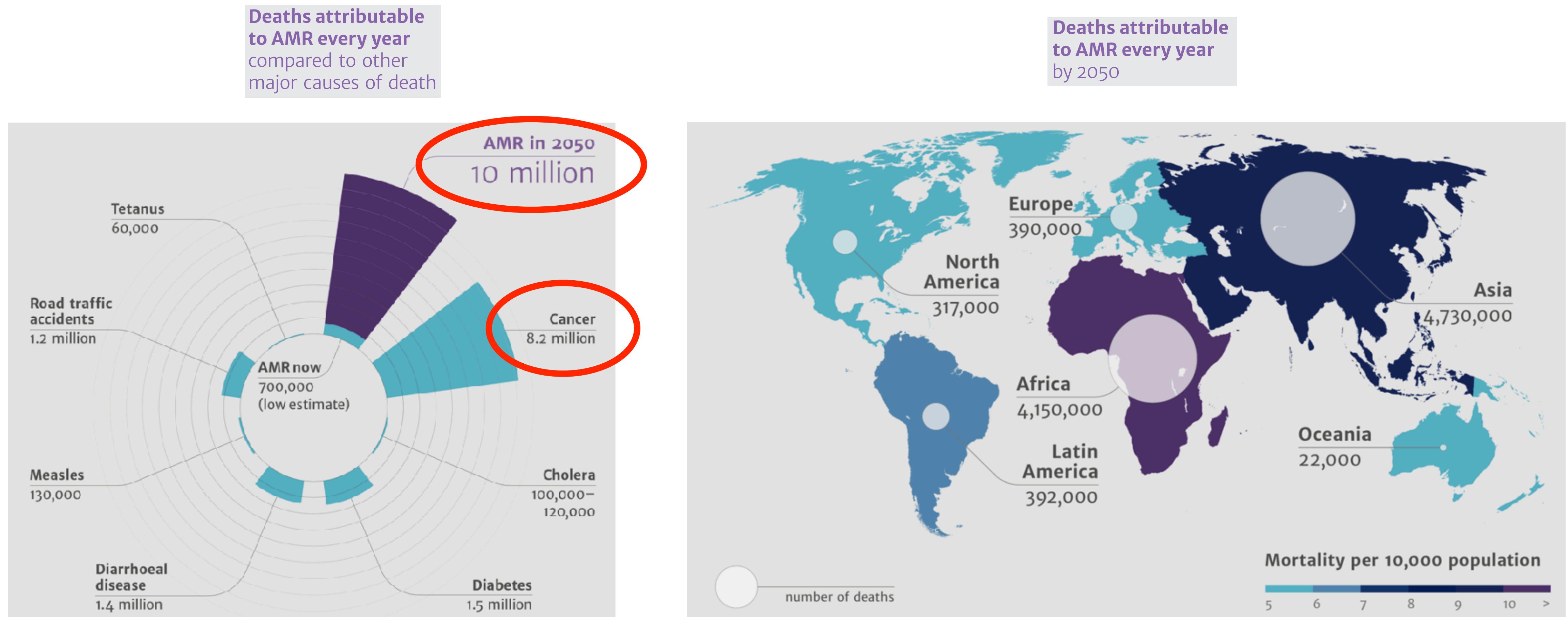
**Resistoma:** Colección de genes resistentes a antibióticos presentes en la comunidad bacteriana de una muestra medioambiental

**ARG:** Gen resistente a antibiótico (Antibiotic Resistant Gene)

**Compost:** Producto de un proceso basado en la descomposición aeróbica de la materia orgánica que ocurre espontáneamente bajo ciertas condiciones ambientales

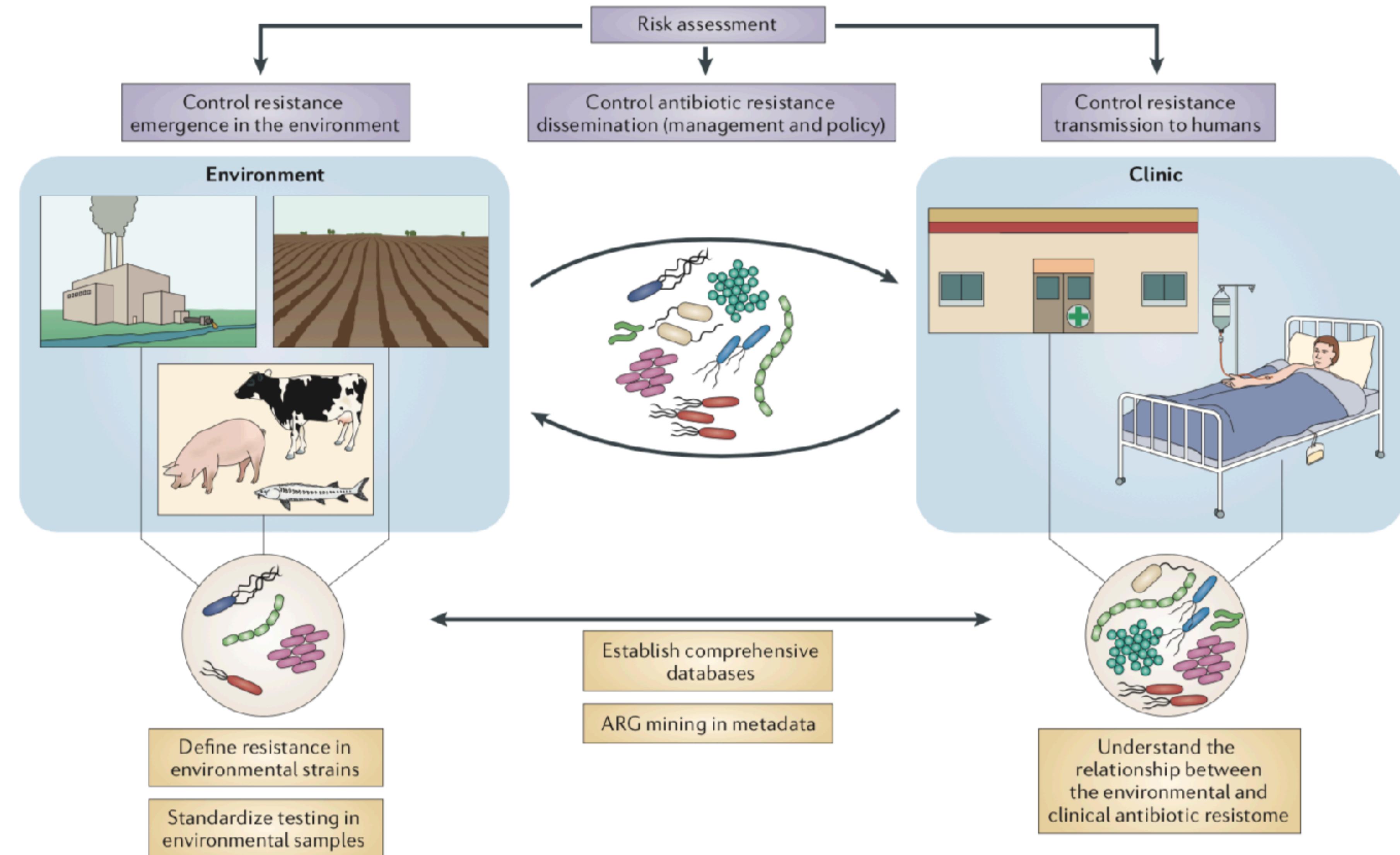
**Metagenoma:** Conjunto de genes de una muestra medioambiental concreta

# Introducción - ARGs son un problema sanitario



O'Neill, J. Review on Antimicrobial Resistance: Tackling A Crisis for the Health and Wealth of Nations. 2014. HM Government, London, 2016.  
<https://wellcomecollection.org/works/rdpck35v>

# Introducción - Control de la proliferación de ARGs



- NGS engloba cualquier técnica de alto rendimiento para la secuenciación de ADN que utilice la secuenciación masiva en paralelo
- Es posible secuenciar genomas en un corto espacio de tiempo y con un coste económico reducido
- En la secuenciación se generan muchos millones de secuencias muy cortas (reads), que posteriormente se alinean mediante herramientas bioinformáticas para lograr los genomas presentes en la muestra
- En este TFM se utilizarán metagenomas de muestras de compost para identificar los ARGs presentes en los mismos.

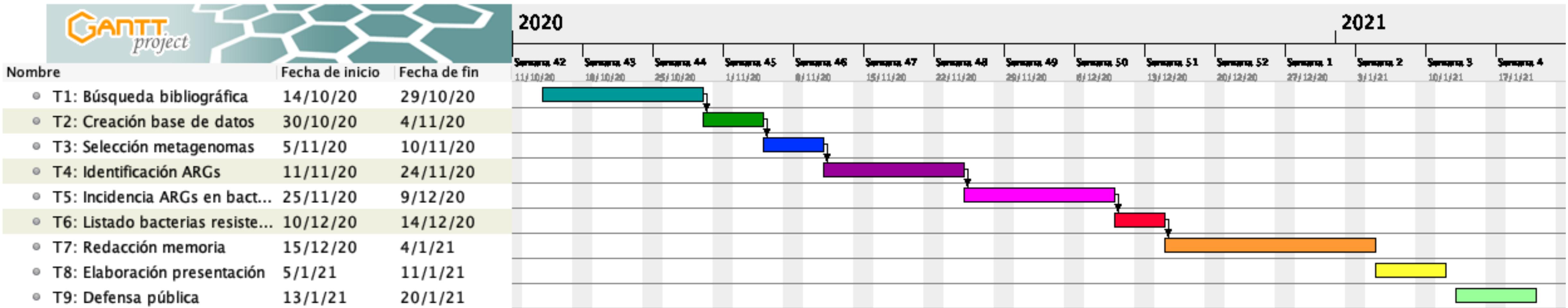
## ***Objetivos generales***

- Identificar genes bacterianos resistentes a antibiótico (ARGs) presentes en muestras de compost
- Definir la incidencia de los ARGs en la comunidad microbiana del compost
- Estudiar la relación entre los ARGs del compost y la presencia de bacterias resistentes a antibiótico en el medioambiente

## ***Objetivos específicos***

- Generar una base de datos de ARGs presentes en el compost
- Seleccionar datos de secuenciación masiva de muestras de compost y comparar los metagenomas obtenidos con la base de datos de ARGs
- Identificar las bacterias resistentes más abundantes en los metagenomas seleccionados

# Introducción - Cronograma



## TAREAS:

- T1: Realizar una búsqueda bibliográfica de ARGs en artículos científicos y en bases de datos
- T2: Generar una base de datos de ARGs presentes en el compost
- T3: Seleccionar metagenomas de muestras de compost en bases de datos de secuenciación masiva
- T4: Identificar los ARGs presentes en los metagenomas
- T5: Observar la incidencia de los ARGs en la distribución de la comunidad microbiana de las muestras
- T6: Generar un listado de las bacterias resistentes más abundantes en los metagenomas

# Resultados - Búsqueda bibliográfica (T1)

Wang et al. *Microbiome* (2017) 5:103  
 DOI 10.1186/s40168-017-0324-0

**RESEARCH**

**Microbial phylogeny determines transcriptional response of resistome to dynamic com**

Cheng Wang<sup>1†</sup>, Da Dong<sup>2†</sup>, P.



**Microbiome**

**Open Access**



*Journal of Hazardous Materials* 389 (2020) 122082

Contents lists available at ScienceDirect

*Journal of Hazardous Materials*

**npj Clean Water**

**REVIEW ARTICLE OPEN**

**Antibiotic resistance genes from livestock waste: occurrence, dissemination and transfer**

*Environment International* 136 (2020) 105458

Contents lists available at ScienceDirect

*Environment International*

journal homepage: [www.elsevier.com/locate/envint](http://www.elsevier.com/locate/envint)



## MÁS DE 100 ARTÍCULOS REVISADOS

Changes in microbial community structure during pig manure composting and its relationship to the fate of antibiotics and antibiotic resistance genes

**Diverse and abundant antibiotic resistance genes in Chinese swine farms**

Yong-Guan Zhu<sup>a,b,1,2</sup>, Timothy A. Johnson<sup>c,d,1</sup>, Jian-Qiang Su<sup>a</sup>, Min Qiao<sup>b</sup>, Guang-Xia Guo<sup>b</sup>, Robert D. Stedtfeld<sup>c,e</sup>, Syed A. Hashsham<sup>c,f</sup>, and James M. Tiedje<sup>c,d,2</sup>



<sup>a</sup>Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China; <sup>b</sup>Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China; and <sup>c</sup>Center for Microbial Ecology, Departments of <sup>d</sup>Plant, Soil and Microbial Sciences, and <sup>e</sup>Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824

Contributed by James M. Tiedje, December 31, 2012 (sent for review October 31, 2012)

Discharge of swine wastes risks water quality and food safety: Antibiotics and antibiotic resistance genes from swine sources to the receiving environments

Liang-Ying He, Guang-Guo Ying <sup>\*</sup>, You-Sheng Liu, Hao-Chang Su, Jun Chen, Shuang-Shuang Liu, Jian-Liang Zhao

<sup>a</sup>State Key Laboratory of Organic Geochemistry, CAS Centre for Pearl River Delta Environmental Pollution and Control Research, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China



**RESEARCH ARTICLE**

**Resistome metagenomics from plate to farm: The resistome and microbial composition during food waste feeding and composting on a Vermont poultry farm**

Korin Eckstrom<sup>①</sup>, John W. Barlow<sup>②\*</sup>

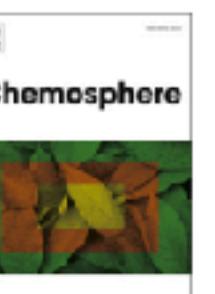
<sup>1</sup> Department of Microbiology and Molecular Genetics, The University of Vermont, Burlington, Vermont, United States of America, <sup>2</sup> Department of Animal and Veterinary Sciences, The University of Vermont, Burlington, Vermont, United States of America

\* [john.barlow@uvm.edu](mailto:john.barlow@uvm.edu)



Jun Chen<sup>a,b</sup>,

<sup>a</sup>State Key Laboratory of Theoretical



**the antibiotic resistance genes (ARGs) by the addition of clay during poultry manure composting**

Mukesh Kumar Awasthi <sup>a, \*\*, 1</sup>, Hongyu Chen <sup>a, 1</sup>, Sanjeev Kumar Awasthi <sup>a</sup>, Yumin Duan <sup>a</sup>, Tao Liu <sup>a</sup>, Ashok Pandey <sup>b</sup>, Sunita Varjani <sup>c</sup>, Zengqiang Zhang <sup>a, \*</sup>

<sup>a</sup> College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi Province, 712100, China

<sup>b</sup> CSIR-Indian Institute of Toxicology Research, Vishvigan Bhawan, 31 Mahatma Gandhi Marg, Lucknow, 226001, Uttar Pradesh, India

<sup>c</sup> Gujarat Pollution Control Board, Sector-10A, Gandhinagar, 382010, Gujarat, India

# Resultados - Base de datos (T2)

A	B	C	D	E	F	G	H	I	J	K	
1	Nombre del gen	Nombres alternativos	Clase	Antibiotico	Mecanismo	Database	Organismo	RefSeq Sequence	Genbank Accession	CARD Accession	Enlace
2	aac(3)-II		Aminoglycoside	Gentamicin	Inactivation	NCBI	Sphingopyxis alaskensis RB2256	NC_018573.1	CP000356.1		<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
3	aac(3)-IV	aacC4, AAC(3)-IVa	Aminoglycoside	Apramycin, Gentamicin, Tobramycin	Inactivation	CARD	Escherichia coli	NG_047253.1	DQ241380.1	ARO:3002539	<a href="https://www.card.mrc.ac.uk">https://www.card.mrc.ac.uk</a>
4	aac(6')-31		Aminoglycoside	Aminoglycoside	Inactivation	CARD	Pseudomonas putida		AM283489.1	ARO:3002585	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
5	aac(6')-Ia	aacA1, aac(6')-I	Aminoglycoside	Aminoglycoside	Inactivation	CARD	Plasmid R		M18967.1	ARO:3002545	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
6	aac(6')-Ib	aacA4, AAC(6')-4	Aminoglycoside	Aminoglycoside	Inactivation	CARD	Klebsiella pneumoniae		JQ808129.1	ARO:3002546	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
7	aac(6')-Ib-cr		Aminoglycoside	Aminoglycoside, Fluoroquinolone	Inactivation	CARD	Escherichia coli		DQ303918.1	ARO:3002547	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
8	aac(6')-Ie-aph(2")-Ia	aac(6')-bifunctional aa	Aminoglycoside	Amikacin, Kanamycin, Tobramycin	Inactivation	CARD	Staphylococcus aureus	NG_047055.1	GU565967.1	ARO:3002597	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
9	aac(6')-Ii		Aminoglycoside	Aminoglycoside	Inactivation	CARD	Enterococcus faecium	NG_047298.1	L12710.1	ARO:3002556	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
10	aac(6')-Ila		Aminoglycoside	Aminoglycoside	Inactivation	CARD	Salmonella enterica subsp. enterica serovar Typhi		AY123251.1	ARO:3002594	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
11	aac(6')-lid		Aminoglycoside	Aminoglycoside	Inactivation	CARD	Enterococcus hirae	NG_047299.1	AJ584700.2	ARO:3002589	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
12	aac(3)-Ia	aacC-A1, aacC1	Aminoglycoside	Gentamicin	Inactivation	CARD	Pseudomonas aeruginosa	NG_056001.1	U12338.3	ARO:3002528	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
13	aac(3)-Ic	aacC2	Aminoglycoside	Gentamicin	Inactivation	CARD	Escherichia coli	NG_047250.1	X54723.1	ARO:3002535	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
14	aad(6)	ant(6)-Ia, aac(6)	Aminoglycoside	Streptomycin	Inactivation	CARD	Streptococcus oralis		AY712687.1	ARO:3002628	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
15	ANT(9)-Ia	aad(9), spc, spw, aad9	Aminoglycoside	Spectinomycin	Inactivation	CARD	Staphylococcus aureus	NG_047397.1	X02588.1	ARO:3002630	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
16	aadA	ANT(3')-Ia, aadA1-pm	Aminoglycoside	Streptomycin	Inactivation	CARD	Escherichia coli		AF550679.1	ARO:3002601	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
17	aadA2		Aminoglycoside	Streptomycin, Spectinomycin	Inactivation	CARD	Klebsiella pneumoniae		AF156486.1	ARO:3002602	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
18	aadA24		Aminoglycoside	Streptomycin, Spectinomycin	Inactivation	CARD	Salmonella enterica subsp. enterica serovar Newport		DQ677333.1	ARO:3002621	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
19	aadA3		Aminoglycoside	Streptomycin	Inactivation	CARD	Plasmid NR79	NG_047353.1	AF047479.2	ARO:3002603	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
20	aadA4		Aminoglycoside	Streptomycin	Inactivation	CARD	Acinetobacter baumannii	NG_047356.1	AY138986.1	ARO:3002604	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
21	aadA5		Aminoglycoside	Streptomycin, Spectinomycin	Inactivation	CARD	Escherichia coli	NG_047357.1	AF137361.1	ARO:3002605	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
22	aadA6		Aminoglycoside	Streptomycin	Inactivation	CARD	Pseudomonas aeruginosa		AM087411.1	ARO:3002606	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
23	aadA9	ANT(9)-Ia	Aminoglycoside	Streptomycin	Inactivation	CARD	Corynebacterium sp. L2-79-05		DQ390458.1	ARO:3002609	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
24	ANT(2")-Ia	aadB	Aminoglycoside	Gentamicin, Kanamycin, Tobramycin	Inactivation	CARD	Pseudomonas aeruginosa		AF078527.1	ARO:3000230	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
25	ANT(4')-Ia	aadD, aadD2	Aminoglycoside	Amikacin, Kanamycin, Tobramycin	Inactivation	CARD	Bacillus clausii	NG_047392.1	EF540343.1	ARO:3002623	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
26	ANT(6)-Ia	ant6, aadE	Aminoglycoside	Streptomycin	Inactivation	CARD	Exiguobacterium sp. S3-2		KF648874.1	ARO:3002626	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
27	abeM		Fluoroquinolone	Multidrug	Efflux	CARD	Acinetobacter baumannii		AB204810.2	ARO:3000753	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
28	abeS		Macrolide, Amino	Multidrug	Efflux	CARD	Acinetobacter baumannii AB307-0294		CP001172.1	ARO:3000768	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
29	acrA		Tetracycline, Pen	Multidrug	Efflux	CARD	Klebsiella pneumoniae		AJ318073.1	ARO:3000207	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
30	acrB	ECK0456, JW0451	Tetracycline, Pen	Multidrug	Efflux	CARD	Escherichia coli str. K-12 substr. MG1655		U00096.3	ARO:3000216	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
31	acrD		Aminoglycoside	Multidrug	Efflux	CARD	Escherichia coli str. K-12 substr. W3110		AP009048.1	ARO:3000491	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>

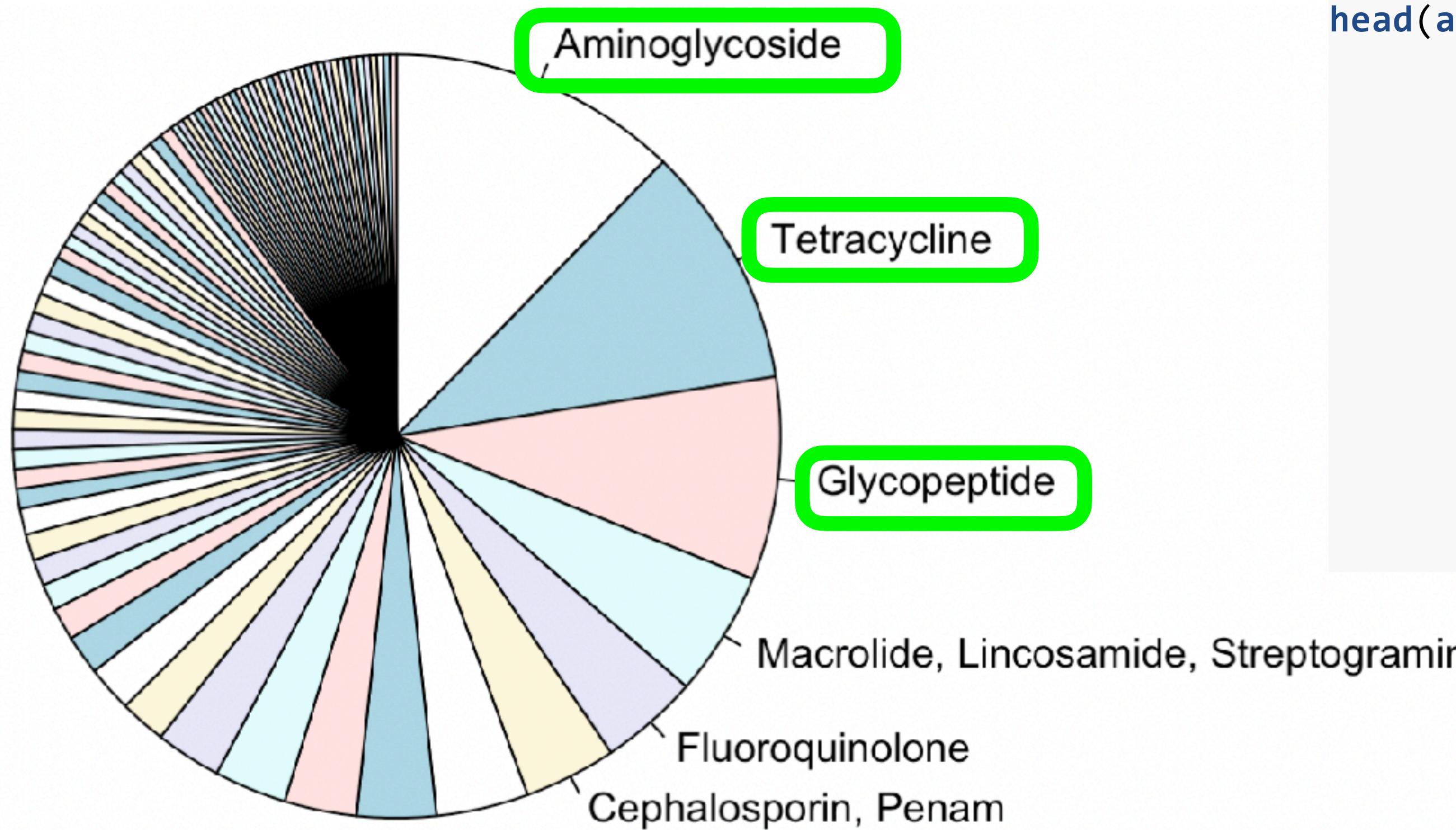
Base de datos con 360 genes

# Resultados - Base de datos (T2)

#Piechart con las clases de antibiótico de la base de datos

```
labels_cl <- rownames(head(sort(table(BD_ARGS$Clase), decreasing = T),  
6))
```

```
pie(sort(table(BD_ARGS$Clase), decreasing = T),  
init.angle = 90, clockwise = T, labels = labels_cl)
```



## Clase de antibióticos

#Tabla con las primeras clases de antibiótico de la base de datos

```
head(as.data.frame(sort(table(BD_ARGS$Clase), decreasing = T)), 10)
```

Var1	Freq
Aminoglycoside	44
Tetracycline	37
Glycopeptide	31
Macrolide, Lincosamide, Streptogramin	19
Fluoroquinolone	15
Cephalosporin, Penam	14
Peptide	14
Penam	12
Diaminopyrimidine	11
Phenicol	11

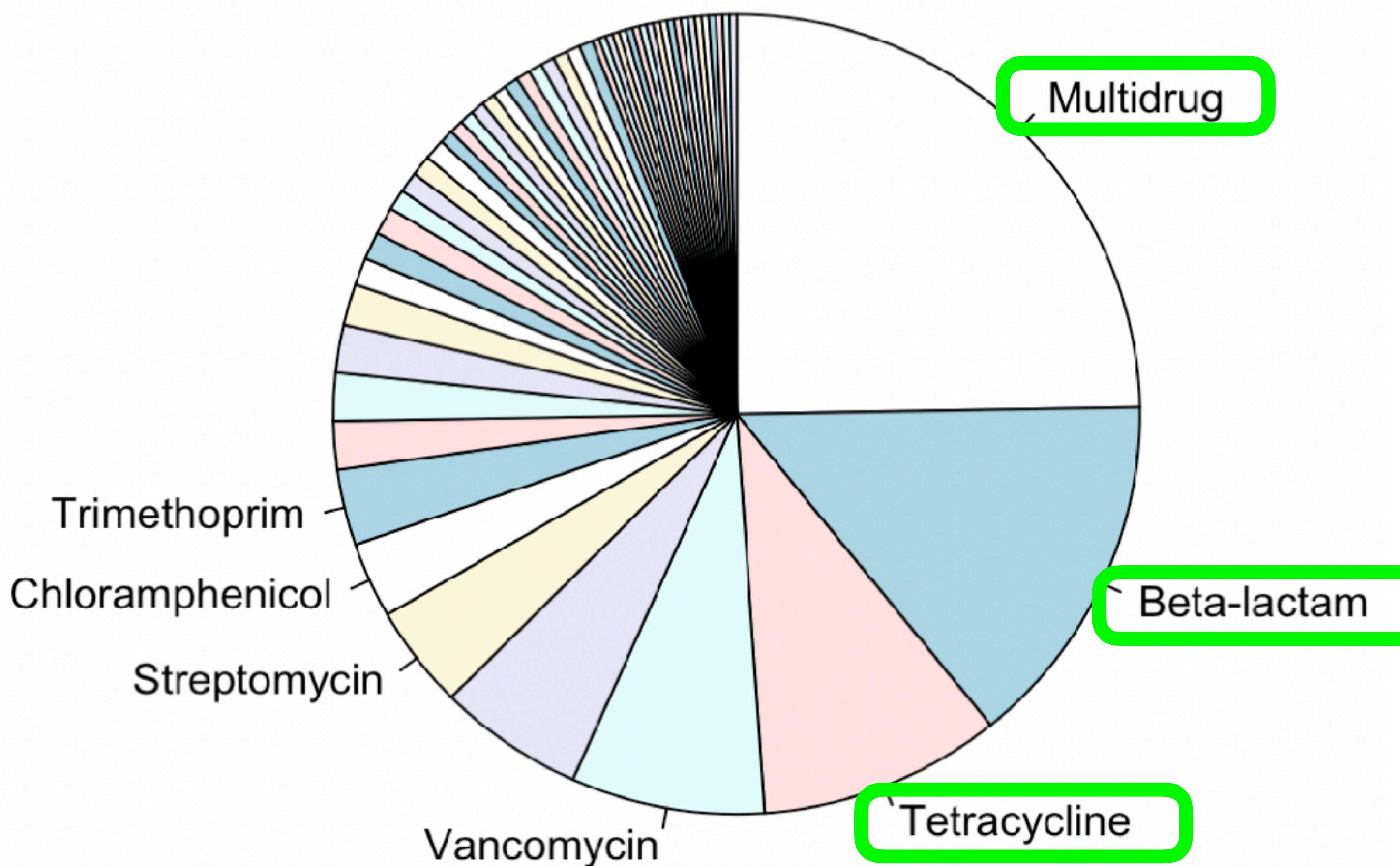
```
##  
## 1  
## 2  
## 3  
## 4 Macrolide, Lincosamide, Streptogramin  
## 5  
## 6  
## 7  
## 8  
## 9  
## 10
```

# Resultados - Base de datos (T2)

#Piechart con los antibióticos de la base de datos

```
labels_ab <- rownames(head(sort(table(BD_ARGS$Antibiotico), decreasing = T), 8))
```

```
pie(sort(table(BD_ARGS$Antibiotico), decreasing = T),  
    init.angle = 90, clockwise = T, labels = labels_ab)
```



## Antibiótico concreto

#Tabla con los primeros antibióticos de la base de datos

```
head(as.data.frame(sort(table(BD_ARGS$Antibiotico), decreasing = T)), 10)
```

##	Var1	Freq
## 1	Multidrug	89
## 2	Beta-lactam	52
## 3	Tetracycline	35
## 4	Vancomycin	28
## 5		21
## 6	Streptomycin	15
## 7	Chloramphenicol	11
## 8	Trimethoprim	11
## 9	Erythromycin	7
## 10	Fluoroquinolone	7

# Resultados - Selección de metagenomas (T3)

**MG-RAST** metagenomics analysis server

search

compost

study ▾

dataset (metagenome name) ▾

sequence type ▾

biome ▾

material ▾

location ▾

metagenome id ▾

project id ▾

Your search returned 147 results. Showing the first 20 matches.

download search results

+

study	dataset (metagenome name)	sequence type	biome	material	location	metagenome id	project id
compost thermophiles	compost-MC1_	shotgun metagenome	high temperature environment	soil	tambaram	mgm4848386.3	mpg89723
compostthermophiles	compost-MC1_	shotgun metagenome	high temperature environment	soil	tambaram	mgm4847936.3	mpg89696
Compost Soil	interleaved_MILM55_merge	shotgun metagenome	terrestrial biome	compost soil	Kompost werk, Goettingen GmbH, Germany	mgm4729903.3	mpg21223
compost psychrophiles	2P_R	shotgun metagenome	high temperature environment	soil	Rekong Peo	mgm4849831.3	mpg89833
Sao Paulo Zoo Composting	Zoo Compost 2 (ZC2)	shotgun metagenome	Subtropical and temperate rainforests or woodlands	compost	Sao Paulo Zoo Park	mgm4479944.3	mpg3997

# Resultados - Selección de metagenomas (T3)

```
#Número de metagenomas en cada proyecto en la búsqueda 1  
data.frame(sort(table(mgblastweb_1$project_id)), decreasing = T)
```

	##	Var1	Freq
	## 1	mgp5435	104
	## 2	mgp80207	12
	## 3	mgp12953	6
	## 4	mgp6843	6
	## 5	mgp12863	2
	## 6	mgp12864	2
	## 7	mgp18601	2
	## 8	mgp3997	2
	## 9	mgp80101	2
	## 10	mgp89870	2
	## 11	mgp12908	1
	## 12	mgp21223	1
	## 13	mgp3456	1
	## 14	mgp45	1
	## 15	mgp89696	1
	## 16	mgp89723	1
	## 17	mgp89833	1

```
#Cargar el fichero de la búsqueda 1 en MG-RAST
mgrastweb_1 <- read.delim("~/TFM/mgrastweb_1.txt", header = T

#Tamaño de la búsqueda 1
length(mgrastweb_1$project_id)

## [1] 147
```

```
#Cargar el fichero de la búsqueda 2 en MG-RAST
mgrastweb_2 <- read.delim("~/TFM/mgrastweb_2.txt", header = T)

#Tamaño de la búsqueda 2
length(mgrastweb_2$project_id)

## [1] 139
```

```
#Proyectos resultantes de la unión de las dos búsquedas
union(mgrastweb_1$project_id, mgrastweb_2$project_id)

## [1] "mgp89870"  "mgp89833"  "mgp89723"  "mgp89696"  "mg
## [7] "mgp6843"    "mgp5435"   "mgp45"      "mgp3997"   "mg
## [13] "mgp18601"   "mgp12953"  "mgp12908"  "mgp12864"  "mg
## [19] "mgp12931"   "mgp12866"
```

# Resultados - Selección de metagenomas (T3)

**Listado de metagenomas seleccionados  
(160 en total de 20 proyectos)**

**Ficheros en formato FASTA de las  
regiones de codificación**

A	B	C	D	E
Project id	Metagenome id		Project id	Metagenome id
2 mgp89870	mgm4850709.3		mgp5435	mgm4565311.3
3 mgp89870	mgm4850710.3		mgp5435	mgm4556395.3
4 mgp89833	mgm4849831.3		mgp5435	mgm4549531.3
5 mgp89723	mgm4848386.3		mgp5435	mgm4549526.3
6 mgp89696	mgm4847936.3		mgp5435	mgm4544128.3
7 mgp80207	mgm4742545.3		mgp5435	mgm4544129.3
8 mgp80207	mgm4742547.3		mgp5435	mgm4533996.3
9 mgp80207	mgm4742537.3		mgp5435	mgm4604928.3
10 mgp80207	mgm4742543.3		mgp5435	mgm4587465.3
11 mgp80207	mgm4742541.3		mgp5435	mgm4570096.3
12 mgp80207	mgm4742540.3		mgp5435	mgm4565313.3
13 mgp80207	mgm4742536.3		mgp5435	mgm4556404.3
14 mgp80207	mgm4743981.3		mgp5435	mgm4549530.3
15 mgp80207	mgm4743979.3		mgp5435	mgm4538739.3
16 mgp80207	mgm4749377.3		mgp5435	mgm4538734.3
17 mgp80207	mgm4743980.3		mgp5435	mgm4587467.3
18 mgp80207	mgm4742538.3		mgp5435	mgm4587458.3
19 mgp80207	mgm4749376.3		mgp5435	mgm4544135.3
20 mgp80207	mgm4749374.3		mgp5435	mgm4544125.3
21 mgp80207	mgm4749373.3		mgp5435	mgm4538740.3
22 mgp80207	mgm4749375.3		mgp5435	mgm4538738.3
23 mgp80207	mgm4742544.3		mgp5435	mgm4533997.3
24 mgp80207	mgm4742542.3		mgp5435	mgm4534000.3
25 mgp80101	mgm4741983.3		mgp5435	mgm4530681.3
26 mgp80101	mgm4741982.3		mgp5435	mgm4530682.3
27 mgp6843	mgm4544117.3		mgp5435	mgm4604927.3
28 mgp6843	mgm4544118.3		mgp5435	mgm4587468.3
29 mgp6843	mgm4544116.3		mgp5435	mgm4587462.3
30 mgp6843	mgm4544115.3		mgp5435	mgm4587455.3
31 mgp6843	mgm4559244.3		mgp5435	mgm4549527.3

# Resultados - Identificación ARGs (T4)

Genomics 111 (2019) 883–885



Contents lists available at ScienceDirect

Genomics

journal homepage: [www.elsevier.com/locate/ygeno](http://www.elsevier.com/locate/ygeno)



Short Communication

## MetaHMM: A webserver for identifying novel genes with specified functions in metagenomic samples

Balázs Szalkai<sup>a,\*</sup>, Vince Grolmusz<sup>a,b,\*\*</sup>

<sup>a</sup> PIT Bioinformatics Group, Eötvös University, H-1117 Budapest, Hungary

<sup>b</sup> Uratim Ltd., H-1118 Budapest, Hungary



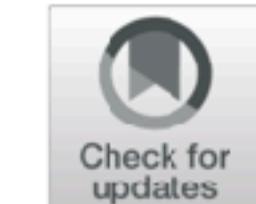
LaPierre et al. *Genome Biology* (2020) 21:242  
<https://doi.org/10.1186/s13059-020-02159-0>

SOFTWARE

Open Access

## Metalign: efficient alignment-based metagenomic profiling via containment min hash

Nathan LaPierre<sup>1\*</sup>, Mohammed Alser<sup>2</sup>, Eleazar Eskin<sup>1,3,4</sup>, David Koslicki<sup>5,6,7†</sup> and Serghei Mangul<sup>8\*\*</sup> 



Genome Biology

## Fast and sensitive protein alignment using DIAMOND

Benjamin Buchfink<sup>1</sup>, Chao Xie<sup>2,3</sup> & Daniel H Huson<sup>1,2</sup>

The alignment of sequencing reads against a protein reference database is a major computational bottleneck in metagenomics and data-intensive evolutionary projects. Although recent tools offer improved performance over the gold standard BLASTX, they exhibit only a modest speedup or low sensitivity. We introduce DIAMOND, an open-source algorithm based on double indexing that is 20,000 times faster than BLASTX on short reads and has a similar degree of sensitivity.

<sup>1</sup>Department of Computer Science and Center for Bioinformatics, University of Tübingen, Tübingen, Germany. <sup>2</sup>Singapore Centre on Environmental Life Sciences Engineering, School of Biological Sciences, Nanyang Technological University, Singapore. <sup>3</sup>Life Sciences Institute, National University of Singapore, Singapore. Correspondence should be addressed to B.B. (buchfink@gmail.com) or D.H.H. (daniel.huson@uni-tuebingen.de).

RECEIVED 29 APRIL; ACCEPTED 20 OCTOBER; PUBLISHED ONLINE 17 NOVEMBER 2014; DOI:10.1038/NMETH.3176

NATURE METHODS | VOL.12 NO.1 | JANUARY 2015 | 59

# Resultados - Identificación ARGs (T4)

A) Código de la consulta: Código del read del metagenoma que se consulta contra la base de datos

B) Código del objetivo: Código del gen de la base de datos contra el que se alinea la consulta

C) Identidad de secuencia: Porcentaje de aminoácidos idénticos alineados en un alineamiento local de las dos secuencias

K) E-valor: Es el valor esperado del número de alineamientos de similar o mejor calidad que se espera encontrar al realizar esta consulta frente a una base de datos de secuencias aleatorias del mismo tamaño que la base de datos real. Un valor de 0.001 significa que un alineamiento de esta calidad se puede encontrar al azar en promedio en una de cada 1000 consultas

L) Bit score: Es una medida independiente de la matriz de puntuación de la similitud de las dos secuencias alineadas. Números altos significan mayor similitud entre secuencias

	A	B	C	D	E	F	G	H	I	J	K	L
1	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	ceoB	72.7	33	9	0	1	33	746	778	1.1e-10	55.1
2	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	adeF	71.9	32	9	0	2	33	745	776	1.9e-10	54.3
3	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	MexF	68.8	32	10	0	2	33	750	781	2.5e-10	53.9
4	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	acrD	51.5	33	16	0	1	33	737	769	9.3e-05	35.4
5	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	mdtF	48.5	33	17	0	1	33	737	769	9.3e-05	35.4
6	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	TtgB	45.5	33	18	0	1	33	737	769	2.1e-04	34.3
7	M01677:7:000000000-A4LYE:1:1107:16186:3624_1_100_-	acrF	50.0	30	15	0	1	30	738	767	2.7e-04	33.9
8	M01677:7:000000000-A4LYE:1:2109:10905:8766_1_188_-	smeE	87.1	62	8	0	1	62	865	926	1.2e-26	109.0
9	M01677:7:000000000-A4LYE:1:2109:10905:8766_1_188_-	acrD	67.7	62	20	0	1	62	861	922	4.6e-21	90.5
10	M01677:7:000000000-A4LYE:1:2109:10905:8766_1_188_-	MexB	62.9	62	23	0	1	62	862	923	3.0e-20	87.8

## Genes más frecuentes con identidad de secuencia superior al 98%

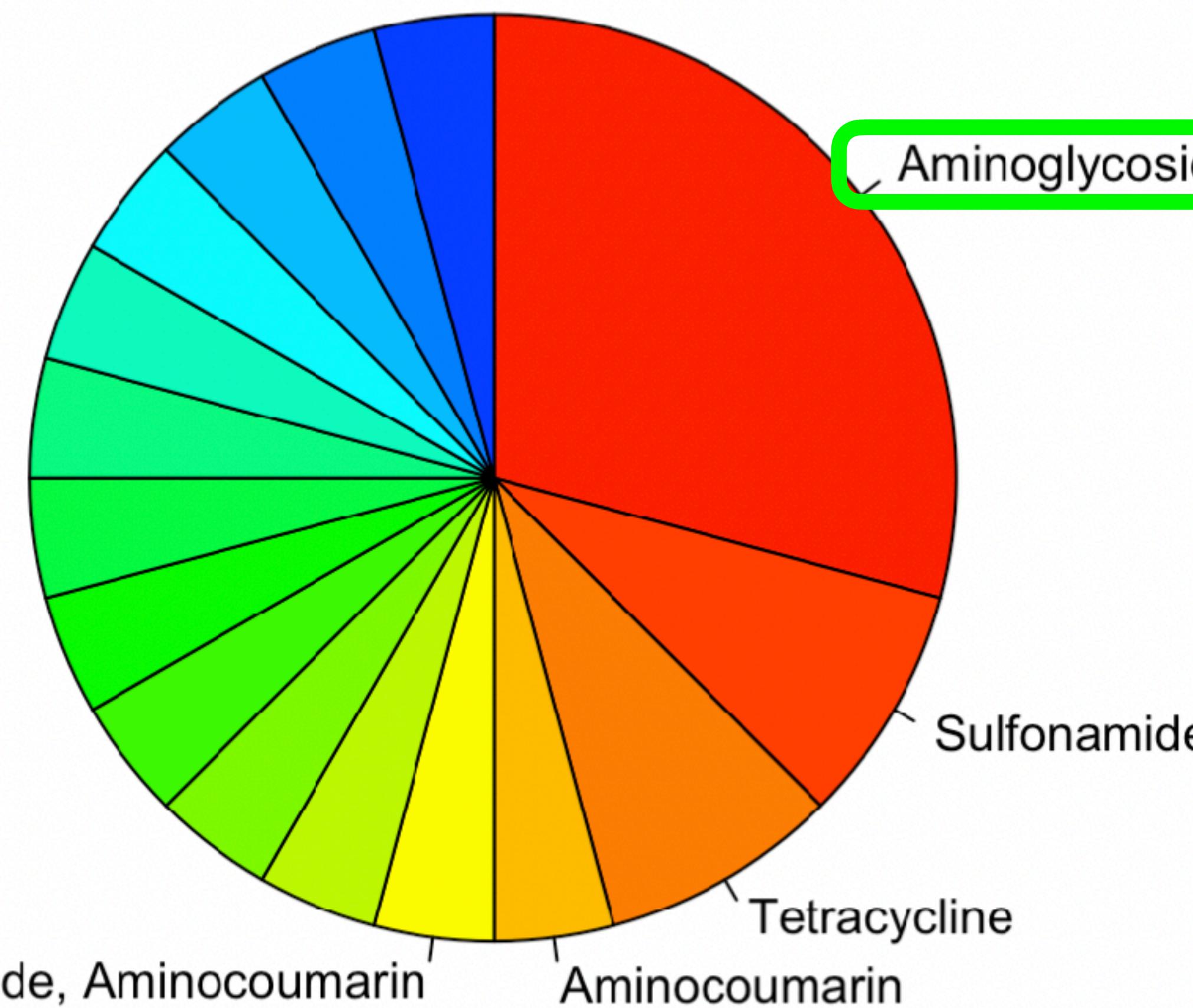


A	B	C	D	E
1	Gen	Frecuencia	Clase	
2	rrnS	80,62	Aminoglycoside	Streptomycin
3	aadA	59,38	Aminoglycoside	Streptomycin
4	sul2	58,13	Sulfonamide	Sulfonamide
5	aadA24	56,25	Aminoglycoside	Streptomycin, Spectinomycin
6	aph(6')-Id	54,37	Aminoglycoside	Streptomycin
7	sul1	54,37	Sulfonamide	Sulfonamide
8	APH(3")-Ib	44,38	Aminoglycoside	Streptomycin
9	aadA6	40,62	Aminoglycoside	Streptomycin
10	acrB	38,12	Tetracycline, Penam, Rifamycin, Glycylcycline, Cephalosporin, Cephem, Cephamycin	Multidrug
11	acrD	34,38	Aminoglycoside	Multidrug
12	acrF	33,75	Fluoroquinolone, Penam, Cephalosporin, Cephem, Cephemycin	Multidrug
13	mtrA	31,87	Macrolide, Penam	Multidrug
14	smeE	31,25	Fluoroquinolone, Macrolide, Tetracycline, Phenicol	Multidrug
15	ttgB	29,38	Tetracycline	Tetracycline
16	adeJ	29,38	Diaminopyrimidine, Rifamycin, Penem, Tetracycline, Phosphotransferase system	Multidrug
17	CRP	27,5	Penam, Macrolide, Fluoroquinolone	Multidrug
18	MexB	27,5	Peptide, Penam, Diaminopyrimidine, Sulfonamide, Fluorophenicol	Multidrug
19	mdtB	27,5	Aminocoumarin	Multidrug
20	TEM-1	26,88	Cephalosporin, Penam, Penem, Monobactam	Beta-lactam
21	emrB	26,88	Fluoroquinolone	Multidrug
22	cpxA	26,25	Aminoglycoside, Aminocoumarin	Escherichia coli O157:H7 str. Sakai
23	adeF	25,62	Tetracycline, Fluoroquinolone	Multidrug
24	tet(L)	25,62	Tetracycline	Tetracycline
25	mdtF	25,62	Fluoroquinolone, Macrolide, Penam	Multidrug

# Resultados - Identificación ARGs (T4)

```
#Piechart con las clases de antibiótico de los genes con 98% identidad
labels_cl_98 <- rownames(head(sort(table(id98final$Clase), decreasing = T), 5))

pie(sort(table(id98final$Clase), decreasing = T),
    init.angle = 90, clockwise = T, labels = labels_cl_98, col =
rainbow(nrow(id98final)))
```



## Clase de antibióticos con identidades >98%

#Tabla con las primeras clases de antibiótico de los genes con 98% identidad

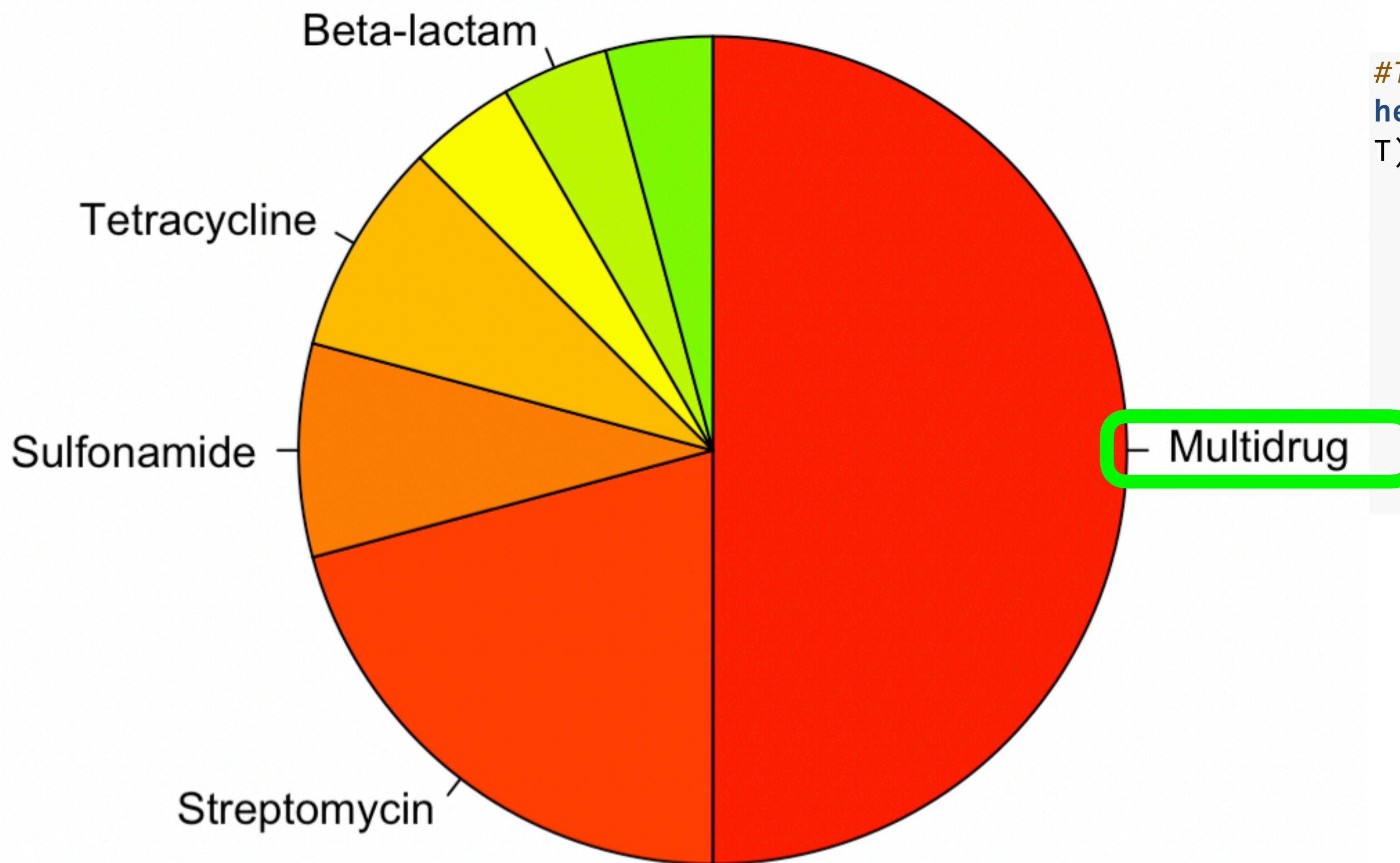
```
head(as.data.frame(sort(table(id98final$Clase), decreasing = T)), 10)
```

	Var1	Freq
## 1	Aminoglycoside	7
## 2	Sulfonamide	2
## 3	Tetracycline	2
## 4	Aminocoumarin	1
## 5	Aminoglycoside, Aminocoumarin	1
## 6	Cephalosporin, Penam, Penem, Monobactam	1
## 7	Diaminopyrimidine, Rifamycin, Penem, Tetracycline, Phenicol, Carbapenem, Macrolide, Lincosamide, Fluoroquinolone, Cephalosporin	1
## 8	Fluoroquinolone	1
## 9	Fluoroquinolone, Macrolide, Penam	1
## 10	Fluoroquinolone, Macrolide, Tetracycline, Phenicol	1

# Resultados - Identificación ARGs (T4)

```
#Piechart con Los antibióticos de los genes con 98% identidad
labels_ab_98 <- rownames(head(sort(table(id98final$Antibiotico),
decreasing = T), 6))

pie(sort(table(id98final$Antibiotico), decreasing = T),
    init.angle = 90, clockwise = T, labels = labels_ab_98, col =
rainbow(nrow(id98final)))
```



**Antibiótico concreto  
con identidades >98%**

#Tabla con Los primeros antibióticos de los genes con 98% identidad
head(as.data.frame(sort(table(id98final\$Antibiotico), decreasing =
T)), 10)

	Var1	Freq
## 1	Multidrug	12
## 2	Streptomycin	5
## 3	Sulfonamide	2
## 4	Tetracycline	2
## 5		1
## 6	Beta-lactam	1
## 7	Streptomycin, Spectinomycin	1

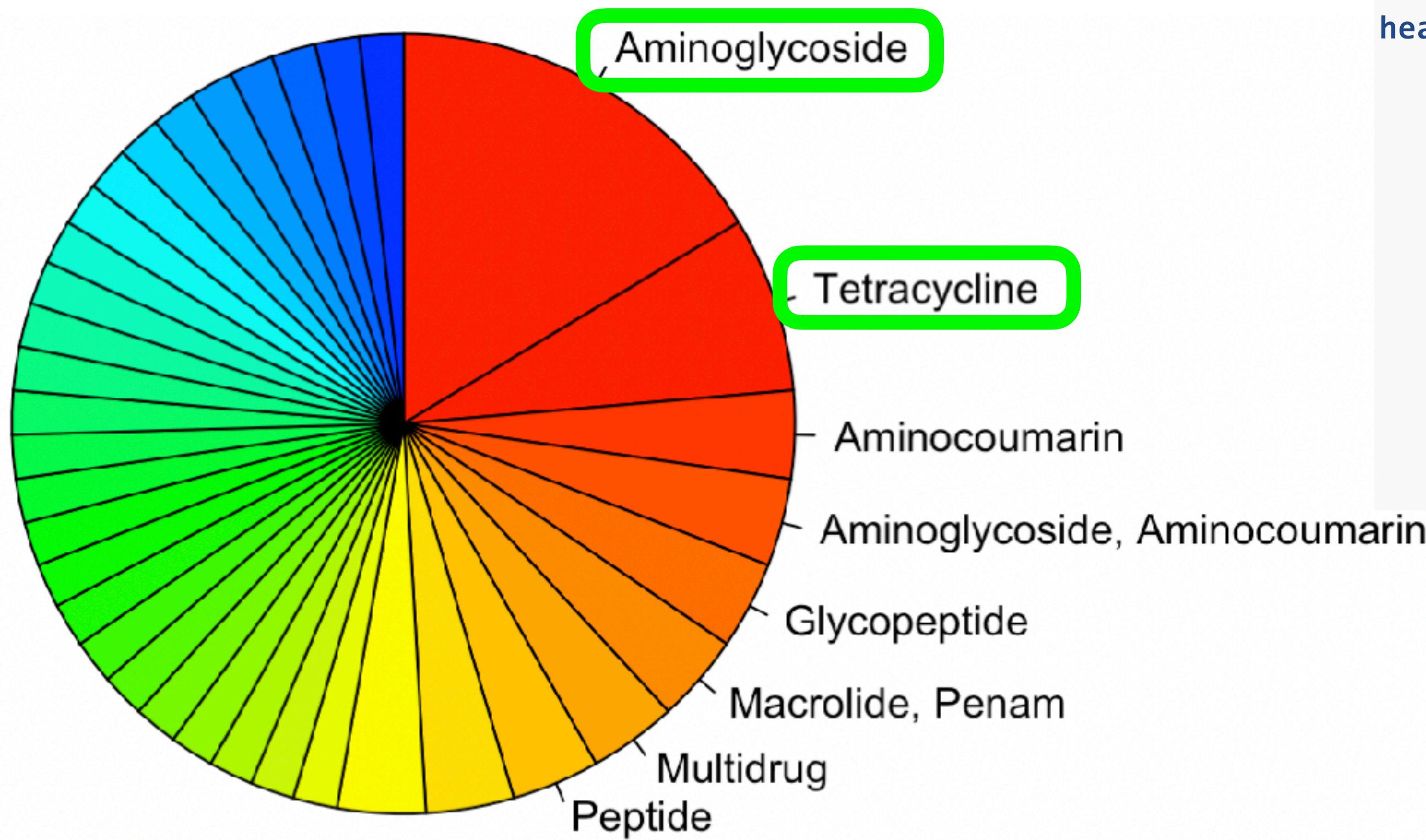
## Genes más frecuentes con identidad de secuencia superior al 90%

A	B	C	D	E
1	Gen	Frecuencia Clase	Antibiotico	Organismo
2	rrnS	88,75 Aminoglycoside	Streptomycin	Chlamydomonas reinhardtii
3	acrB	80,63 Tetracycline, Penam, Rifamycin, Glycylcycline, Cephalosporin, Phenicol, Triclosan, Fluoroquinolone	Multidrug	Escherichia coli str. K-12 substr. MG1655
4	SdeY	76,88 Multidrug	Multidrug	Serratia marcescens
5	mtrA	76,25 Macrolide, Penam	Multidrug	Mycobacterium tuberculosis H37Rv
6	MexB	75,00 Peptide, Penam, Diaminopyrimidine, Sulfonamide, Fluoroquinolone, Monobactam, Phenicol, Carbapenem	Multidrug	Pseudomonas aeruginosa
7	acrF	72,50 Fluoroquinolone, Penam, Cephalosporin, Cephemycin	Multidrug	Escherichia coli str. K-12 substr. MG1655
8	smeE	71,25 Fluoroquinolone, Macrolide, Tetracycline, Phenicol	Multidrug	Stenotrophomonas maltophilia
9	adeF	71,25 Tetracycline, Fluoroquinolone	Multidrug	Acinetobacter baumannii AYE
10	MexF	70,00 Diaminopyrimidine, Phenicol, Fluoroquinolone	Multidrug	Pseudomonas aeruginosa PAO1
11	mexK	69,38 Macrolide, Tetracycline, Triclosan	Multidrug	Pseudomonas aeruginosa PAO1
12	acrD	68,13 Aminoglycoside	Multidrug	Escherichia coli str. K-12 substr. W3110
13	ttgb	66,88 Tetracycline	Tetracycline	Pseudomonas putida
14	ceoB	65,00 Aminoglycoside, Fluoroquinolone	Multidrug	Burkholderia cepacia
15	smeB	64,38 Cephalosporin, Cephemycin, Penam, Aminoglycoside	Multidrug	Stenotrophomonas maltophilia
16	mexW	62,50 Tetracycline, Acridine, Macrolide, Phenicol, Fluoroquinolone		Pseudomonas aeruginosa PAO1
17	adeJ	62,50 Diaminopyrimidine, Rifamycin, Penem, Tetracycline, Phenicol, Carbapenem, Macrolide, Lincosamide	Multidrug	Acinetobacter baumannii
18	aadA	61,88 Aminoglycoside	Streptomycin	Escherichia coli
19	aadA24	61,88 Aminoglycoside	Streptomycin, Spectinomycin	Salmonella enterica subsp. enterica serovar Newport
20	sul2	58,13 Sulfonamide	Sulfonamide	Vibrio cholerae
21	mdtF	58,13 Fluoroquinolone, Macrolide, Penam	Multidrug	Escherichia coli str. K-12 substr. MG1655
22	APH(6)-Id	56,25 Aminoglycoside	Streptomycin	Pseudomonas aeruginosa
23	aadA2	56,25 Aminoglycoside	Streptomycin, Spectinomycin	Klebsiella pneumoniae
24	pncA	55,63 Pyrazinamide	Pyrazinamide	Mycobacterium tuberculosis H37Rv
25	aadA3	55,63 Aminoglycoside	Streptomycin	Plasmid NR79
26	sul1	55,63 Sulfonamide	Sulfonamide	Vibrio fluvialis
27	msbA	54,38 Nitroimidazole	Multidrug	Escherichia coli str. K-12 substr. MG1655
28	mdtB	54,38 Aminocoumarin	Multidrug	Escherichia coli str. K-12 substr. MG1655
29	aadA6	53,75 Aminoglycoside	Streptomycin	Pseudomonas aeruginosa
30	macB	50,00 Macrolide		Neisseria gonorrhoeae
31	mdtC	48,75 Aminocoumarin	Multidrug	Escherichia coli str. K-12 substr. MG1655

# Resultados - Identificación ARGs (T4)

```
#Piechart con las clases de antibiótico de los genes con 90% identidad
labels_cl_90 <- rownames(head(sort(table(id90final$Clase), decreasing = T), 8))

pie(sort(table(id90final$Clase), decreasing = T),
    init.angle = 90, clockwise = T, labels = labels_cl_90, col =
rainbow(nrow(id90final)))
```



## Clase de antibióticos con identidades >90%

```
#Tabla con las primeras clases de antibiótico de los genes con 90% identidad
head(as.data.frame(sort(table(id90final$Clase), decreasing = T)), 10)
```

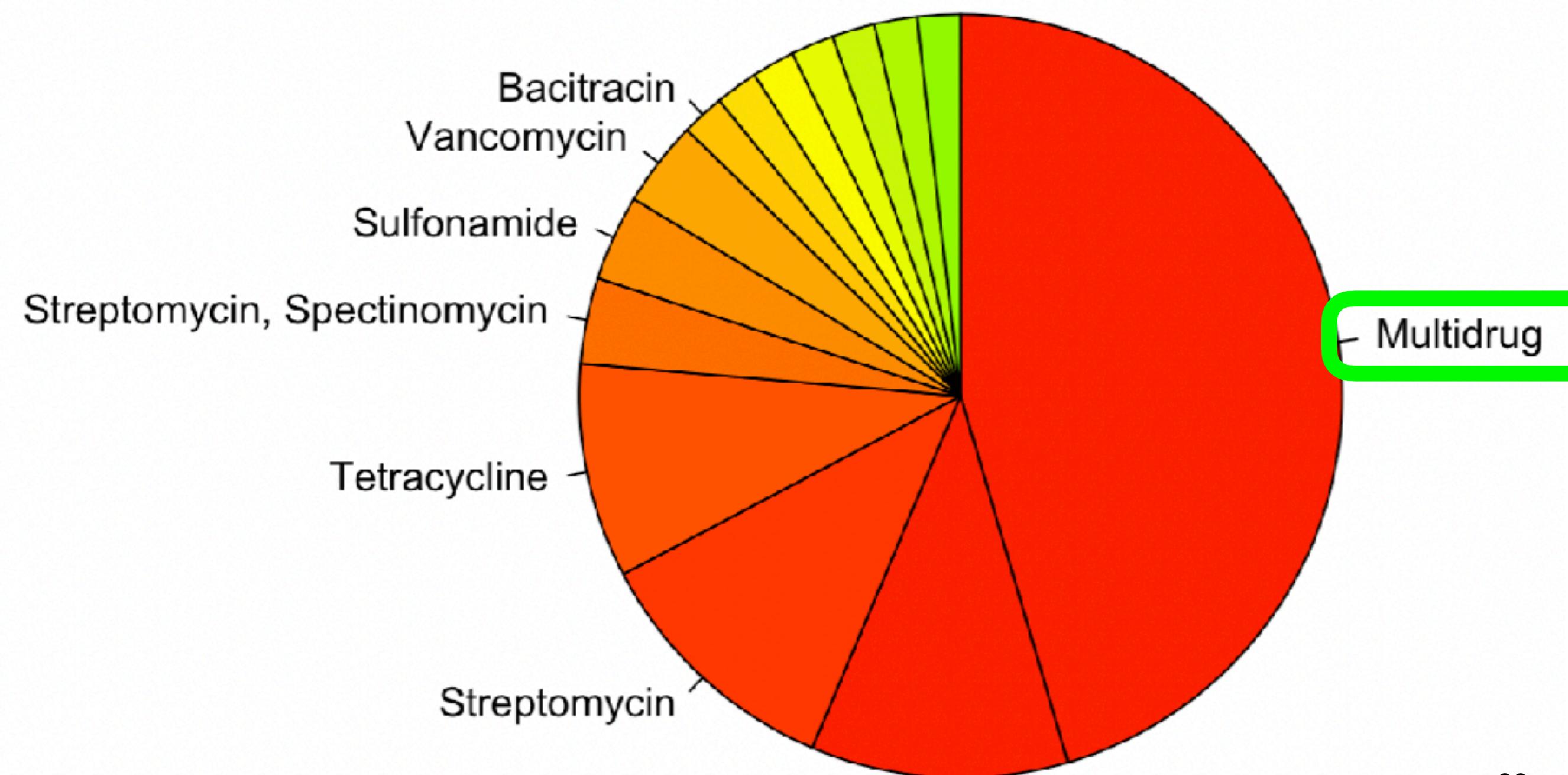
	Var1	Freq
## 1	Aminoglycoside	9
## 2	Tetracycline	4
## 3	Aminocoumarin	2
## 4	Aminoglycoside, Aminocoumarin	2
## 5	Glycopeptide	2
## 6	Macrolide, Penam	2
## 7	Multidrug	2
## 8	Peptide	2
## 9	Sulfonamide	2
## 10	Tetracycline, Glycylcycline	2

# Resultados - Identificación ARGs (T4)

```
#Piechart con los antibióticos de los genes con 90% identidad
labels_ab_90 <- rownames(head(sort(table(id90final$Antibiotico),
decreasing = T), 8))

pie(sort(table(id90final$Antibiotico), decreasing = T),
    init.angle = 90, clockwise = T, labels = labels_ab_90, col =
rainbow(nrow(id90final)))
```

**Antibiótico concreto  
con identidades >90%**



```
#Tabla con los primeros antibióticos de los genes con 90% identidad
head(as.data.frame(sort(table(id90final$Antibiotico), decreasing =
T)), 10)
```

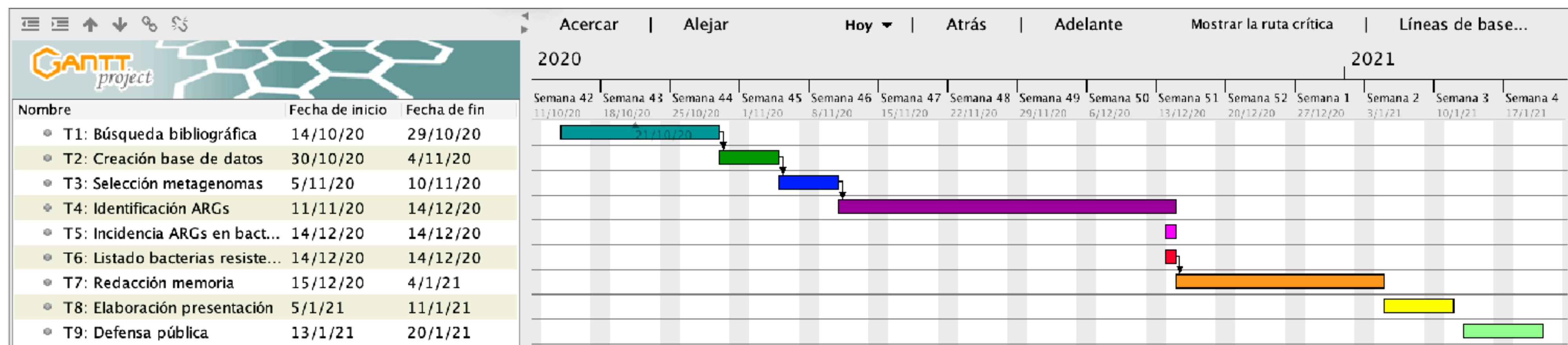
Var1	Freq
Multidrug	25
	6
Streptomycin	6
Tetracycline	5
Streptomycin, Spectinomycin	2
Sulfonamide	2
Vancomycin	2
Bacitracin	1
Beta-lactam	1
Chloramphenicol	1

# Conclusiones

- Las resistencias de los genes más frecuentes de la base de datos y de los metagenomas con diferentes porcentajes de identidad de secuencia son bastante similares: aminoglucósidos, tetraciclinas y multirresistentes
- Tipos de resistencias similares pueden indicar que la composición de las diferentes muestras de compost son similares, al menos en cuanto al tipo de ARGs que se encuentra en las mismas
- Se han encontrado un número elevado de genes en al menos el 50% de los metagenomas (6 para identidades de secuencia superiores al 98% y 29 para identidades de secuencia superiores al 90%). La ubicuidad de estos genes evidencia la gran extensión que tienen los ARGs en las diferentes muestras de compost

# Conclusiones

- Se ha logrado el objetivo general de identificar los ARGs presentes en muestras de compost. Además, se han logrado otros objetivos secundarios como la creación de una base de datos y la selección de metagenomas de muestras de compost y su comparación con la base de datos. Por desgracia, no se han logrado el resto de objetivos debido al retraso en la tarea T4
- Se han cumplido las cuatro primeras tareas (T1-T4) y los tres primeros hitos descritos en la planificación del trabajo. Ha sido imposible completar las tareas T5-T6 y el cuarto hito
- El cronograma final del proyecto refleja el aumento de tiempo para completar la tarea T4 y el nulo tiempo asignado para las tareas T5 y T6



# **Estudio del resistoma del compost: identificación de genes bacterianos resistentes a antibióticos**

**Aritz Durana**

**Master Universitario en Bioinformática y Bioestadística UOC-UB**  
**Área 4, Subárea 6: Microbiología, biotecnología y biología molecular**  
**Consultora: Paloma Pizarro Tobias**

Enero 2021