



# Desarrollo de una aplicación web para la visualización dinámica de duplicaciones génicas en bacterias

Bacteria

Duplicates

Genes

Shiny Apps

R Circos

web html

Blast

Fasta

<b>Antecedentes</b>	<b>Contexto biológico, continuación de BacDup, objetivo del proyecto, temporización TFM</b>
<b>Planificación y diseño</b>	<b>Estudio previo de necesidades, archivos de partida, selección de herramientas, objetivos detallados.</b>
<b>Resultados</b>	<b>Presentación de todos los apartados de la aplicación, sus funcionalidades y usabilidad</b>
<b>Conclusiones</b>	<b>Valoración de la aplicación, observaciones sobre el desarrollo, limitaciones, evolutivos futuros, cierre.</b>



THE LANCET

Submit Article

ARTICLES | VOLUME 399, ISSUE 10325, P629-655, FEBRUARY 12, 2022

## Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis

Antimicrobial Resistance Collaborators <sup>†</sup> • [Show footnotes](#)

Open Access • Published: January 19, 2022 • DOI: [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)

 Check for updates

Summary

Introduction

Methods

Results

Discussion

### Summary

#### Background

Antimicrobial resistance (AMR) poses a major threat to human health around the world. Previous publications have estimated the effect of AMR on incidence, deaths, hospital length of stay, and health-care costs for specific pathogen–drug combinations in select locations. To our knowledge, this study presents the most comprehensive estimates of AMR burden to date.

- Resistencia a los antibióticos (AMR antimicrobial resistance).
- 23 patógenos y 88 combinaciones patógeno-tratamiento.
- Problema de salud pública en todo el mundo. 204 países.
- Muertes y casos graves por infecciones de origen nosocomial.
- Costes económicos: mayor estancia en hospital y gastos directos.



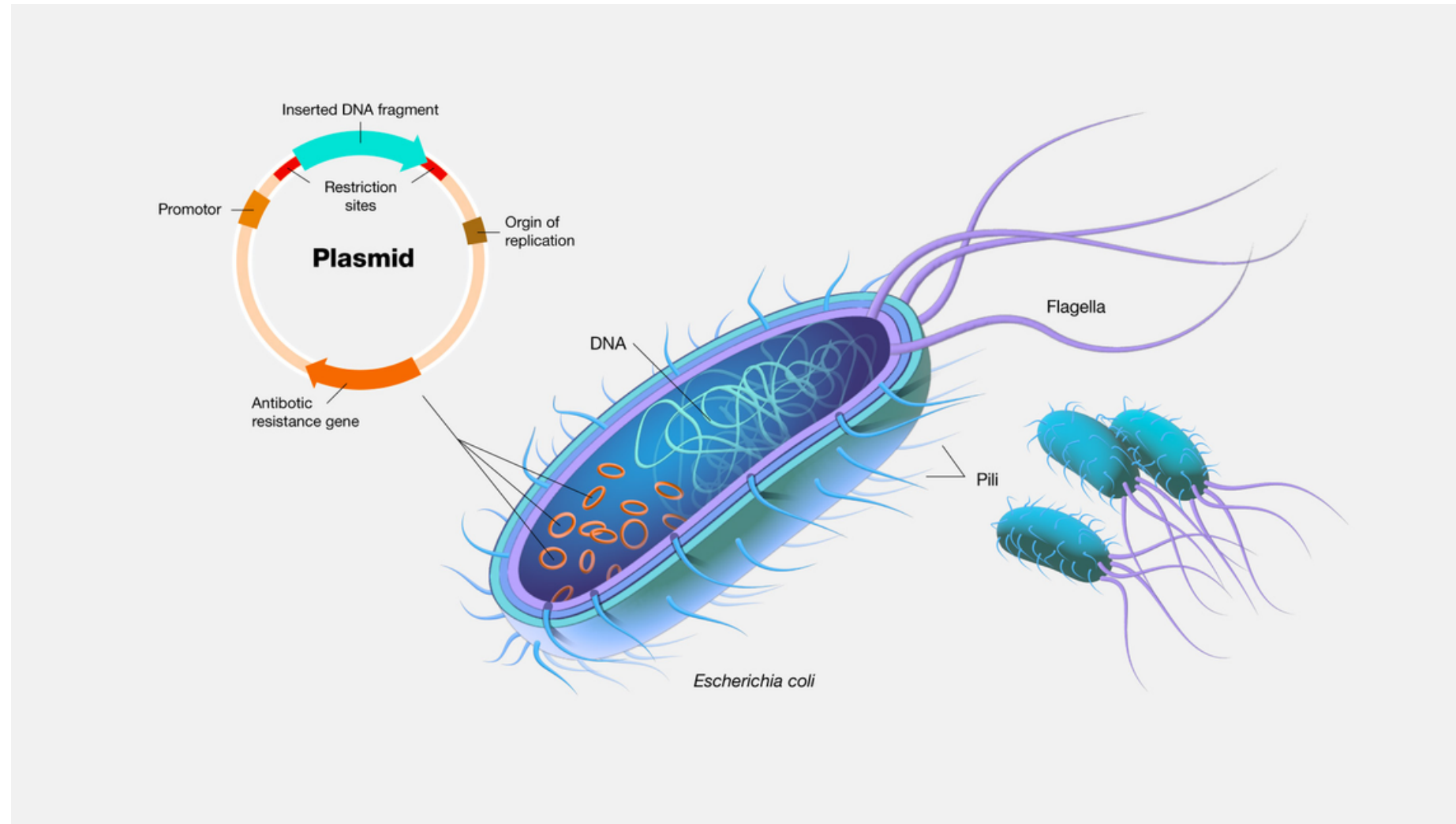
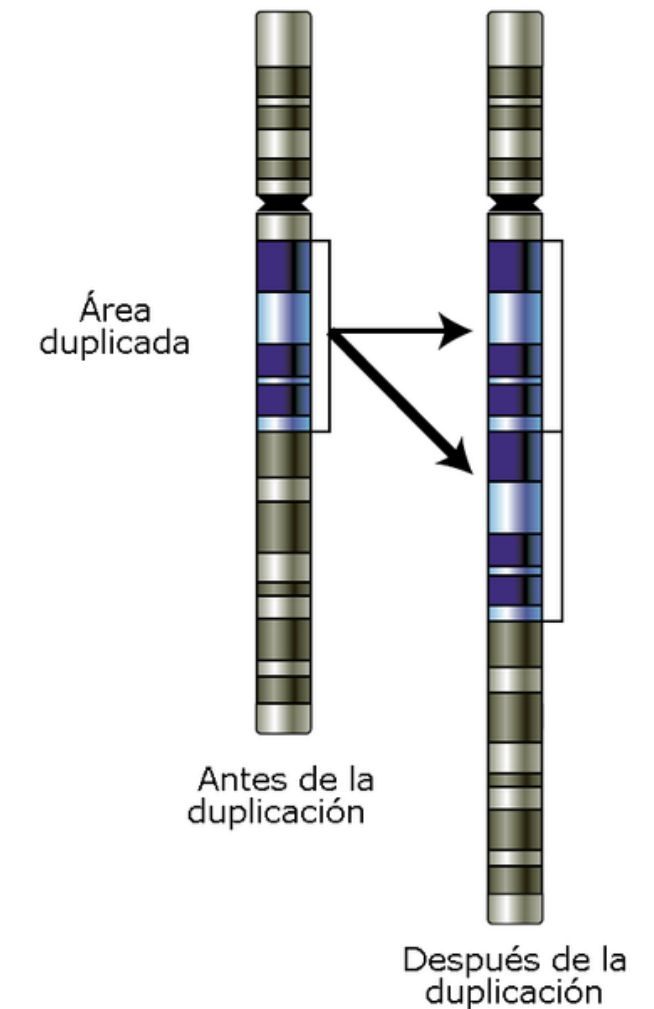
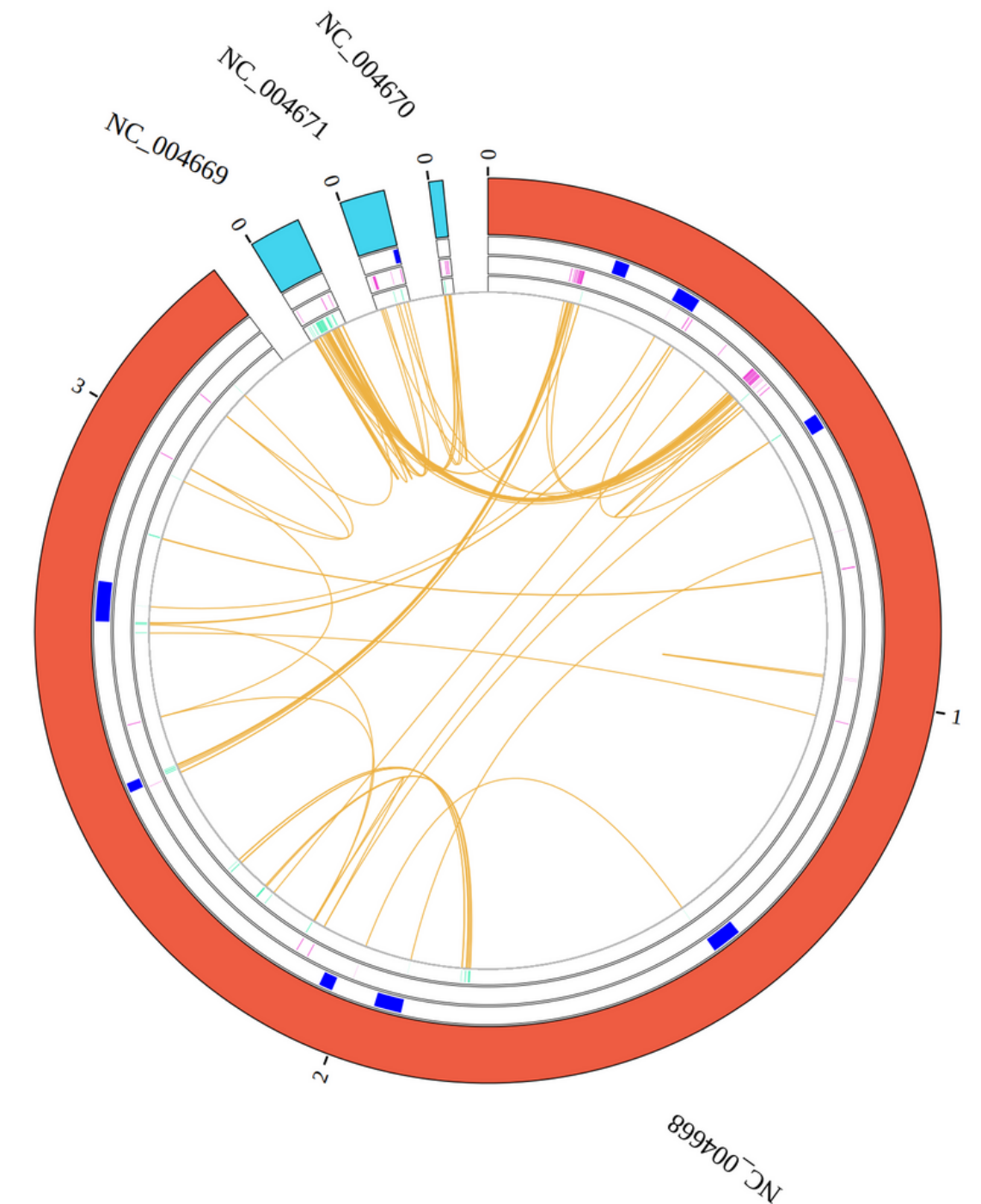
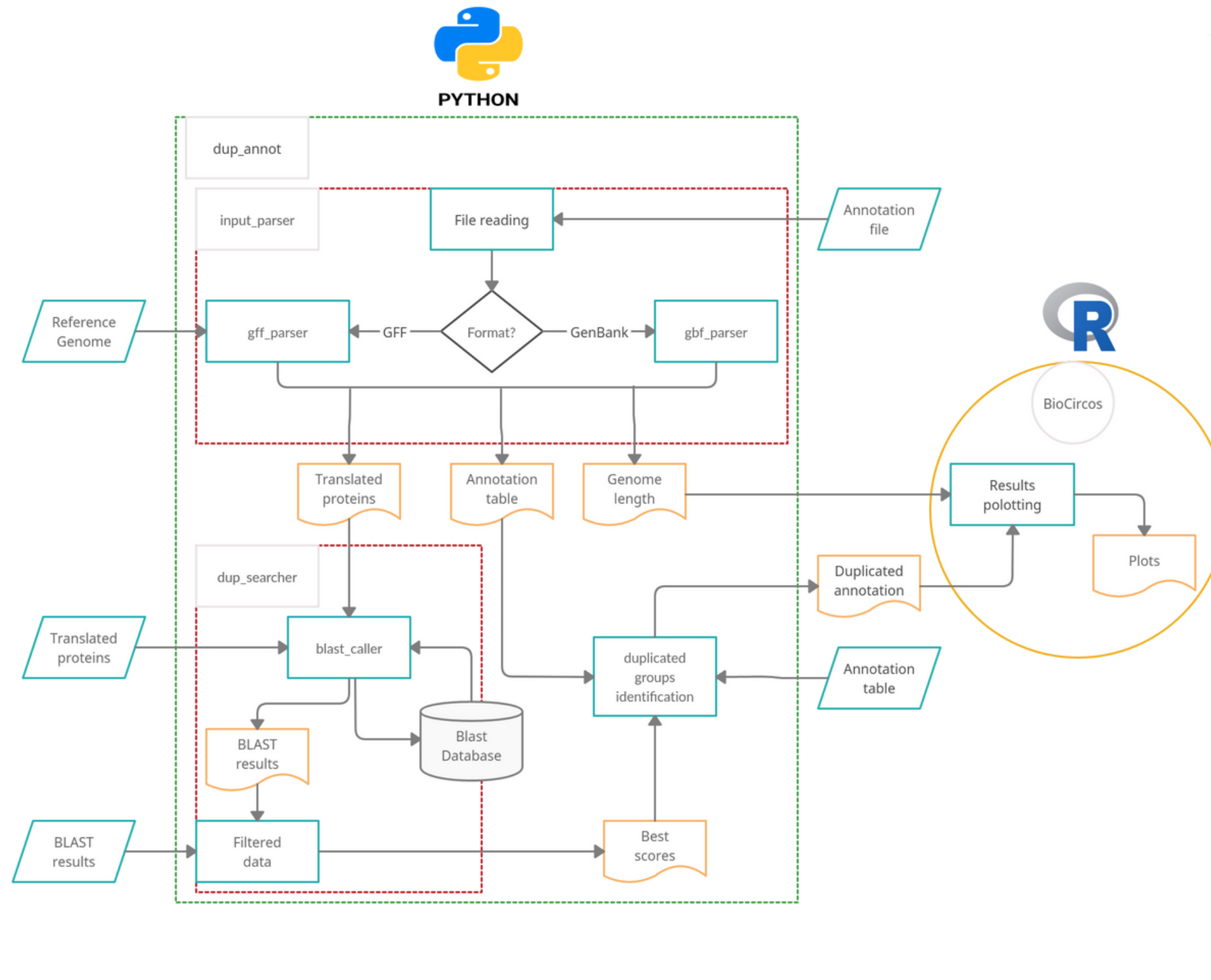


Imagen de NIH National Human Genome Research Institute



- Estudio genoma sobre cromosoma central y plásmidos.
- Duplicaciones génicas como potencial característica de estudio.
- Mayor número de genes duplicados presentes en cepas patógenas.
- Posible creación de resistencia por ligeras mutaciones adaptativas.
- Posible virulencia por múltiples copias del mismo gen.

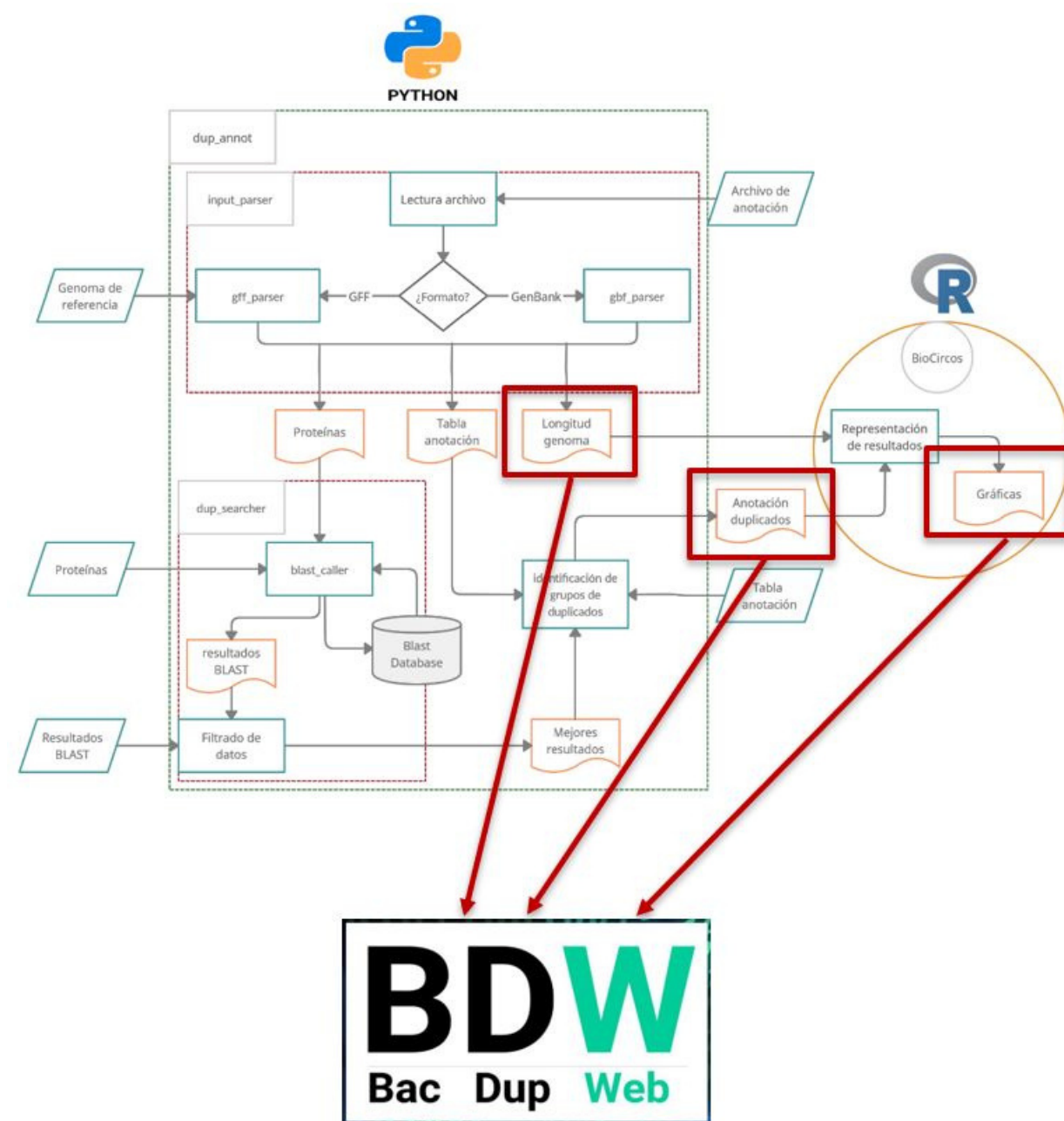




**TFM de Alba Moya con el profesor José Francisco Sánchez**



Aplicación web para poder mostrar el gráfico biocircos y los datos de duplicaciones de cualquier cepa bacteriana con la ventaja de poder clasificar, filtrar, modificar las características de estudio, viendo los resultados en tiempo real.



Original Data		Filtered Data	
Name	Value	Name	Value
Records	612	Records	431
max_dups	11	max_dups	11
min_dups	2	min_dups	2

TFM Daniel Ibáñez  
Tutor José Francisco Sánchez









Column Name	Description
<b>rec_id</b>	Contig or chromosome ID
<b>locus_tag</b>	Unique gene ID provided by Genbank annotation
<b>protein_id</b>	Protein ID provided by Genbank annotation
<b>gene</b>	Gene name (if available)
<b>start</b>	Coordinate Start position (in rec_id)
<b>end</b>	Coordinate End position (in rec_id)
<b>strand</b>	Strand value (+/-)
<b>pseudo</b>	Pseudogene True/False
<b>product</b>	Gene description (if available)
<b>Dbxref</b>	External reference to third party database (if available)
<b>inference</b>	Inference details (if available)
<b>EC_number</b>	Enzyme family unique ID (if available)
<b>old_locus_tag</b>	Former Unique gene ID provided by Genbank annotation



### GCF\_000299455\_length.csv

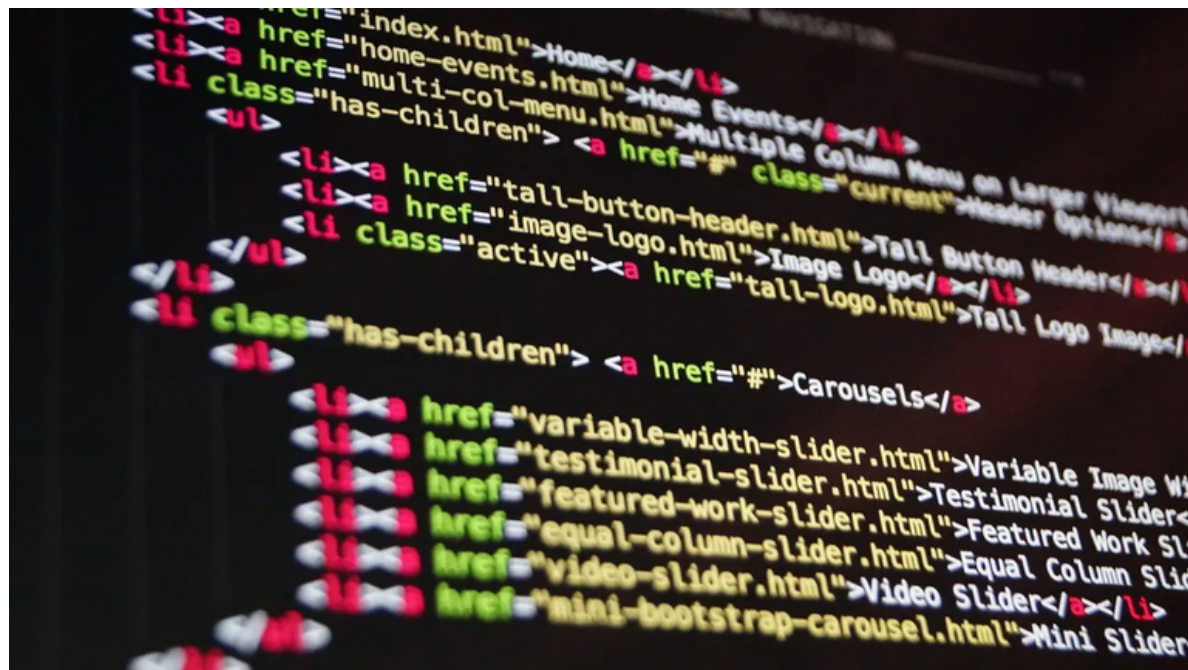
```
GCF_000299455_length_df.csv: Bloc de ...
Archivo Edición Formato Ver Ayuda
NC_018658.1,5273097
NC_018666.1,74217
NC_018659.1,88544
NC_018660.1,1549
```

### GCF\_000299455\_dup\_annot.csv

```
GCF_000299455_dup_annot.csv: Bloc de notas
Archivo Edición Formato Ver Ayuda
|,rec_id,locus_tag,protein_id,gene,start,end,strand,pseudo,product,Dbxref,inference,EC_number,old_locus_tag,dup_id,count_dups,dup_id_pseudo_free,count_dups_pseudo_free,dup_id_mobile_free,count_dups_mobile_free
CDS_NC_018658.1_2898196_2901676_neg,NC_018658.1,O3K_RS14150,WP_000515776.1,,2898196,2901676,neg,,host specificity protein J,,COORDINATES: similar to AA sequence:RefSeq:NP_706646.1,,O3K_14090,1,3,1.0,3.0,1.0,3.0
CDS_NC_018658.1_2189610_2193006_pos,NC_018658.1,O3K_RS10615,WP_000515718.1,,2189610,2193006,pos,,host specificity protein J,,COORDINATES: similar to AA sequence:RefSeq:NP_706646.1,,O3K_10565,1,3,1.0,3.0,1.0,3.0
CDS_NC_018658.1_3586294_3589792_neg,NC_018658.1,O3K_RS17620,WP_000515639.1,,3586294,3589792,neg,,host specificity protein J,,COORDINATES: similar to AA sequence:RefSeq:NP_308869.1,,O3K_17525,1,3,1.0,3.0,1.0,3.0
CDS_NC_018666.1_55668_59763_pos,NC_018666.1,O3K_RS26470,WP_001034126.1,sepA,55668,59763,pos,,serine protease autotransporter toxin SepA,,COORDINATES: similar to AA sequence:RefSeq:NP_858203.1,,O3K_26432,2,2,,,
CDS_NC_018666.1_60057_60246_neg,NC_018666.1,O3K_RS26475,,,60057,60246,neg,True,transposase domain-containing protein,,COORDINATES: similar to AA sequence:RefSeq:WP_000080195.1,,O3K_26437,2,2,,,
CDS_NC_018658.1_4587764_4588589_pos,NC_018658.1,O3K_RS22405,WP_001114712.1,,4587764,4588589,pos,,MurR/RpiR family transcriptional regulator,,COORDINATES: similar to AA sequence:RefSeq:WP_019077578.1,,O3K_22225,3,2,,,
CDS_NC_018658.1_4588760_4589458_neg,NC_018658.1,O3K_RS22410,,,4588760,4589458,neg,True,IS1 family transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_076612230.1,,O3K_22230,3,2,,,
CDS_NC_018658.1_4178987_4179758_pos,NC_018658.1,O3K_RS20520,WP_001118029.1,yafV,4178987,4179758,pos,,2-oxoglutaramate amidase,,COORDINATES: similar to AA sequence:RefSeq:NP_308273.1,3.5.1.111,O3K_20365,4,3,,,
CDS_NC_018658.1_4181038_4182165_neg,NC_018658.1,O3K_RS20530,,,4181038,4182165,neg,True,ISAs1 family transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_000027427.1,,,4,3,,,
CDS_NC_018658.1_4179799_4180936_neg,NC_018658.1,O3K_RS20525,,,4179799,4180936,neg,True,ISAs1 family transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_000420980.1,,O3K_20370,4,3,,,
CDS_NC_018666.1_7762_7945_pos,NC_018666.1,O3K_RS26150,,,7762,7945,pos,True,IS1 family transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_012477181.1,,,5,4,,,
CDS_NC_018666.1_6864_7662_pos,NC_018666.1,O3K_RS26145,WP_000769457.1,aggR,6864,7662,pos,,aggregative adherence transcriptional regulator AggR,,COORDINATES: similar to AA sequence:RefSeq:WP_000769457.1,,O3K_26097,5,4,,,
CDS_NC_018666.1_8409_9413_neg,NC_018666.1,O3K_RS26160,,,8409,9413,neg,True,IS110 family transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_013315785.1,,O3K_26107,5,4,,,
CDS_NC_018666.1_7945_8134_pos,NC_018666.1,O3K_RS26155,,,7945,8134,pos,True,transposase,,COORDINATES: similar to AA sequence:RefSeq:WP_076611134.1,,,5,4,,,
CDS_NC_018658.1_2170073_2170880_pos,NC_018658.1,O3K_RS10480,WP_000731196.1,,2170073,2170880,pos,,YdFR family protein,,COORDINATES: similar to AA sequence:RefSeq:WP_000731189.1,,O3K_10425,6,2,2.0,2.0,2.0,2.0
CDS_NC_018658.1_2921053_2921860_neg,NC_018658.1,O3K_RS14285,WP_000731197.1,,2921053,2921860,neg,,YdFR family protein,,COORDINATES: similar to AA sequence:RefSeq:WP_000731189.1,,O3K_14235,6,2,2.0,2.0,2.0,2.0
CDS_NC_018658.1_3619388_3620174_pos,NC_018658.1,O3K_RS17850,WP_000100847.1,bet,3619388,3620174,pos,,phage recombination protein Bet,,COORDINATES: similar to AA sequence:RefSeq:NP_311028.1,,O3K_17755,7,2,3.0,2.0,,
CDS_NC_018658.1_2528195_2528981_neg,NC_018658.1,O3K_RS12360,WP_000100829.1,bet,2528195,2528981,neg,,phage recombination protein Bet,,COORDINATES: similar to AA sequence:RefSeq:NP_309202.1,,O3K_12320,7,2,3.0,2.0,,
CDS_NC_018666.1_21618_22377_neg,NC_018666.1,O3K_RS26240,WP_000702455.1,aggD,21618,22377,neg,,aggregative adherence fimbria I chaperone AggD,,COORDINATES: similar to AA sequence:RefSeq:WP_032154288.1,,O3K_26212,8,2,,,
CDS_NC_018666.1_22824_23059_pos,NC_018666.1,O3K_RS26245,,,22824,23059,pos,True,hypothetical protein,,COORDINATES: similar to AA sequence:RefSeq:WP_001366937.1,,,8,2,,,
CDS_NC_018658.1_2151814_2151993_neg,NC_018658.1,O3K_RS10330,,,2151814,2151993,neg,True,hypothetical protein,,COORDINATES: similar to AA sequence:RefSeq:NP_305741.1,,,9,2,,,
CDS_NC_018658.1_2150722_2151481_neg,NC_018658.1,O3K_RS10325,WP_000628772.1,,2150722,2151481,neg,,phage protein,,COORDINATES: similar to AA sequence:RefSeq:WP_000628772.1,,O3K_10290,9,2,,,

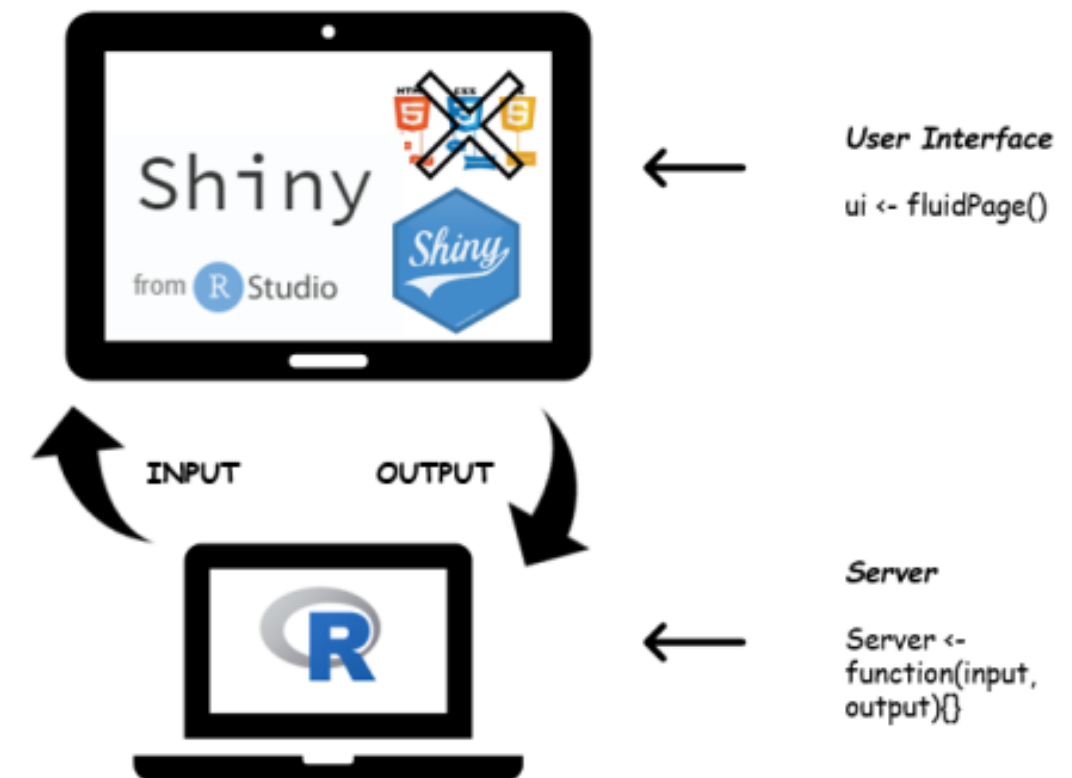
```





- Descarte de mysql y motor bbdd
- Descarte html, php, mysql
- framework python django / flask
- framework R shiny (ui,server)
- shiny + (css,js)

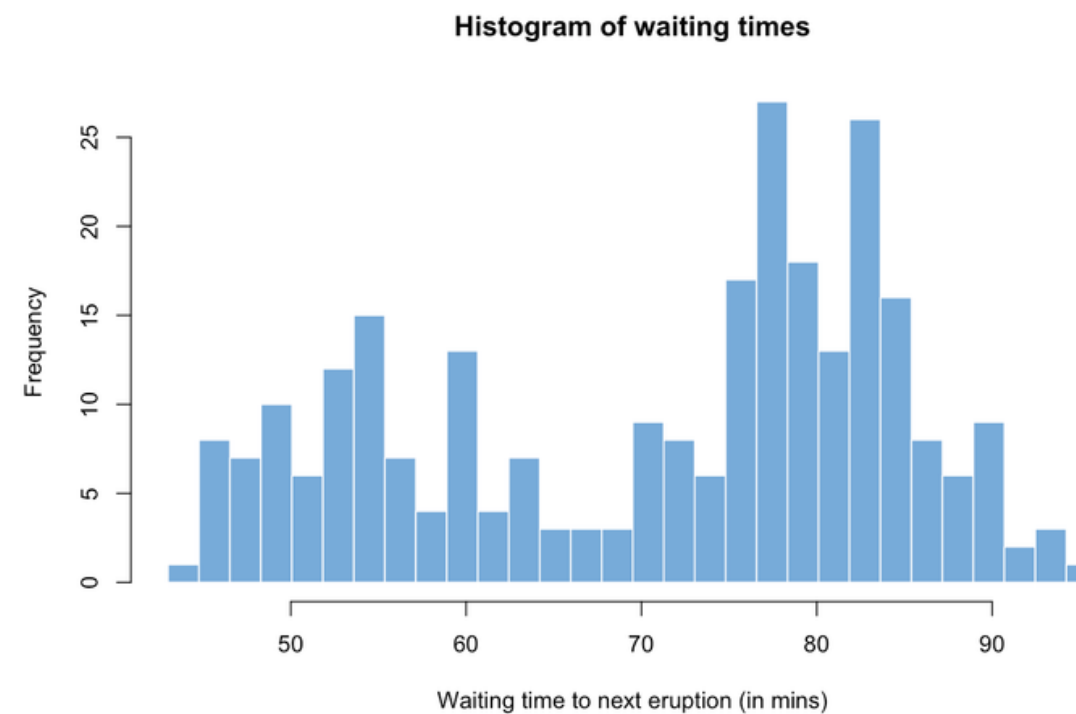
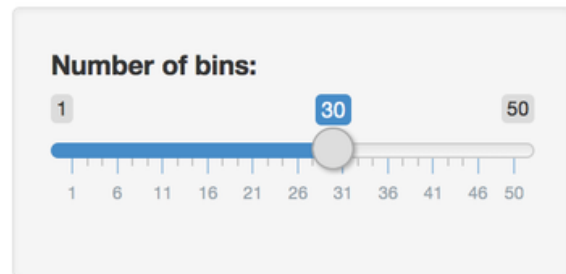
- frontend: Interfaz usuario  
inputs y outputs
- backend: funciones y cálculos
- bbdd: motor de bases de datos





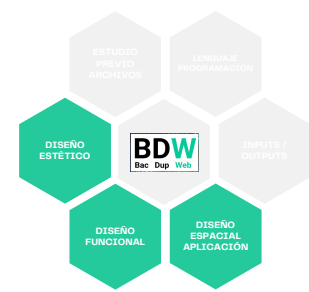
http://127.0.0.1:4397 [Open in Browser](#) [Publish](#)

Hello Shiny!



- Shiny ui/server y shinyapp()
- Shiny estética estándar
- Shiny reactivo
- La facilidad inicial de Shiny. Histograma con bins
- La complejidad de avanzar con Shiny
- Shiny con html, css y js





**User:** Daniba  
**Session:** 1.4  
3/4/2022

Choose annot\_dup file:  
   
Upload complete

Choose length file:  
   
Upload complete

---

Select Sequence / plasmid:  
 NC\_018658.1  
 NC\_018666.1  
 NC\_018659.1  
 NC\_018660.1

n\_dups Range:  
  
2 3 4 5 6 7 8 9 10 11

select strand

include pseudogenes

### Original Data

Name	Value
Records :	612
max_dups :	11
min_dups :	2

### Filtered Data

Name	Value
Records :	612
max_dups :	11
min_dups :	2

LINK NCBI:  
GCF\_000299455

REPORT

Table
Circos
Plasmids
Data

Show  entries Search:

	rec_id	gene	strand	pseudo	product	dup_id	count_dups
1	NC_018658.1		neg		host specificity protein J	1	3
2	NC_018658.1		pos		host specificity protein J	1	3
3	NC_018658.1		neg		host specificity protein J	1	3
4	NC_018666.1	sepA	pos		serine protease autotransporter toxin SepA	2	2
5	NC_018666.1		neg	True	transposase domain-containing protein	2	2





Choose annot\_dup file:

Browse... GCF\_000299455\_dup\_anno

Upload complete

Choose length file:

Browse... No file selected

Go

Abrir

> Este equipo > Windows (C:) > TFM > SHINY > BDW1 > BacDupWeb8 > arxius

Organizar Nueva carpeta

Nombre	Fecha de modificación	Tipo	Tamaño
GCF_000010765.1	08/05/2022 18:36	Carpeta de archivos	
GCF_000017765.1	08/05/2022 18:36	Carpeta de archivos	
GCF_000026245.1	08/05/2022 18:36	Carpeta de archivos	
GCF_000299455.1	08/05/2022 18:36	Carpeta de archivos	
GCF_000829985.1	08/05/2022 18:36	Carpeta de archivos	
GCF_000010765_dup_annot.csv	25/03/2022 22:56	Archivo de valores se...	201 KB
GCF_000010765_length_df.csv	29/03/2022 17:57	Archivo de valores se...	1 KB
GCF_000017765_dup_annot.csv	05/04/2022 21:26	Archivo de valores se...	62 KB
GCF_000017765_length_df.csv	29/03/2022 17:57	Archivo de valores se...	1 KB
GCF_000026245_dup_annot.csv	30/03/2022 17:10	Archivo de valores se...	101 KB
GCF_000026245_length_df.csv	29/03/2022 17:57	Archivo de valores se...	1 KB
GCF_000299455_dup_annot.csv	05/04/2022 21:29	Archivo de valores se...	123 KB
GCF_000299455_length_df.csv	29/03/2022 17:57	Archivo de valores se...	1 KB
GCF_000829985_dup_annot.csv	30/03/2022 17:17	Archivo de valores se...	85 KB
GCF_000829985_length_df.csv	29/03/2022 17:57	Archivo de valores se...	1 KB

Nombre: GCF\_000299455\_length\_df.csv

Archivos personalizados (\*.csv;\*.csv)

Abrir Cancelar





Go

Select Sequence / plasmid:

- NC\_018658.1
- NC\_018666.1
- NC\_018659.1
- NC\_018660.1

n\_dups Range:

2 11

2 3 4 5 6 7 8 9 10 11

select strand

both

- include pseudogenes
- include mobile elements

Filtros condicionados por la cepa

Campos de la tabla configurables en tiempo real

Filtros fijos

compact view

Select Fields:

- X
- rec\_id
- locus\_tag
- protein\_id
- gene
- start
- end
- strand
- pseudo
- product
- Dbxref
- inference
- EC\_number
- old\_locus\_tag
- dup\_id
- count\_dups
- dup\_id\_pseudo\_free
- count\_dups\_pseudo\_free
- dup\_id\_mobile\_free
- count\_dups\_mobile\_free



Upload complete

Go

Select Sequence / plasmid:

- NC\_018658.1
- NC\_018666.1
- NC\_018659.1
- NC\_018660.1

n\_dups Range:

2 11

select strand

neg

- include pseudogenes
- include mobile elements

compact view

Select Fields:

- X
- rec\_id
- locus\_tag
- protein\_id
- gene
- start
- end
- strand
- pseudo
- product

max\_dups : 11  
min\_dups : 2

max\_dups : 11  
min\_dups : 2

REPORT

Table    **Circos**    Plasmids    Data

Show 25 entries

Search:

rec_id	gene	strand	pseudo	product	dup_id	count_dups	
5	NC_018666.1	neg	True	transposase domain-containing protein	2	2	
7	NC_018658.1	neg	True	IS1 family transposase	3	2	
9	NC_018658.1	neg	True	ISAs1 family transposase	4	3	
10	NC_018658.1	neg	True	ISAs1 family transposase	4	3	
13	NC_018666.1	neg	True	IS110 family transposase	5	4	
18	NC_018658.1	bet	neg	phage recombination protein Bet	7	2	
19	NC_018666.1	aggD	neg	aggregative adherence fimbria I chaperone AggD	8	2	
21	NC_018658.1	neg	True	hypothetical protein	9	2	
22	NC_018658.1	neg		phage protein	9	2	
27	NC_018666.1	traX	neg	conjugal transfer pilus acetylase TraX	12	4	
29	NC_018666.1	tral	neg	True	conjugative transfer relaxase/helicase Tral	12	4
30	NC_018666.1	neg	True	PAS domain-containing protein	12	4	
32	NC_018658.1	neg		hypothetical protein	13	2	
36	NC_018658.1	neg		IS6-like element IS26 family transposase	15	3	
39	NC_018658.1	neg		phage tail protein	16	3	



Go

---

Select Sequence / plasmid:

- NC\_018658.1
- NC\_018666.1
- NC\_018659.1
- NC\_018660.1

n\_dups Range:

2  11

select strand

both

- include pseudogenes
- include mobile elements

---

compact view

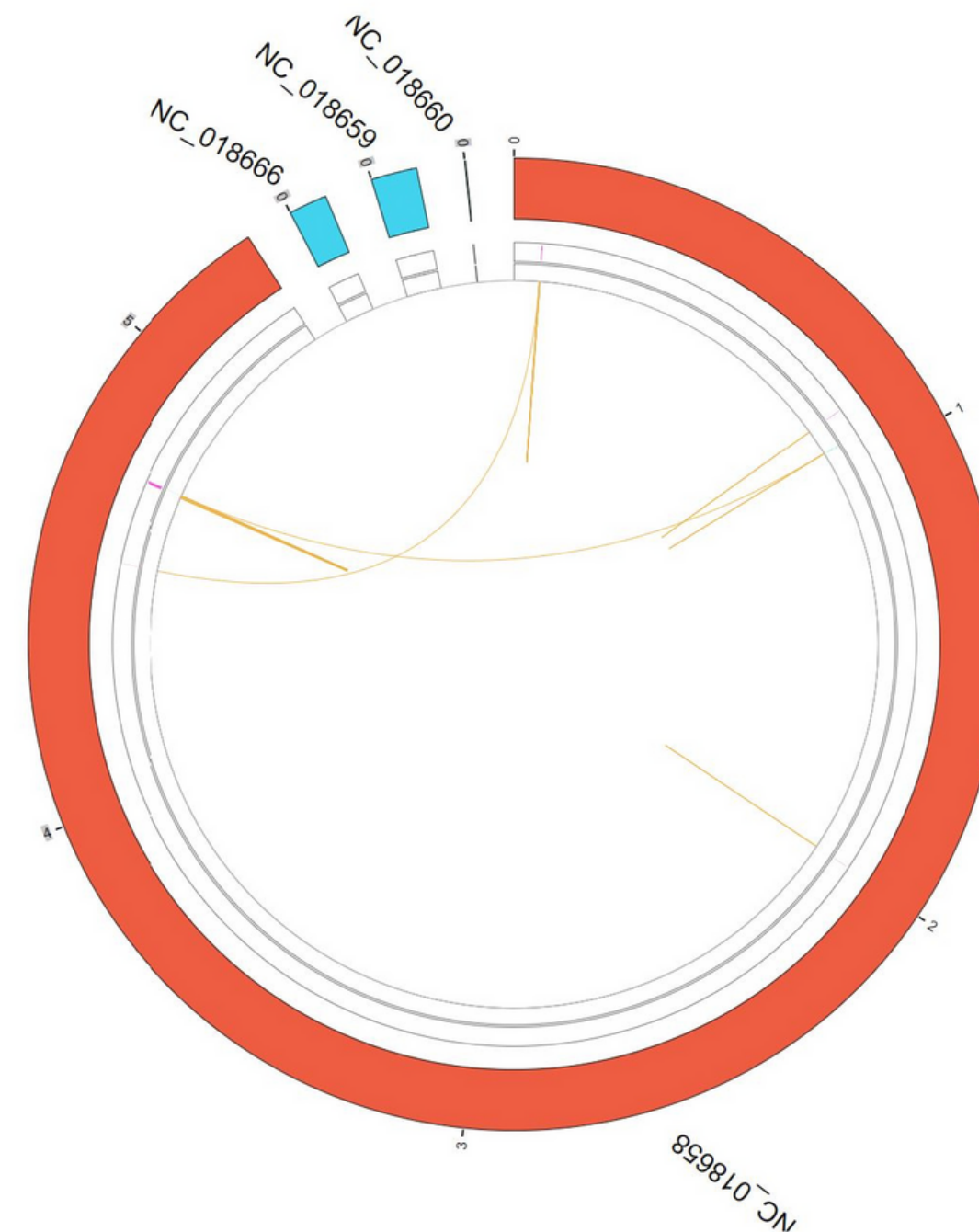
Select Fields:

- X
- rec\_id
- locus\_tag
- protein\_id
- gene
- start
- end

min\_dups : 2

min\_dups : 7

- Table
- Circos**
- Plasmids
- Data







Choose annot\_dup file:  
 GCF\_000299455\_dup\_annr

Choose length file:  
 GCF\_000299455\_length\_df

---

Select Sequence / plasmid:  
 NC\_018658.1  
 NC\_018666.1  
 NC\_018659.1  
 NC\_018660.1

n\_dups Range:  
  
 2 3 4 5 6 7 8 9 10 11

select strand

include pseudogenes  
 include mobile elements

Original Data

Name	Value
Records :	612
max dups :	11
min_dups :	2

Filtered Data

Name	Value
Records :	431
max dups :	11
min_dups :	2

[LINK NCBI:](#)  
GCF\_000299455

[REPORT](#)

- Table
- Circos
- Plasmids
- Data

Taxonomy and Chromosome table

Taxonomy
Escherichia coli O104:H4 str. 2011C-3493
Bacteria
Proteobacteria
Gammaproteobacteria
Enterobacterales
Enterobacteriaceae
Escherichia

Id sequence	Size	plasmid/Chr
NC_018658.1	5273097	Chromosome
NC_018659.1	88544	Plasmid
NC_018666.1	74217	Plasmid
NC_018660.1	1549	Plasmid



Choose annot\_dup file:  
 GCF\_000299455\_dup\_annr

Choose length file:  
 GCF\_000299455\_length\_df

---

Select Sequence / plasmid:  
 NC\_018658.1  
 NC\_018666.1  
 NC\_018659.1  
 NC\_018660.1

n\_dups Range:

select strand

include pseudogenes  
 include mobile elements

---

compact view

Select Fields:

Original Data

Name	Value
Records :	612
max dups :	11
min_dups :	2

Filtered Data

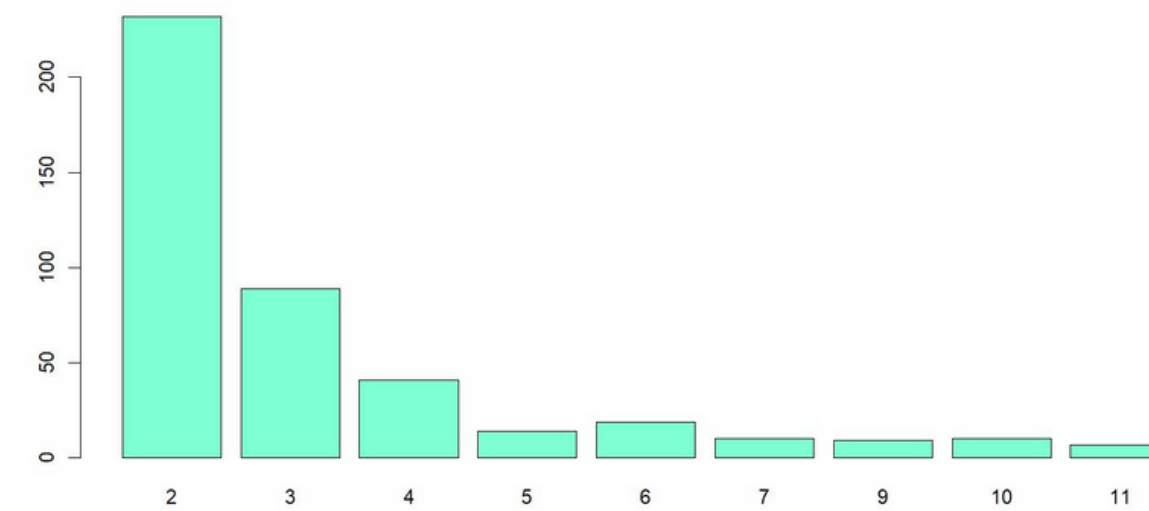
Name	Value
Records :	431
max dups :	11
min_dups :	2

LINK NCBI:  
[GCF\\_000299455](#)

- Table
- Circos
- Plasmids
- Data

N_Dups	Total	Real Dups
2	232	116
3	89	29
4	41	10
5	14	2
6	19	3
7	10	1
9	9	1
10	10	1
11	7	0

Distribution of duplicate genes





User: Daniba  
Session: 1.4  
3/4/2022

Filtered Data

Name	Value
Records :	612
max_dups :	11
min_dups :	2

LINK NCBI:  
GCF\_000299455

REPORT

Data

Search:

pseudo	product	dup_id	count_dups
	host specificity protein J	1	3
	host specificity protein J	1	3
	host specificity protein J	1	3
	serine protease autotransporter toxin SepA	2	2

U.S. Department of Health and Human Services

**NIH** National Library of Medicine  
National Center for Biotechnology Information

Search NCBI ...

Datasets / Assembly / ASM29945v1

## Genome assembly ASM29945v1

Download

Reference sequence	RefSeq GCF_000299455.1	⋮
Submitted sequence	GenBank GCA_000299455.1	⋮
Taxon	<i>Escherichia coli</i> O104:H4 str. 2011C-3493	
Strain	2011C-3493	
Submitter	Los Alamos National Laboratory	
Date	Sep 27, 2012	

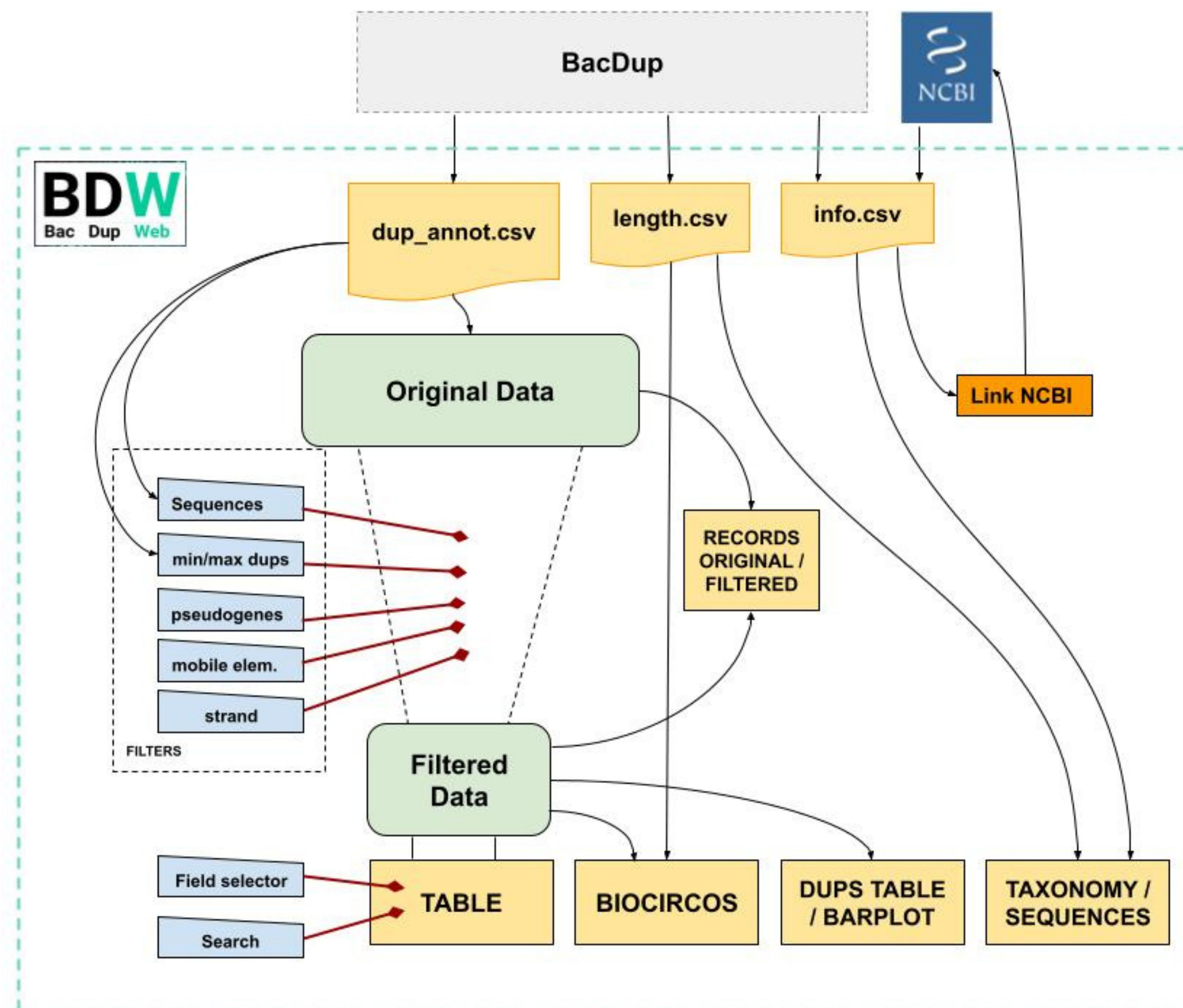
View the [legacy Assembly page](#)

### Assembly statistics

These statistics describe the nuclear genome of the reference sequence, GCF\_000299455.1

Genome size	5.4 Mb
Number of chromosomes	4
Number of scaffolds	4
Scaffold N50	5.2 Mb







- **Resultado final satisfactorio. Objetivo conseguido.**
- **Aplicación con 4 apartados y velocidad reactiva instantánea.**
- **Recopilatorio de conceptos y herramientas aprendidas del máster.**
- **Memoria, presentación, código R, github y shinyapps.**
  
- **Integrar BacDup en la aplicación web. Entorno Conda.**
- **Gestión de usuarios.**
- **Trabajo con varias cepas. Comparación de cepas.**
- **Aplicación web para el estudio de genomas de procariotas.**





[Bacdupweb en Github](#)

[BacDupWeb en Shinyapps.io](#)