

# Práctica 1

# GenBank

Laura Oliva - 702756

Irene Sánchez - 702692

# Índice

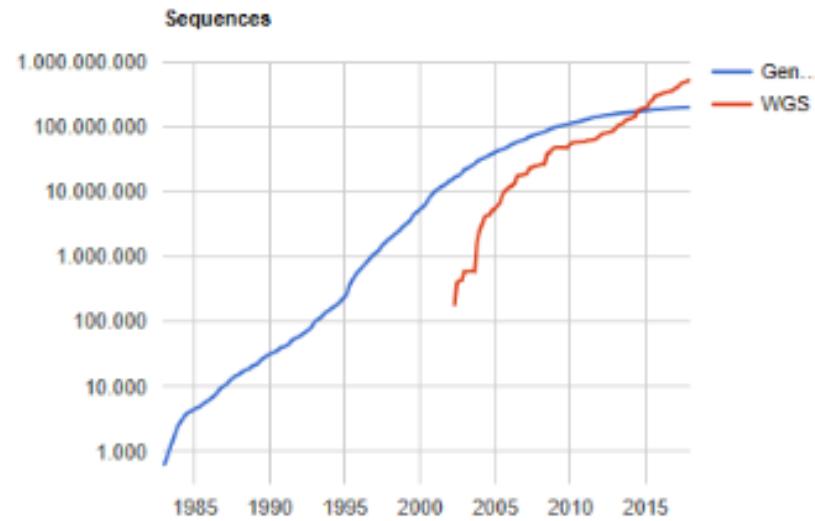
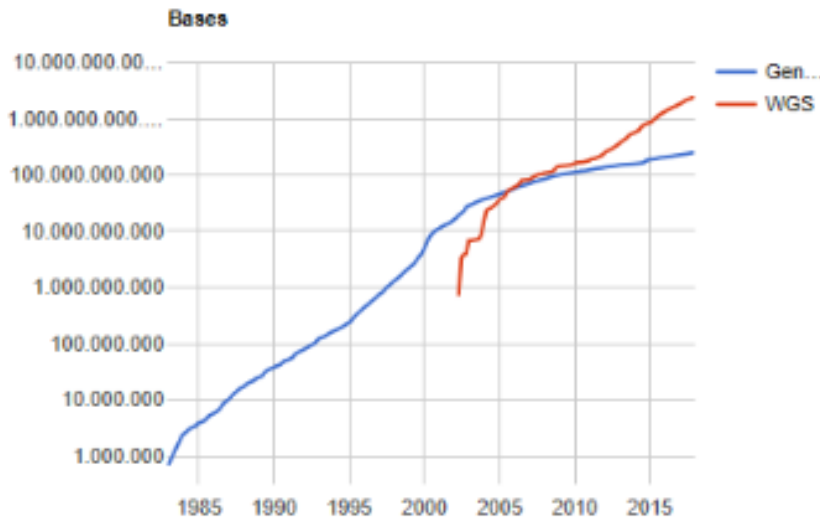
- ▶ ¿Qué es?
- ▶ Tamaño y actualizaciones
- ▶ Búsquedas simples
- ▶ Búsquedas avanzadas
- ▶ Resultados
- ▶ Filtros
- ▶ Guardar resultados

# ¿Qué es?

- ▶ Base de datos americana de secuencias genéticas del NIH (Instituto Internacional de Salud).
- ▶ Contiene las secuencias de ADN de acceso público. Incluye anotaciones.
- ▶ Es parte del INSDC.
- ▶ Su base de datos está dividida en 3:
  - ▶ NCBI
  - ▶ ENA
  - ▶ DDBJ
- ▶ Bases de datos actualizadas diariamente.

# Tamaño y actualizaciones

- ▶ Empezó en 1982: 606 secuencias y 680.338 bases nitrogenadas.
- ▶ En 2017: 206.293.625 secuencias y 249.722.163.594 bases nitrogenadas.
- ▶ El número de bases nitrogenadas se duplica cada 18 meses.
- ▶ Cada 2 meses sale una nueva versión.



# Búsquedas simples

Nucleotide  Search

All

- All Databases
- Assembly
- BioCollections
- BioProject
- BioSample
- BioSystems
- Books
- ClinVar
- Clone
- Conserved Domains
- dbGaP
- dbVar
- EST**
- Gene
- Genome
- GEO DataSets
- GEO Profiles
- GSS
- GTR

HomoloGene

- Identical Protein Groups
- MedGen
- MeSH
- NCBI Web Site
- NLM Catalog
- Nucleotide
- OMIM
- PMC
- PopSet
- Probe
- Protein
- Protein Clusters
- PubChem BioAssay
- PubChem Compound
- PubChem Substance
- PubMed
- PubMed Health
- SNP
- Sparcle

SRA

- Structure
- Taxonomy
- ToolKit
- ToolKitAll
- ToolKitBookgh**
- UniGene

# Búsqueda avanzada

Builder

All Fields ▾ genomic dna ⊖ [Hide index list](#)

- genomic dna (1534423)
- genomic dna 1218 1 (2)
- genomic dna 5 (3)
- genomic dna amplification (78044)
- genomic dna amplification products (29)
- genomic dna analyses (5)
- genomic dna analysis (38)
- genomic dna bac (95)
- genomic dna bac library (6)
- genomic dna bacteriophage p1 genome systems, inc (1)

[Previous 200](#)  
[Next 200](#)  
[Refresh index](#)

**AND**  
**OR**  
**NOT**

AND ▾ All Fields ▾ ⊖ ⊕ [Show index list](#)

**Search** or [Add to history](#)

Accession	Gene Name	Page Number
All Fields	Isolate	Primary Accession
Assembly	Issue	Primary Organism
Author	Journal	Properties
BioProject	Keyword	Protein Name
BioSample	Modification Date	Publication Date
Breed	Organism	SeqID String
Component Accession		Sequence Length
Cultivar		Strain
Division		Substance Name
EC/RN Number		Text Word
Feature key		Title
Filter		Volume

# Resultados

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide all[filter] Search

Create alert Advanced

Species  
Animals (133,008,729)  
Plants (38,937,698)  
Fungi (7,194,962)  
Protists (5,665,863)  
Bacteria (39,129,221)  
Archaea (624,770)  
Viruses (2,668,826)  
Customize ...

Molecule types  
genomic  
DNA/RNA (133,354,539)  
mRNA (45,166,934)  
rRNA (182,320)  
Customize ...

Source databases  
INSDC (GenBank)  
(206,613,288)  
RefSeq (48,574,371)  
Customize ...

Genetic compartments  
Chloroplast (1,064,864)

Summary 20 per page Sort by Default order

Format  
● Summary  
○ GenBank  
○ GenBank (full)  
○ FASTA  
○ FASTA (text)  
○ ASN.1  
○ Revision History  
○ Accession List  
○ GI List

Send to: Filters: Manage Filters

Find related data  
Database: Select

Find items

Search details  
all[filter]

Search See more...

Recent activity  
Turn Off Clear

all[filter] (255450202)

Items: 1 to 20 of 255450202

Found 372398013 nucleotide sequences. Nucleotide (255450202) EST (76973976) GSS (39973835)

- [Homo sapiens Sharpr-MPRA regulatory region 9595 \(LOC111982872\) on chromosome 10](#)
  - 495 bp linear DNA  
Accession: NG\_056365.1 GI: 1343071312  
[GenBank](#) [FASTA](#) [Graphics](#)
- [Homo sapiens Sharpr-MPRA regulatory region 7934 \(LOC111982871\) on chromosome 10](#)
  - 495 bp linear DNA  
Accession: NG\_056364.1 GI: 1343071311  
[GenBank](#) [FASTA](#) [Graphics](#)
- [Homo sapiens Sharpr-MPRA regulatory region 14427 \(LOC111982870\) on chromosome 10](#)
  - 495 bp linear DNA  
Accession: NG\_056363.1 GI: 1343071310  
[GenBank](#) [FASTA](#) [Graphics](#)

# Resultados

LOCUS AM282870 718 bp DNA linear PLN 27-JAN-2010  
DEFINITION Prunus armeniaca chloroplast atpB-rbcL intergenic spacer, cultivar Alfred.  
ACCESSION AM282870  
VERSION AM282870.1  
KEYWORDS atpB-rbcL intergenic spacer.  
SOURCE chloroplast Prunus armeniaca (apricot)  
ORGANISM Prunus armeniaca  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Rosales; Rosaceae; Maloideae; Amygdaleae; Prunus.

REFERENCE 1  
AUTHORS Reales,A., Sargent,D.J., Tobutt,K.R. and Rivera,D.  
TITLE Phylogenetics of Eurasian plums, Prunus L. section Prunus (Rosaceae), according to coding and non-coding chloroplast DNA sequences  
JOURNAL Tree Genet. Genomes 6 (1), 37-45 (2009)  
REFERENCE 2 (bases 1 to 718)  
AUTHORS Sargent,D.J.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2006) Sargent D.J., Plant Breeding and Genetics, East Malling Research, New Road, East Malling, Kent, ME19 6BJ, UNITED KINGDOM

FEATURES Location/Qualifiers  
source 1..718  
/organism="Prunus armeniaca"  
/organelle="plastid:chloroplast"  
/mol\_type="genomic DNA"  
/cultivar="Alfred"  
/db\_xref="taxon:36596"  
misc feature 1..718  
/note="atpB-rbcL intergenic spacer"

ORIGIN  
1 ggatttattc tcatataaaa tatgtcgaaa tttttgcca aaattactga aatcaaaaat  
61 aaattcgata gcaaaacaag ttaagttgat cgattaattc aataagaaat gggcttagc  
121 gctcgatttc gttggtacca tccaactgaa tccaattcaa ttgtttactt attcagtgaa  
181 ttgaaaaaatt caaacaaaac ccatttccaa aatagcaagt gtatgaataa aaattttgag  
241 aaagtctttt atttgccat cattatagac aataccttcc atattatcta tggaattcga  
301 accccattta cgatttcttt tttctatctc attggtcctt atttacgata tcagcatatc



# Filtros I

## Species

Animals (133,008,729)  
Plants (38,937,698)  
Fungi (7,194,962)  
Protists (5,665,863)  
Bacteria (39,129,221)  
Archaea (624,770)  
Viruses (2,668,826)  
Customize ...

## Molecule types

genomic  
DNA/RNA (133,354,539)  
mRNA (45,166,934)  
rRNA (182,320)  
Customize ...

## Enzyme types

Hydrolases  
Isomerases  
Ligases  
Lyases  
Oxidoreductases  
Transferases

## Source databases

INSDC (GenBank)  
(206,613,288)  
RefSeq (48,574,371)  
Customize ...

## Genetic compartments

Chloroplast (1,064,864)  
Mitochondrion (4,501,317)  
Plasmid (175,448)  
Plastid (1,170,708)

## Sequence length

Custom range...

## Release date

Custom range...

## Revision date

Custom range...

## Search fields

Choose ...


## Results by taxon

### Top Organisms [\[Tree\]](#)

Homo sapiens (15542647)  
Escherichia coli (1594245)  
uncultured bacterium (771548)  
Acinetobacter baumannii (698696)  
Klebsiella pneumoniae (555581)  
All other taxa (7173513)

More...

# Filtros II


**Find related data** 


Database:

Option:

Link to r


- Component Of
- Annotated Genomic Components (Core)
- Identical RefSeq
- Order cDNA clone
- RNA
- Genomic Neighbours
- Identical GenBank Sequence
- RefSeq Genome Sequences
- Other INSDC Genome Sequences
- All segments from this genome

**Recent** 

 allf

ar

tid

**Find related data** 

Database:

Option:

RNA sequence records that are annotated on the current genomic nucleotide sequences. The RNA records represent aligned or predicted transcripts of the current records.

# Guardar resultados

Send to: ▾ **Filters:** [Manage Filters](#)

Complete Record  
 Coding Sequences  
 Gene Features

**Choose Destination**

File  Clipboard  
 Collections

Download 26336230 items.

Format  
Summary ▾

Sort by  
Default order ▾

Create File

on  
[Tree]  
(15542647)  
li (1594245)  
terium (771548)  
baumannii (6986  
umoniae (55558:  
7173513)

ta  
otide

RNA sequence records that are ar

**Format**  
Summary ▾  
Summary  
GenBank  
GenBank (full)  
FASTA  
ASN.1  
XML  
INSDSeq XML  
TinySeq XML  
Feature Table  
Accession List  
GI List  
GFF3

**Sort by**  
Default order ▾  
Default order  
Accession  
Date Modified  
Date Released  
Organism Name  
Taxonomy ID  
Sequence Length

# Preguntas

