



I Curso de Bioinformática de la Estación Experimental de La Mayora

sol genomics network

Boyce Thompson Institute for Plant Research
Tower Road
Ithaca, New York 14853-1801
U.S.A.

by

Aureliano Bombarely Gomez



MATERIAL DEL CURSO:

***ftp://ftp.solgenomics.net/user_requests/
aubombarely/courses/BioinfoLaMayora/***



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Ux y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Ux y líneas de comandos.

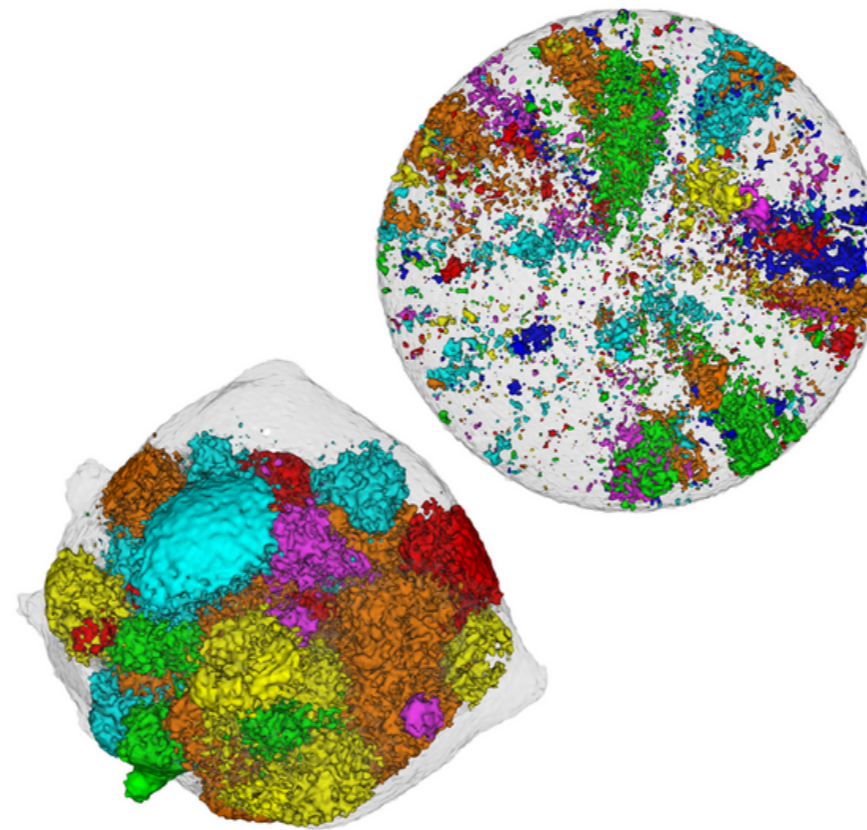
15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Introducción a la Bioinformática



Predicted distribution of methylation patterns in a non-hierarchical versus a cancer stem cell driven tumor.

Image Credit: Andrea Sottoriva, Cambridge University, United Kingdom

Source: PLOS Computational Biology





Introducción a la Bioinformática:

1. ¿ Qué es la bioinformática ?
2. Campos de estudio.
3. Herramientas utilizadas.
4. Perspectivas de futuro.





Introducción a la Bioinformática:

1. ¿ Qué es la bioinformática ?
2. Campos de estudio.
3. Herramientas utilizadas.
4. Perspectivas de futuro.



I. ¿ Qué es la bioinformática ?



Bioinformática:

**DATOS BIOLÓGICOS, HERRAMIENTAS
COMPUTACIONALES, VISUALIZACIÓN, ANÁLISIS,
ALGORITMOS, PROGRAMACIÓN,
SECUENCIAS, ENSAMBLAJES, ANOTACIÓN, ANÁLISIS
FILOGENÉTICO, VARIABILIDAD, ANÁLISIS POBLACIONALES,
ANÁLISIS DE EXPRESION, ÓMICAS, GRAN ESCALA,
ANÁLISIS DE SEGREGACIÓN DE UN CARACTER FENOTÍPICO, MAPAS
GENÉTICOS...**



I. ¿ Qué es la bioinformática ?



Bioinformática:

Aplicación de estadística y ciencia computacional al campo de la biología molecular (wikipedia).

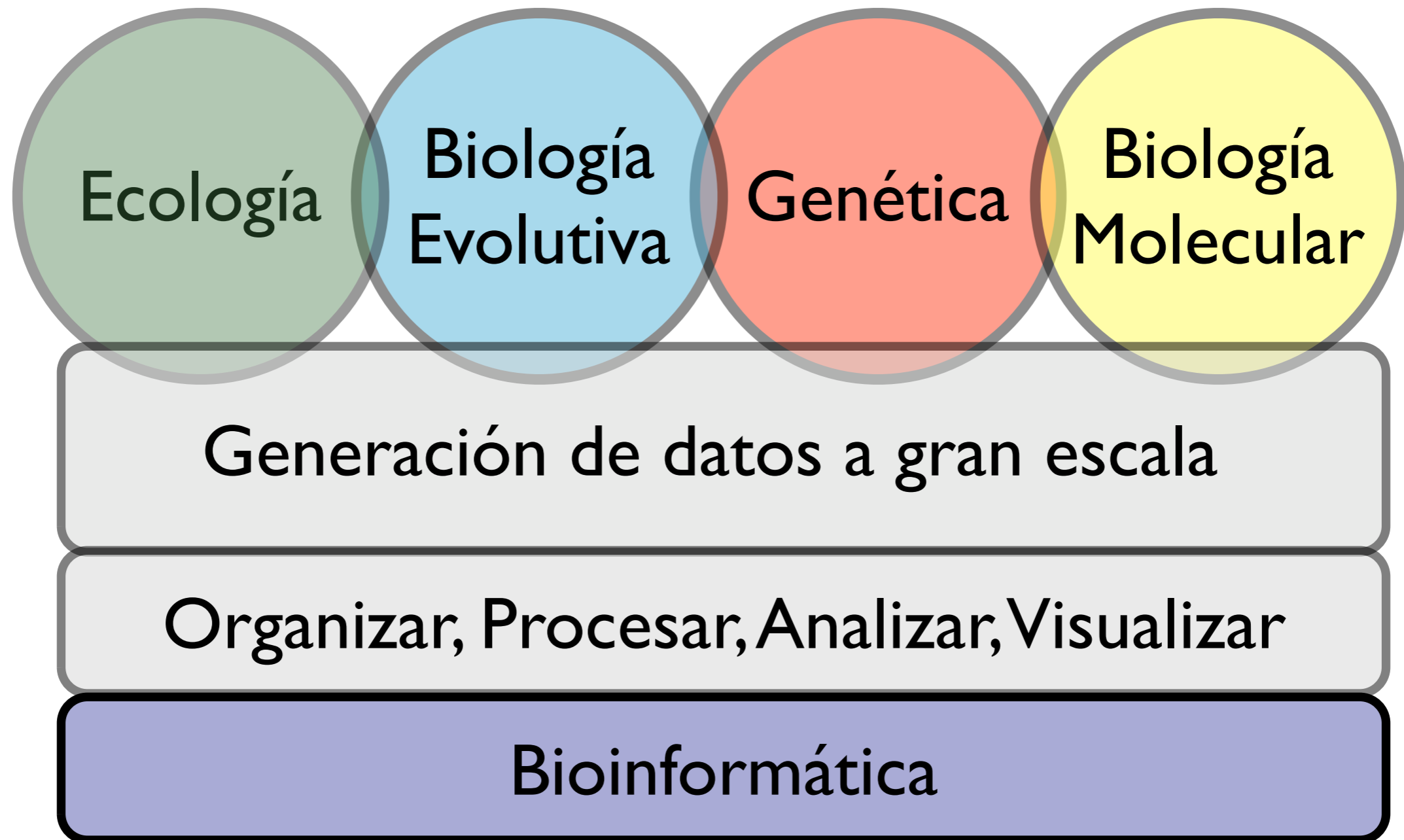
Conceptualización biológica en términos de macromoléculas (in el sentido físico-químico) y aplicación de técnicas “informáticas” (derivadas de otras disciplinas tales como matemáticas aplicadas, ciencia computacional y estadísticas) para entender y organizar la información asociada a estas moléculas, a gran escala. (Luscombe NM. et al. 2001)

Investigación, desarrollo o aplicación de herramientas computacionales y aproximaciones para expandir el uso de datos biológicos, médicos, de comportamiento o de salud, incluyendo aquellas herramientas para adquirir, guardar, organizar, analizar o visualizar tales datos (NIH, 2001).

I. ¿Qué es la bioinformática ?



Bioinformática:





Introducción a la Bioinformática:

1. ¿ Qué es la bioinformática ?
2. Campos de estudio.
3. Herramientas utilizadas.
4. Perspectivas de futuro.



2. Campos de estudio.



Campos de estudio → Desarrollo de herramientas para...

Biología Molecular:

Ensamblaje, anotación y análisis de secuencias.

Modelización estructural de moléculas complejas.

Análisis de interacciones biomoleculares.

Análisis de expresión y traducción génica.

Almacenamiento y administración de secuencias.

Genética:

Análisis de recombinación y ligamiento.

Representación de mapas de ligamiento.



2. Campos de estudio.

Campos de estudio → Desarrollo de herramientas para...

Biología evolutiva:

Creación de árboles filogenéticos.

Análisis de selección.

Análisis de duplicación de genes.

Ecología:

Almacenamiento de datos ecológicos.

Modelización poblacional.





Introducción a la Bioinformática:

1. ¿ Qué es la bioinformática ?
2. Campos de estudio.
- 3. Herramientas utilizadas.**
4. Perspectivas de futuro.



3. Herramientas utilizadas.



Actualmente existen miles de herramientas bioinformáticas distintas.
(385 páginas de software listadas en la wiki: <http://www.bioinformatics.org/wiki/Category:Software>)

Clasificación compleja, dependiente de muchas características:

- **Función.**

Ejemplo: “Búsqueda de Secuencias por Homología”

- **Lenguaje de programación.**

Ejemplo: C++

- **Organización/Grupo.**

Ejemplo: NCBI

- **Licencia y disponibilidad.**

Ejemplo: Software de dominio público.



3. Herramientas utilizadas.



“TOP TEN” DE HERRAMIENTAS BIOINFORMÁTICAS:

1. BLAST (Altschul SF. *et al.* 1990; 30,202 citaciones)
búsqueda de secuencias homologas.
2. CLUSTALW (Thompson JD. *et al.* 1994; 32,681 citaciones).
alineamiento múltiple de secuencias.
3. PAML (Yang ZH, 1997; 2,642 citaciones)
análisis filogenético mediante “Maximum Likelihood”
4. GBROWSE (Stein LD, *et al.* 2002; 428 citaciones),
visualización de genomas.
5. BLAST2GO (Conesa A. *et al.* 2005; 363 citaciones),
programa para anotación de secuencias.



“TOP TEN” DE HERRAMIENTAS BIOINFORMÁTICAS:

6. VELVET (Zerbino DR, *et al.* 2008; 323 citasiones),
Ensamblaje de secuencias mediante *Bruijn Graphs*.
7. SAMTOOLS (Li H. *et al.* 2009; 172 citasiones),
procesamiento de archivos de alineamiento de millones
secuencias con una secuencia de referencia
8. SOAP2 (Li RQ *et al.* 2009; 76 citasiones).
Alineamiento y ensamblaje de secuencias (*short reads*).
9. MAKER (Cantarel BL, *et al.* 2008; 23 citasiones),
Pipeline de anotación de genomas.
10. GALAXY (Goecks J. *et al.* 2010; 20 citasiones),
Plataforma web de análisis genómico.

3. Herramientas utilizadas.

Una gran parte de las herramientas utilizadas son pequeños ***scripts***, piezas de código escritas por programadores para realizar una tarea sencilla como convertir un formato en otro o traducir una secuencia de nucleótidos a proteína.

Los lenguajes de programación mas utilizados para la creación de *scripts* son Perl, Python, Java, Ruby y R.



3. Herramientas utilizadas.



Existen **módulos**, piezas de código prefabricadas por otros programadores para guardar funciones de manera que no sea necesario escribir la misma función desde cero. Por ejemplo, el módulo Bio::Seq de perl posee una función para traducir una secuencia de nucleótidos a proteína.

Estos lenguajes de programación poseen módulos específicos para trabajar con datos biológicos. El más conocido es Bioperl, pero también existen Biopython, Biojava, Bioruby y Bioconductor (R).

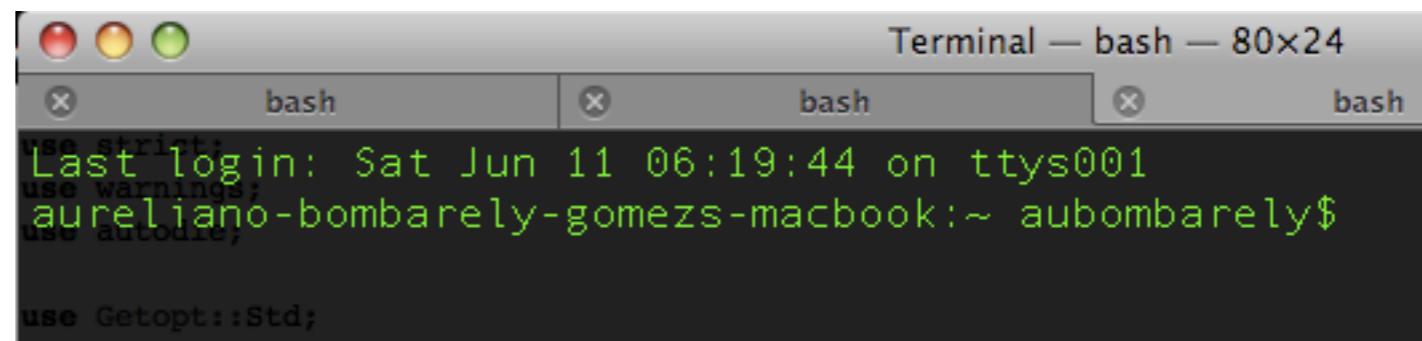


3. Herramientas utilizadas.



scripts,

```
95 use strict;
96 use warnings;
97 use autodie;
98
99 use Getopt::Std;
100 use Bio::SeqIO;
101
102 our ($opt_s, $opt_f, $opt_o, $opt_i, $opt_l, $opt_r, $opt_v, $opt_h);
103 getopts("s:f:o:i:l:r:vh");
104 if (!$opt_f && !$opt_o && !$opt_i && !$opt_l && !$opt_r && !$opt_v && !$opt_h) {
105     print "There are n't any tags. Print help\n\n";
106     help();
107 }
108 if ($opt_h) {
109     help();
110 }
111
112 ## Get the arguments and check them
113
114 my $infasta_file = $opt_s ||
115     die("INPUT ARGS ERROR: Input option was not supplied (-f <seq_file>).\n");
116
117 my $format = $opt_f || 'fasta';
118 my $id_file = $opt_i;
119 my $length_cf = $opt_l;
120 my $regexp;
121 if (defined $opt_r) {
122     $regexp = qr/$opt_r/;
123 }
```



3. Herramientas utilizadas.

A lo largo de un mismo análisis, pueden utilizarse más de una herramienta.

Ejemplo: Análisis filogenético de Receptores de Etileno.

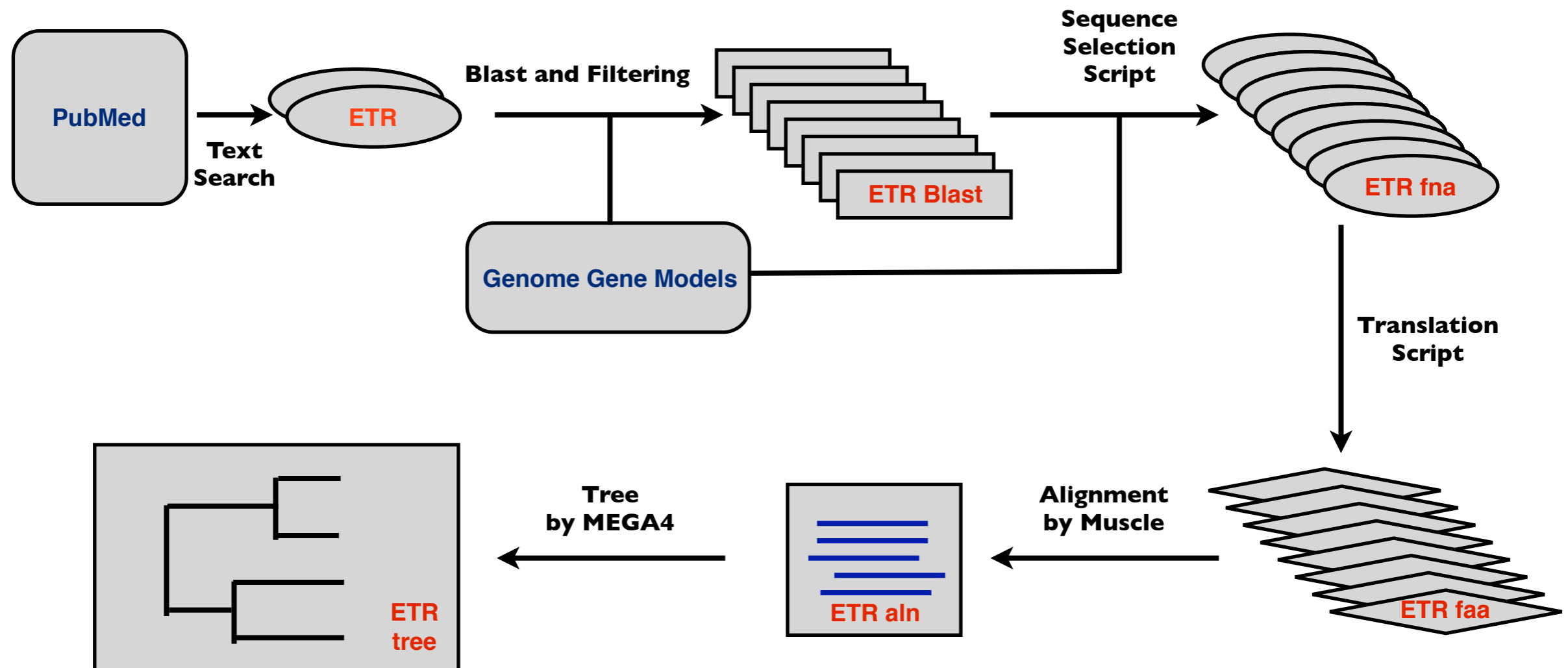
1. *Data Mining*: Búsqueda del receptor de etileno de arabidopsis y tomate por palabras clave, o búsqueda del ID de la base de datos a través de la literatura. (formato .fasta).
2. *Data Mining*: Búsqueda de secuencias homólogas usando blast en los genomas secuenciados. (formato .m8).
3. *Data Processing*: Filtro de los resultados de blast usando un script (formato .txt).
4. *Data Processing*: Toma de las secuencias homólogas de los genomas analizados usando un script. (formato .fasta).
5. *Data processing*: Traducción de las secuencias a proteínas usando un script. (formato .fasta).
6. *Data processing*: Alineamiento de secuencias usando Muscle (formato .aln).
7. *Data processing*: Generación del bootstrap y del árbol filogenético usando MEGA4 (formato .newick).
8. *Data visualizing*: Visualización del árbol filogenético.

3. Herramientas utilizadas.



A lo largo de un mismo análisis, pueden utilizarse más de una herramienta.

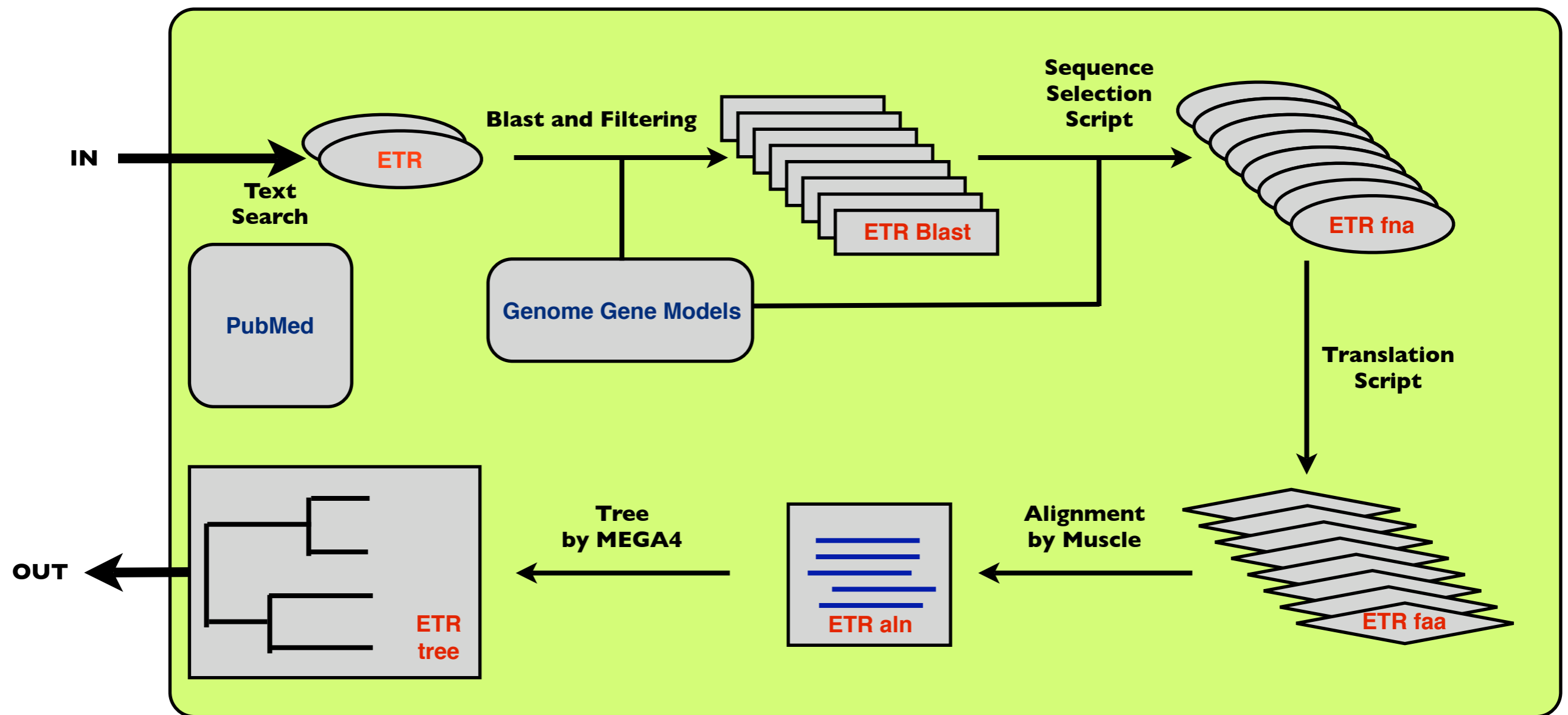
WORKFLOW: Análisis filogenético de Receptores de Etileno.



3. Herramientas utilizadas.



Si todas las herramientas se integran bajo el mismo programa o script se define como **Pipeline** (Set de Procesamiento de Datos Conectados en Serie)





Introducción a la Bioinformática:

1. ¿ Qué es la bioinformática ?
2. Campos de estudio.
3. Herramientas utilizadas.
4. Perspectivas de futuro.





4. Perspectivas de futuro.

Las nuevas tecnologías de secuenciación (***Next Generation Sequencing, NGS***) están produciendo enormes cantidades de datos genómicos y transcriptómicos.

Incremento de el poder de computación a traves del hardware standard (Uso de GPU en lugar de CPU)

Análisis integrativo donde se combinen distintos tipos de datos al mismo tiempo.

(Análisis de expresión y traducción de distintos genes en distintas especies y tejidos al mismo tiempo)

Secuenciación y análisis de variantes (incluidos mutaciones) en cuestión de horas.





PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Uinx y líneas de comandos.

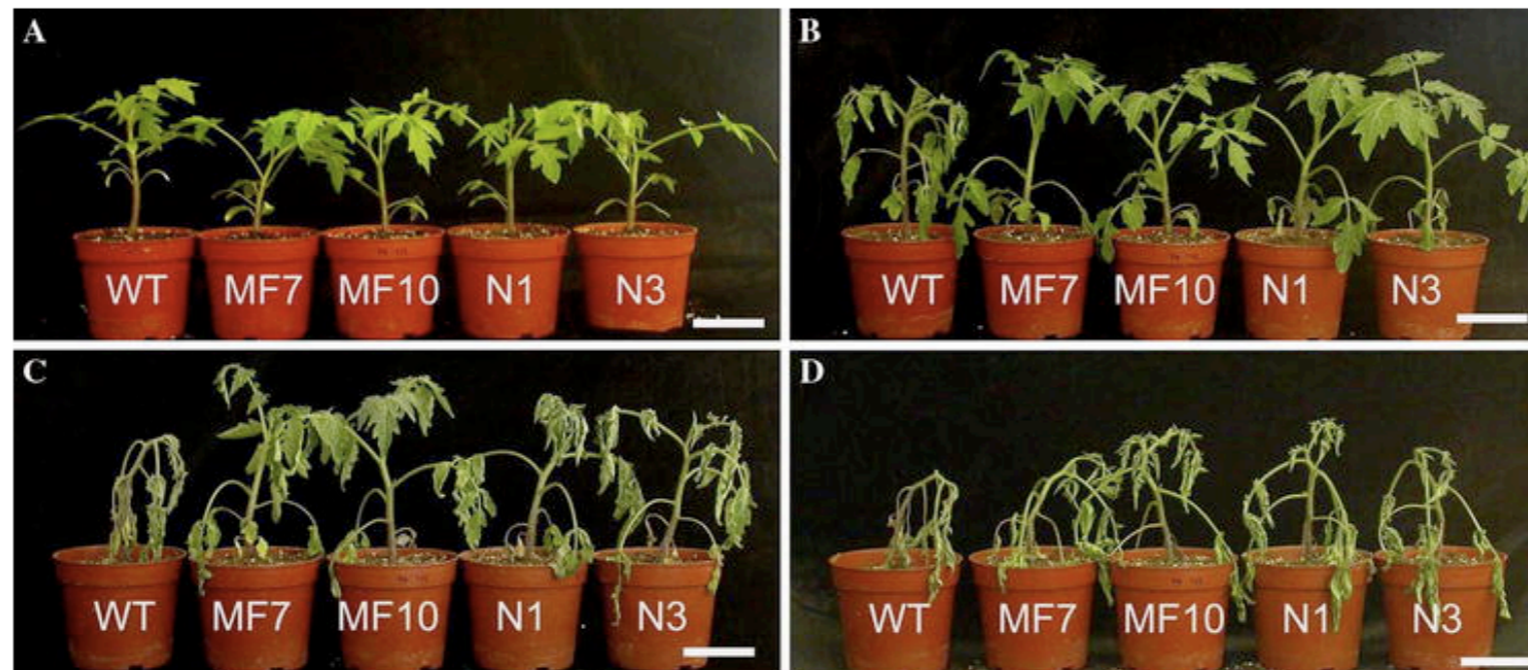
15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Caso de Estudio: Estrés por Sequía en Tomate.



Source: Zhang Y. et al. Generation of selectable marker-free transgenic tomato resistant to drought, cold and oxidative stress using the Cre/loxP DNA excision system. *Transgenic Research* 2009, 18 (4): 607-619





Caso de Estudio: Estrés por Sequía en Tomate

1. Background.
2. Diseño Experimental.
3. Material y Herramientas.





Caso de Estudio: Estrés por Sequía en Tomate

1. Background.

2. Diseño Experimental.

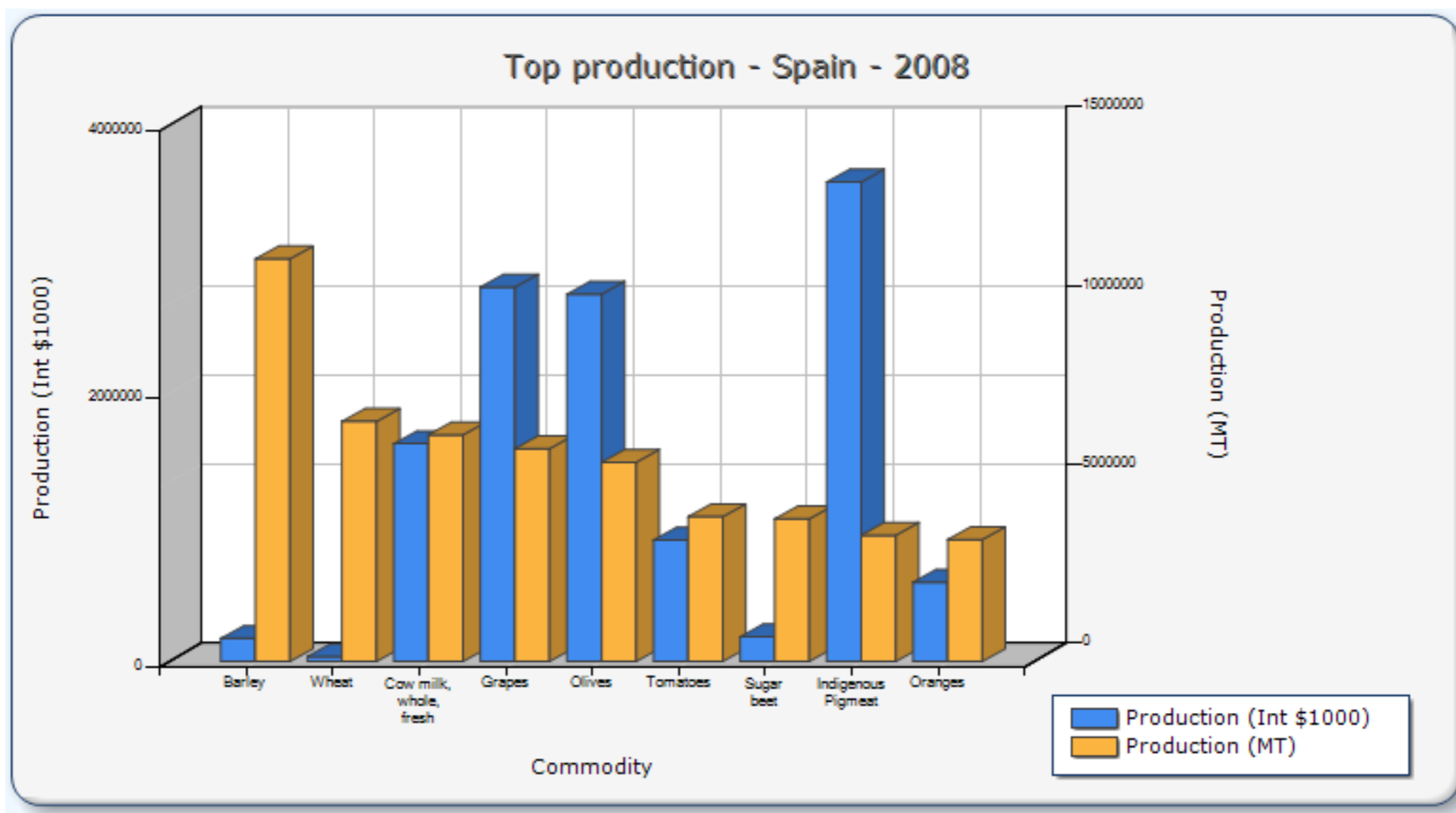
3. Material y Herramientas.





I) Importancia del Tomate.

El tomate es una de las plantas más cultivadas en el mundo, con más de 142 millones de toneladas (FAOSTAT, 2008).

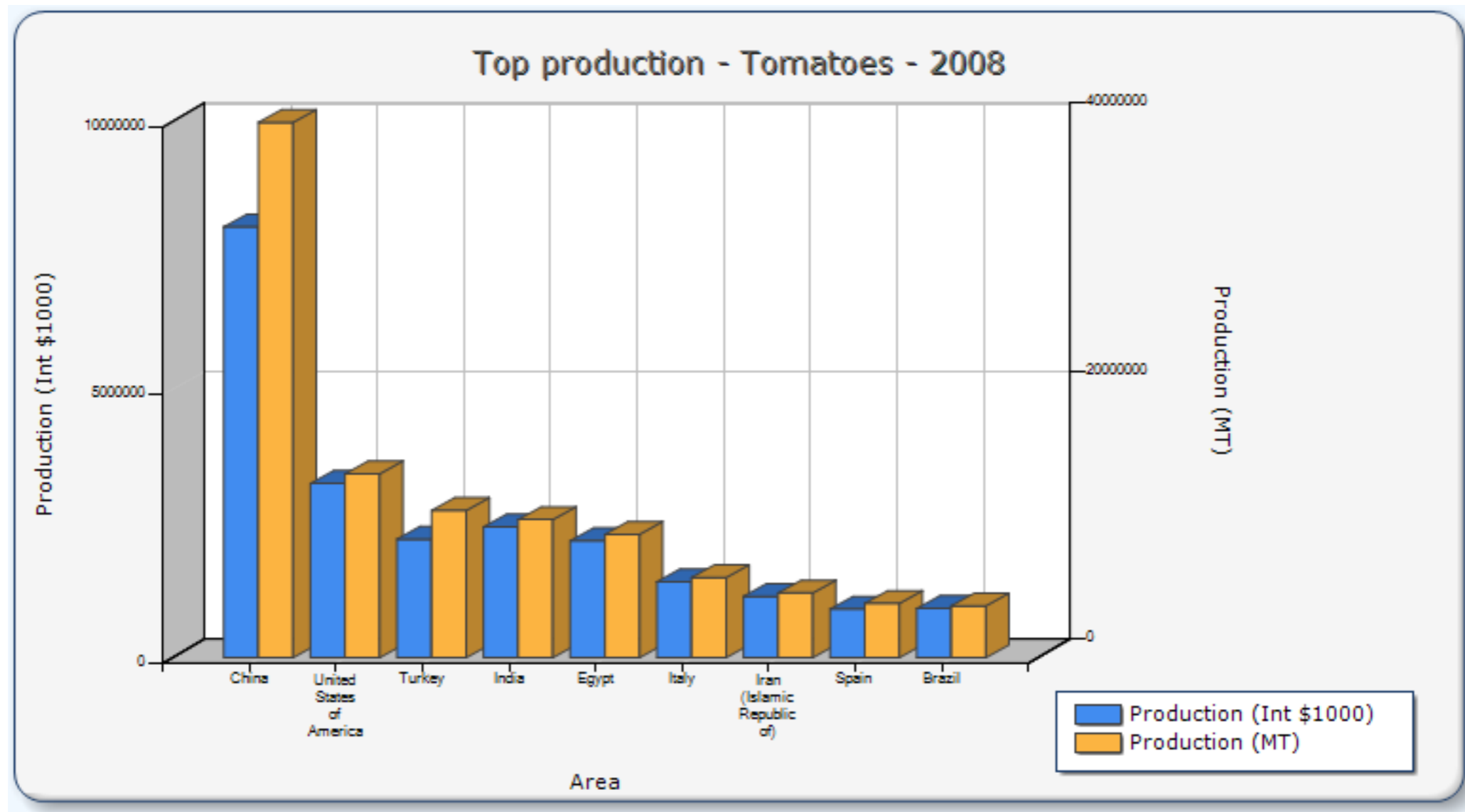


I. Background.



I) Importancia del Tomate.

España es el octavo productor mundial de tomate con 4 millones de toneladas (FAOSTAT, 2008).



I. Background.

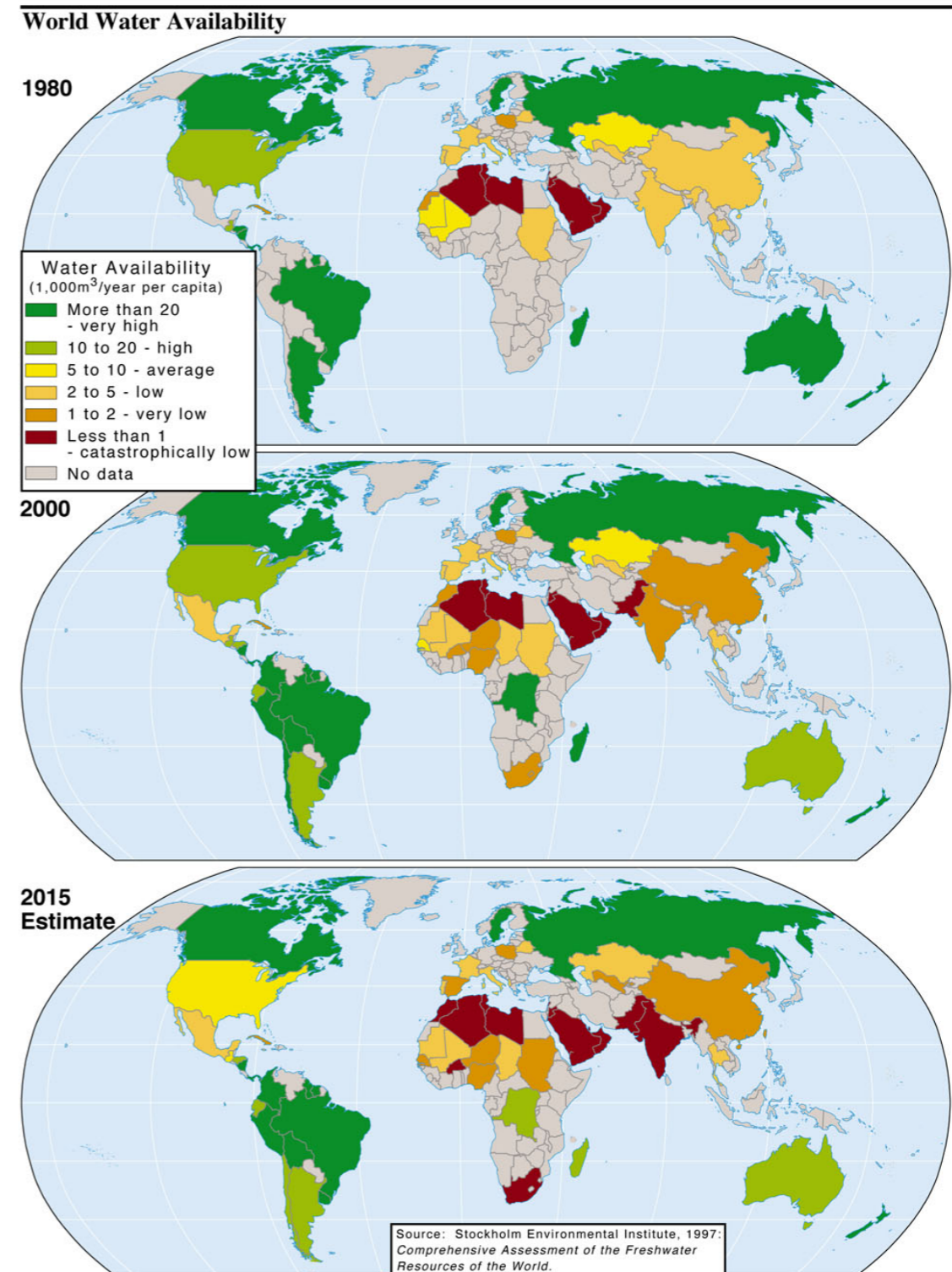


2) Problemas de la Sequía.

El aumento de la demanda y la reducción en la disponibilidad en la cantidad de agua utilizable para consumo es uno de los principales problemas al que se enfrenta la agricultura.

España esta clasificada como un país donde la disponibilidad de agua fresca es baja (2000-5000 m³/año-cápita).

Se estima que en 2015, la disponibilidad de agua será muy baja





I. Background.

3) Recursos científicos disponibles.

- ◆ Secuencias de varios genomas de tomate:
 - *Solanum lycopersicum* v.2.40.
 - *Solanum pimpinellifolium* draft.
 - *Solanum pennellii* (secuencia en progreso)

- ◆ Mapas genéticos.
 - 19 mapas en la base de datos: SGN

- ◆ Líneas de introgresión (IL)
 - IL *S. pennellii* (http://solgenomics.net/maps/pennellii_il/index.pl)

- ◆ Datos Transcriptómicos (GEO-ArrayExpress).
 - 22 experimentos, 347 ensayos.

- ◆ Literatura científica (PubMed).
 - 117 publicaciones relacionadas con tomate y sequía



Caso de Estudio: Estrés por Sequía en Tomate

1. Background.

2. Diseño Experimental.

3. Material y Herramientas.



2. Diseño experimental.



Diseño experimental, opciones:

1. Genes candidatos de respuesta a sequía.

2. Fenotipado de una población segregante y estudio de QTLs.

3. Mutagenesis al azar, búsqueda de mutantes de respuesta a sequía, y mapeado de la mutación.



2. Diseño experimental.



Diseño experimental, opciones:

1. Genes candidatos de respuesta a sequía.

2. Fenotipado de una población segregante y estudio de QTLs.

3. Mutagenesis al azar, búsqueda de mutantes de respuesta a sequía, y mapeado de la mutación.





Caso de Estudio: Estrés por Sequía en Tomate

1. Background.

2. Diseño Experimental.

3. Material y Herramientas.



3. Herramientas.



Material y Herramientas:

I. Genes candidatos de respuesta a sequía

I. Material vegetal:

- Plantas susceptibles versus resistentes.
- Plantas tratadas versus control.
- Selección del tejido (raíces versus hojas).

2. Estudios de expresión.

- Microarrays versus RNAseq.
- Análisis de expresión.
- Verificación candidatos (QRT-PCR).



3. Herramientas.



Material y Herramientas:

I. Genes candidatos de respuesta a sequía

I. Material vegetal:

- Plantas susceptibles versus resistentes.
- Plantas tratadas versus control.
- Selección del tejido (raíces versus hojas).

2. Estudios de expresión.

- Microarrays versus RNAseq.
- Análisis de expresión.
- Verificación candidatos (QRT-PCR).



Bases de datos.

Bibliografía

(Pubmed, Web Of Science).

Expresión

(GEO, ArrayExpress).



3. Herramientas.



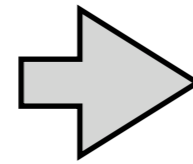
Bases de datos.

Bibliografía

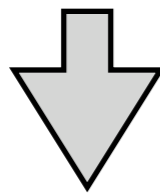
(Pubmed, Web Of Science).

Expresión

(GEO, ArrayExpress).



Sin publicación en
PubMed o Web of Science



Series GSE22304		Query DataSets for GSE22304
Status	Public on Jun 12, 2010	
Title	Expression data in response to abiotic stresses in tomato at flowering stage	
Organism	Solanum lycopersicum	
Experiment type	Expression profiling by array	
Summary	A comparative study were made to know the abiotic stress tolerance machanism between tolerant and susceptible plants at flowering stage. The tolerance in response to abiotic stresses are sum of expression of thousands of genes at a particular stage.	
Overall design	Tomato plants were exposed to drought and heat stress for RNA extraction and hybridization on Affymetrix microarrays. To study the molecular mechanism of abiotic stress tolerance to increase the tolerance in tomato plants, transcripts of tolerant and susceptible plants at flowering stage were compared in response to heat and water stress.	



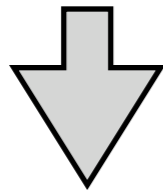
3. Herramientas.



Bases de datos.

Genes

(Sol Genomics Networks).



The screenshot shows the Sol Genomics Network website. At the top left is the logo and the text "sol genomics network". To the right are links for "home | forum | contact | help". Below this is a navigation bar with buttons for "search", "maps", "genomes", and "tools", followed by a search input field containing "sol search" and links for "log in" and "new user".

Below the navigation bar is the "Search SGN" section, which features a row of buttons for different search categories: "Genes", "QTLs & Phenotypes", "Unigenes", "Unigene Families", "Markers", "Genomic Clones", "ESTs", "Images", "People", and "Expression". The "Genes" button is highlighted.

Underneath is the "Gene search" section. It includes the text "Search for any locus or allele (gene search help page)". Below this is a search interface with a dropdown menu set to "contains" and a text input field containing "drought". To the right of the input field is a link "[Submit new locus]". Below the search input is a button labeled "Advanced search options". At the bottom center of the search section is a "Search" button.



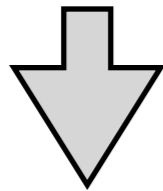
3. Herramientas.



Bases de datos.

Genes

(Sol Genomics Networks).



Gene search results

Gene search results 14 matches

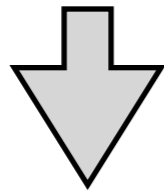
Organism	Locus name	Locus symbol	Allele name	Allele symbol	Allele phenotype
Tomato	AREB	AREB			--
Tomato	cupida	cup		1	Extremely susceptibl...
Tomato	dehydrin	dhn			--
Tomato	impatiens-2	im2		1	Small, flat bush; le...
Tomato	Solyc01g065550	Solyc01g065550			--
Tomato	Solyc01g100140	Solyc01g100140			--
Tomato	Solyc01g100160	Solyc01g100160			--
Tomato	Solyc02g077750	Solyc02g077750			--
Tomato	Solyc03g112160	Solyc03g112160			--
Tomato	Solyc05g015470	Solyc05g015470			--

1 2 >



3. Herramientas.

Lista de Genes Candidatos



Estudios de expresión.

Sobreexpresión-Silenciamiento y caracterización fenotípica.





PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Unix y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Bases de Datos Biológicas.



Source: Contributing Organizations at GMOD





Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.





Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.



I. Tipos de Bases de Datos Biológicas.

Existen 3 tipos de bases de datos biológicas (Rhee SY. *et al.* 2006):

- Repositorios públicos a gran escala.
- Bases de datos específicas de una comunidad.
- Bases de datos específicas de un proyecto.





* Repositorios públicos a gran escala.

- Generalmente mantenidos por grandes agencias gubernamentales o consorcios internacionales.
- Enorme cantidad de datos.
- Datos no curados o pobremente curados.
- Almacenaje de datos a largo plazo.
- Ejemplos: GenBank, Uniprot, Gene Expression Omnibus (GEO), ArrayExpress.





- * Bases de datos específicas de una comunidad.
 - Generalmente mantenidos por grupos de investigación cuyos proyectos implican el desarrollo de la base de datos en si mismo.
 - Cantidad de datos aceptable que corresponde con las necesidades de una comunidad.
 - Datos curados o muy curados.
 - Almacenaje de datos a largo plazo.





* Bases de datos específicas de una comunidad.

- Distintos tipos:
 - Estudio de organismos modelo:
TAIR, MaizeGDB
 - Estudio de una familia de organismos o un clado:
Sol Genomics Network (SGN), Gramine.
 - Estudio de rutas metabólicas:
MetaCyc, KEGG pathways.
 - Estudio de filogenias:
TreeBase.
 - Estudio de Ontologías:
Gene Ontology, Plant Ontology.



I. Tipos de Bases de Datos Biológicas.

- * Bases de datos específicas de un proyecto.
 - Generalmente mantenidos por un grupo o un consorcio
 - Poca cantidad de datos.
 - Gran variabilidad en el nivel de curado de los datos.
 - Tiempo de vida limitado, generalmente asociado con el periodo de financiación de un proyecto.
 - Ejemplos: Plant Genome Network (PGN)





Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.



2. Repositorios Públicos.



NCBI (National Center for Biotechnology Information)

<http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI website homepage. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To' menus, and a 'My NCBI Sign In' link. Below this is a search bar with a dropdown menu set to 'All Databases', a search input field, and 'Search' and 'Clear' buttons. On the left side, there is a vertical navigation menu with links to various sections: NCBI Home, Site Map (A-Z), All Resources, Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation. The main content area features a 'Welcome to NCBI' message, a 'Notice' box with a yellow background, and a 'Popular Resources' list. The 'Notice' box states that NCBI servers will undergo maintenance on Saturday, June 11, at 7 AM EDT, and will return to full operation by 4 PM EDT. It also mentions that during this time, NCBI services like Pubmed, BLAST, and the Entrez search and retrieval systems will be operational but may have slower performance. The 'Popular Resources' list includes BLAST, Bookshelf, Gene, Genome, Nucleotide, OMIM, Protein, PubChem, PubMed, PubMed Central, and SNP. The 'NCBI News' section features a headline about preliminary genomic assemblies from two isolates from the European E. coli outbreak, dated 07 Jun 2011.

NCBI Resources How To My NCBI Sign In

NCBI National Center for Biotechnology Information

Search All Databases Search Clear

NCBI Home
Site Map (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [Research](#) | [RSS Feeds](#)

Notice

NCBI servers will undergo maintenance beginning **Saturday, June 11 at 7 AM EDT** and will return to full operation by 4PM EDT.

During this time, NCBI services, including Pubmed, BLAST and the Entrez search and retrieval systems, will be operational but may have slower performance.

Data submission systems, including GenBank FTP submissions, SequinMacroSend, dbGaP, SRA, Trace and GEO submission systems, will be unavailable. dbGaP and SRA retrieval will not be available during this time as well. Please contact NCBI with concerns: info@ncbi.nlm.nih.gov

Popular Resources

- BLAST
- Bookshelf
- Gene
- Genome
- Nucleotide
- OMIM
- Protein
- PubChem
- PubMed
- PubMed Central
- SNP

NCBI News

[Preliminary genomic assemblies from two isolates from the European E. coli outbreak now available](#)

07 Jun 2011

Preliminary genomic assemblies of two isolates are in the

2. Repositorios Públicos.

NCBI (National Center for Biotechnology Information)

<http://www.ncbi.nlm.nih.gov/>

Partes a destacar:

- GenBank.
- PubMed.
- Gene Expression Omnibus (GEO)
- Sequence Read Archive (SRA)
- Taxonomy



2. Repositorios Públicos: NCBI



GenBank, base de datos del NIH para secuencias genéticas, una colección anotada de todas las secuencias de ADN públicamente disponibles (Benson DA. *et al.* 2011).

<http://www.ncbi.nlm.nih.gov/genbank/>

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>

NCBI Resources How To My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear

Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Using Nucleotide

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

[RefSeq FTP](#)

Nucleotide Tools

[Submit to GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[BLAST](#)

[Batch Entrez](#)

Other Resources

[GenBank Home](#)

[RefSeq Home](#)

[Gene Home](#)

[SRA Home](#)

[INSDC](#)

2. Repositorios Públicos: NCBI



GenBank:

1 → Sección de Búsqueda

Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Found 770004 nucleotide sequences. Nucleotide (15269) EST (754652) GSS (83)

Results: 1 to 20 of 15269

<< First < Prev Page 1 of 764 Next > Last >>

- [Arabidopsis thaliana chromosome 1, complete sequence](#)
1. 30,427,671 bp linear DNA
Accession: CP002684.1 GI: 332189094
[GenBank](#) [FASTA](#) [Graphics](#)
- [Gossypium hirsutum mitogen-activated protein kinase \(MAPK\) gene, promoter region](#)
2. 1,622 bp linear DNA
Accession: HM150999.1 GI: 315258198
[GenBank](#) [FASTA](#) [Graphics](#)
- [Gossypium hirsutum mitogen-activated protein kinase 16 \(MPK16\) gene, complete cds](#)
3. 6,195 bp linear DNA
Accession: FJ966896.1 GI: 297748124
[GenBank](#) [FASTA](#) [Graphics](#)

Filter your results:

- All (15269)
- [Bacteria \(198\)](#)
- [INSDC \(GenBank\) \(14860\)](#)
- [mRNA \(9150\)](#)
- [RefSeq \(401\)](#)

[Manage Filters](#)

3

Top Organisms [Tree]

- Populus tremula x Populus alba (7835)
- Oryza sativa (2155)
- Oryza sativa Indica Group (1366)
- Pinus taeda (577)
- Oryza sativa Japonica Group (524)
- All other taxa (4292)
- More...

4

2

Resultados similares en otras subsecciones

Filtro para tipo de nucleótido

Filtro taxonómico



2. Repositorios Públicos: NCBI



GenBank:

1 → Aplicación de los filtros

Nucleotide
Alphabet of Life

Search: Save search Limits Advanced search Help

Search Clear

Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Found 22653 nucleotide sequences. Nucleotide (559) EST (22094)

Results: 1 to 20 of 49

<< First < Prev Page 1 of 3 Next > Last >>

- [Capsicum annuum chitinase class II \(CACHi2\) mRNA, complete cds](#)
 1. 1,004 bp linear mRNA
Accession: AF091235.1 GI: 3641354
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)
 2. 937 bp linear mRNA
Accession: AF291179.1 GI: 9885805
[GenBank](#) [FASTA](#) [Graphics](#)
 3. 672 bp linear mRNA
Accession: HM068895.1 GI: 298155395
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)
 4. 958 bp linear mRNA

Filter your results:

- [All \(559\)](#)
- Bacteria (0)
- [INSDC \(GenBank\) \(559\)](#)
- mRNA (49)**
- RefSeq (0)

[Manage Filters](#)

▼ Taxonomic Groups [\[List\]](#)

- Solanaceae (49)
 - Nicotiana (27)
 - Capsicum (12)
 - Solanum (10)

Find related data

Database:

2. Repositorios Públicos: NCBI



GenBank:

Enlace a herramientas

Nucleotide
Alphabet of Life

Search: [Limits](#) [Advanced search](#) [Help](#)

[Display Settings:](#) GenBank [Send:](#)

Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence

GenBank: HM068893.1
[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS	HM068893	958 bp	mRNA	linear	PLN 28-DEC-2010
DEFINITION	Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence.				
ACCESSION	HM068893				
VERSION	HM068893.1 GI:298155393				
KEYWORDS	.				
SOURCE	Nicotiana attenuata				
ORGANISM	Nicotiana attenuata Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Nicotianoideae; Nicotianeae; Nicotiana.				
REFERENCE	1 (bases 1 to 958)				
AUTHORS	Re,D.A., Dezar,C.A., Chan,R.L., Baldwin,I.T. and Bonaventure,G.				
TITLE	Nicotiana attenuata NaHD20 plays a role in leaf ABA accumulation during water stress, benzylacetone emission from flowers, and the timing of bolting and flower transitions				
JOURNAL	J. Exp. Bot. 62 (1), 155-166 (2011)				
PUBMED	20713465				
REFERENCE	2 (bases 1 to 958)				
AUTHORS	Bonaventure,G., Re,D. and Baldwin,I.				
TITLE	Analysis of drought and ABA responsive genes in Nicotiana attenuata				
JOURNAL	Unpublished				

Analyze this sequence

- Run BLAST
- Pick Primers
- Find in this Sequence

LinkOut to external resources

Gramene [Gramene]

All links from this record

- Full text in PMC
- PubMed

Recent activity [Turn Off](#) [Clear](#)

- Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence Nucleotide



2. Repositorios Públicos: NCBI



GenBank:

```
FEATURES             Location/Qualifiers
     source            1..958
                        /organism="Nicotiana attenuata"
                        /mol_type="mRNA"
                        /db_xref="taxon:49451"
     gene                1..958
                        /gene="OSM1"
                        /note="NaOSM1"
     misc\_feature       4..738
                        /gene="OSM1"
                        /note="similar to osmotin 1"

ORIGIN
1  aaaatgtcca acaacatggg caacttgaga tcttttgttt tcttccttct tgcctacgta
61  acttatactt atgctgccac tatcgaggtc cgaacaact  gtcctacac  cgtctgggcg
121 gcgtcgacac ccataggcgg tggtcggcgt ctcgatcgag gccaaacttg ggtgatcaac
181 gcgccacgag gtactaagat ggcacgtgta tggggtcgta ctaattgtaa cttcaatgct
241 gcaggtaaag gtacatgcca aaccggtgac tgtggcggag tcctacaatg caccgggtgg
301 ggtaaaccac caaacacctt ggccgaatac gccttggatc aattcagtgg cttagatttc
361 tgggacattt ctttagtgga tggattcaat ataccgatga ctttcgcccc gactaacctt
421 agtggaggga aatgccatgc aattcattgt acggctaata taaacgggtg atgtcccgcg
481 gaacttaggg ttcccggagg atgtaataac cttgtacga  cattcggagg acaacaatat
541 tgttgcacac aaggaccatg tggccaaca  aatttatcaa gatttttcaa acaaagatgc
601 cccgatgcct atagctacc  acaagatgat cctacaagca cttttacttg ccctggtggt
661 agtacaatt  atagggttat cttttgtcct aatgggtcaag ctcacccaaa ttttcccctt
721 ggagatgcct ggaagtgatg aagtggctaa gtagtgtggc tatttctgta ataagatcac
781 cttttggtcg aattattctg tcgaagcgtt agtgtaagac aatctacttt gactcgtttt
841 tatagtttac acactttgtt tgaagtgatc aattcacgat ctttgctgta ataacctaa
901 gacctgaata agagtcacat gtacttttgt cttgatgta  tatagatcaa taatgcat
```

//



2. Repositorios Públicos: NCBI



GenBank:

Formato

Guardar en archivos

Nucleotide
Alphabet of Life

Search: Nucleotide

[Limits](#) [Advanced search](#) [Help](#)

[Display Settings:](#) GenBank

Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence

GenBank: HM068893.1
[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS	HM068893	958 bp	mRNA	linear	PLN 28-DEC-2010
DEFINITION	Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence.				
ACCESSION	HM068893				
VERSION	HM068893.1 GI:298155393				
KEYWORDS	.				
SOURCE	Nicotiana attenuata				
ORGANISM	Nicotiana attenuata Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Nicotianoideae; Nicotianeae; Nicotiana.				
REFERENCE	1 (bases 1 to 958)				
AUTHORS	Re,D.A., Dezar,C.A., Chan,R.L., Baldwin,I.T. and Bonaventure,G.				
TITLE	Nicotiana attenuata NaHD20 plays a role in leaf ABA accumulation during water stress, benzylacetone emission from flowers, and the timing of bolting and flower transitions				
JOURNAL	J. Exp. Bot. 62 (1), 155-166 (2011)				
PUBMED	20713465				
REFERENCE	2 (bases 1 to 958)				
AUTHORS	Bonaventure,G., Re,D. and Baldwin,I.				
TITLE	Analysis of drought and ABA responsive genes in Nicotiana attenuata				
JOURNAL	Unpublished				

Change region shown

Customize view

Analyze this sequence

Run BLAST
Pick Primers
Find in this Sequence

LinkOut to external resources

Gramene [Gramene]

All links from this record

Full text in PMC
PubMed

Recent activity

[Turn Off](#) [Clear](#)

Nicotiana attenuata osmotin 1-like (OSM1) mRNA, complete sequence Nucleotide



2. Repositorios Públicos: NCBI



GenBank:

Guardar en archivos

Display Settings: Summary, 20 per page, Sorted by Default order

Found 21339 nucleotide sequences. Nucleotide (515) EST (20824)

Results: 10

- [Solanum lycopersicum mRNA for SIGRX1 protein, cultivar Hongbaoshi](#)
1. 1,003 bp linear mRNA
Accession: FN646220.1 GI: 308233000
[GenBank](#) [FASTA](#) [Graphics](#)
- [Lycopersicon esculentum ethylene-responsive late embryogenesis-like protein \(ER5\) mRNA, complete cds](#)
2. 748 bp linear mRNA
Accession: U77719.1 GI: 1684829
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)
- [Lycopersicon chilense proline-rich protein \(PRP13\) gene, complete cds](#)
3. 552 bp linear mRNA
Accession: U19098.1 GI: 1001952
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)
- [Lycopersicon esculentum non specific lipid transfer protein \(le16\) mRNA, complete cds](#)
4. 583 bp linear mRNA
Accession: U81996.1 GI: 1816534
[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

Send to: Filter your results:

Choose Destination

File Clipboard
 Collections Analysis Tool

Download 10 items.

Format

- Summary
- GenBank
- GenBank (full)
- FASTA
- ASN.1
- XML
- INSDSeq XML
- TinySeq XML
- Feature Table
- Accession List
- GI List

Analyze these sequences

Run BLAST

Find related data

Database: Select

Find items

2. Repositorios Públicos: NCBI



PubMed, base de datos del NIH para publicaciones y literatura científica.

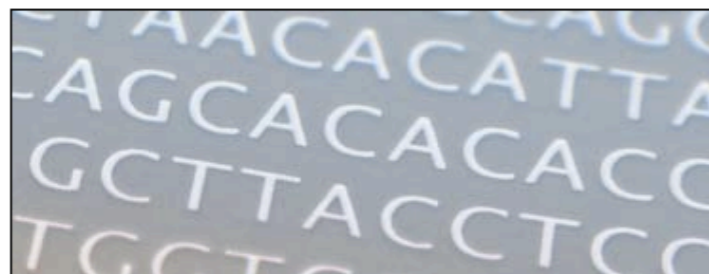
<http://www.ncbi.nlm.nih.gov/pubmed/>

NCBI Resources How To My NCBI Sign In

Nucleotide Alphabet of Life

Search: Nucleotide Limits Advanced search Help

Search Clear



Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Using Nucleotide

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

[RefSeq FTP](#)

Nucleotide Tools

[Submit to GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[BLAST](#)

[Batch Entrez](#)

Other Resources

[GenBank Home](#)

[RefSeq Home](#)

[Gene Home](#)


[SRA Home](#)

[INSDC](#)

2. Repositorios Públicos: NCBI



PubMed,

 U.S. National Library of Medicine
National Institutes of Health

Search: PubMed

[RSS](#) [Save search](#) [Limits](#) [Advanced search](#) [Help](#)

drought tomato

[Display Settings:](#) Summary, 20 per page, Sorted by Recently Added

[Send to:](#)

Results: 1 to 20 of 117

<< First < Prev Page 1 of 6 Next > Last >>

- [An insertional mutagenesis programme with an enhancer trap for the identification and tagging of genes involved in abiotic stress tolerance in the tomato wild-related species *Solanum pennellii*.](#)
1. Atarés A, Moyano E, Morales B, Schleicher P, García-Abellán JO, Antón T, García-Sogo B, Perez-Martin F, Lozano R, Flores FB, Moreno V, Del Carmen Bolarin M, Pineda B.
Plant Cell Rep. 2011 Jun 7. [Epub ahead of print]
PMID: 21647638 [PubMed - as supplied by publisher]
[Related citations](#)
- [Identification and expression pattern of one stress-responsive NAC gene from *Solanum lycopersicum*.](#)
2. Han Q, Zhang J, Li H, Luo Z, Ziaf K, Ouyang B, Wang T, Ye Z.
Mol Biol Rep. 2011 Jun 3. [Epub ahead of print]
PMID: 21637957 [PubMed - as supplied by publisher]
[Related citations](#)
- [Atypical epigenetic mark in an atypical location: cytosine methylation at asymmetric \(CNN\) sites within the body of a non-repetitive tomato gene.](#)
3. Gonzalez RM, Ricardi MM, Iusem ND.
BMC Plant Biol. 2011 May 20;11(1):94. [Epub ahead of print]
PMID: 21599976 [PubMed - as supplied by publisher] **Free Article**
[Related citations](#)

PubMed:

- Relativamente actualizado (tiempo desde aparición de la publicación hasta que aparece en PubMed de no mas de un par de días).
- Algunas de las revistas científicas asociadas a agricultura y mejora de plantas no aparecen (Theoretical Applied and Genetics).
(Más información: <http://wwwcf.nlm.nih.gov/serials/journals/index.cfm>)
- No existen enlaces entre los artículos y secuencias, genes, marcadores u otros datos contenidos en el artículo y en la base de datos.



2. Repositorios Públicos: NCBI



Gene Expression Omnibus (GEO), base de datos del NIH para datos de expresion (microarrays, NGS...)

<http://www.ncbi.nlm.nih.gov/geo/>

The screenshot shows the NCBI GEO website interface. At the top left is the NCBI logo, and at the top center is the GEO logo with the text "Gene Expression Omnibus". Navigation links include "GEO Publications", "FAQ", "MIAME", "Email GEO", and "Login". A breadcrumb trail shows "NCBI » GEO".

The main content area features a description: "Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)"

Below the description are two main sections:

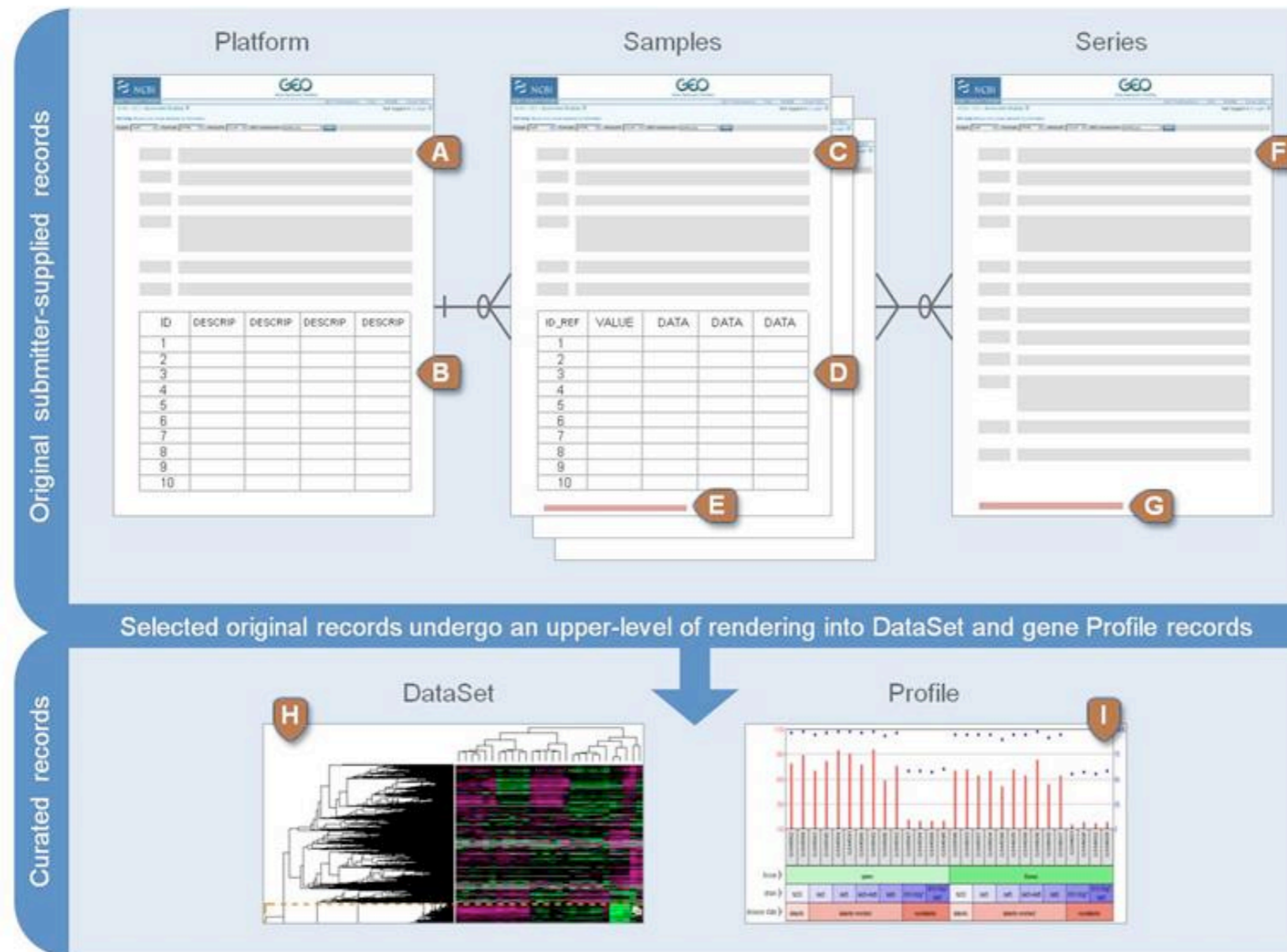
- GEO navigation:** Contains a "QUERY" section with three search options: "DataSets", "Gene profiles", and "GEO accession", each with a text input field and a "GO" button. Below this is a "GEO BLAST" button. The "BROWSE" section is a tree structure: "DataSets" and "GEO accessions" are the primary categories. "GEO accessions" further branches into "Platforms", "Samples", and "Series".
- Site contents:** A sidebar menu with two main categories:
 - Public data:** A table showing statistics:

Platforms	8,880
Samples	577,869
Series	23,357
DataSets	2,720
 - Documentation:** A list of links: Overview | FAQ | Find, Submission guide, Linking & citing, Journal citations, Construct a Query, Programmatic access, DataSet clusters, GEO announce list, and Data disclaimer.

2. Repositorios Públicos: NCBI



Gene Expression Omnibus (GEO), Organización de datos.



Source: <http://www.ncbi.nlm.nih.gov/geo/info/overview.html>

2. Repositorios Públicos: NCBI



Gene Expression Omnibus,

NCBI » GEO » Repository browser » Samples

GEO Publications | FAQ | MIAME | Email GEO | Login

Series | **Samples** | Platforms | DataSets | Advanced search

tomato drought Search 12 samples Export

Page 1 of 1 Page size 200

Accession	Title	Sample type	Organism(s)	Ch	Platform	Series	Supplementary	Contact	Release date
GSM555140	Tomato Drought Susceptible Control rep-1	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555141	Tomato Drought Susceptible Control rep-2	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555142	Tomato Drought Susceptible Control rep-3	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555143	Tomato Drought Susceptible Treated rep-1	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555144	Tomato Drought Susceptible Treated rep-2	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555145	Tomato Drought Susceptible Treated rep-3	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555146	Tomato Drought Tolerant Control rep-1	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555147	Tomato Drought Tolerant Control rep-2	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555148	Tomato Drought Tolerant Control rep-3	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555149	Tomato Drought Tolerant Treated rep-1	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555150	Tomato Drought Tolerant Treated rep-2	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010
GSM555151	Tomato Drought Tolerant Treated rep-3	RNA	<i>Solanum lycopersicum</i>	1	GPL4741	GSE22304	CEL CHP	Ashutosh Rai	Jun 12, 2010

2. Repositorios Públicos: NCBI



Gene Expression Omnibus, registro

The screenshot shows the NCBI GEO website interface. At the top left is the NCBI logo, and at the top right is the GEO logo (Gene Expression Omnibus). Below the logos are navigation links: HOME, SEARCH, SITE MAP, and GEO Publications. The main content area displays the accession number GSM555140. A search bar at the top of the content area shows the following settings: Scope: Self, Format: HTML, Amount: Quick, GEO accession: GSM555140, and a GO button. Below the search bar, the sample information is displayed in a table format.

Sample GSM555140		Query DataSets for GSM555140
Status	Public on Jun 12, 2010	
Title	Tomato Drought Susceptible Control rep-1	
Sample type	RNA	
Source name	Leave at flowering stage in normal growth condition	
Organism	Solanum lycopersicum	
Characteristics	genotype: Drought Susceptible Genotype	
Treatment protocol	Sample 1,2,3,7,8,9,13,14,15,19,20,21 were grown in normal growth conditions without any stress in green house. While sample 4,5,6, and 10,11,12 were subjected to water stress by with holding water for 7 days. Sample 16,17,18 and 22,23,24 were subjected to 40C for 60 min in growth chamber before sample collection.	
Growth protocol	Plants were sown in plastic pots at optimal conditions for tomato plants.	
Extracted molecule	total RNA	
Extraction protocol	Trizol extraction of total RNA was performed according to the manufacturer's instructions.	
Label	biotin	
Label protocol	Biotinylated cRNA were prepared according to the standard Affymetrix protocol from 6 ug total RNA (Expression Analysis Technical Manual, 2001, Affymetrix).	
Hybridization protocol	Following fragmentation, 10 ug of cRNA were hybridized for 16 hr at 45C on GeneChip Tomato Genome Array. GeneChips were washed and stained in the Affymetrix Fluidics Station 400.	

2. Repositorios Públicos: NCBI



Gene Expression Omnibus, bajada de datos.

Total number of rows: **10209**

Table truncated, full table size **267 Kbytes**.

[View full table...](#)

Supplementary file	Size	Download	File type/resource
GSM555140_DSC_1.CEL.gz	920.1 Kb	(ftp) (http)	CEL
GSM555140_DSC_1.rma.chp.gz	87.7 Kb	(ftp) (http)	CHP

Raw data provided as supplementary file

Processed data included within Sample table

Processed data provided as supplementary file

Processed data is available on Series record

2. Repositorios Públicos: NCBI



Sequence Read Archive (SRA), base de datos para guardar las secuencias producidas por los métodos de nueva secuenciación (NGS) tales como Illumina, 454, Solid, Helicos...

<http://www.ncbi.nlm.nih.gov/sra>

The screenshot shows the NCBI Sequence Read Archive (SRA) website interface. At the top, there is a navigation bar with links for 'All Databases', 'PubMed', 'Nucleotide', 'Protein', 'Genome', 'Structure', 'OMIM', 'PMC', 'Journals', and 'Books'. Below this is a search bar with 'SRA' entered and buttons for 'Go' and 'Clear'. A sidebar on the left contains links for 'About Entrez', 'Related Resources', 'Trace Archive', 'SRA', and 'Trace Assembly'. The main content area features a green-bordered box with the following text:

Sequence Read Archive (SRA) is still in service.

Recently, NCBI announced that due to budget constraints, it would be discontinuing its Sequence Read Archive (SRA) and Trace Archive repositories for high-throughput sequence data. However, NIH has since committed interim funding for SRA in its current form until October 1, 2011. In addition, NCBI has been working with staff from other NIH Institutes and NIH grantees to develop an approach to continue archiving a widely used subset of next generation sequencing data after October 1, 2011.

We now plan to continue handling sequencing data associated with:

1. RNA-Seq, ChIP-Seq, and epigenomic data that are submitted to GEO
2. Genomic and Transcriptomic assemblies that are submitted to GenBank
3. Genomic assemblies to GenBank/WGS
4. 16S ribosomal RNA data associated with metagenomics that are submitted to GenBank

In addition, NCBI will continue to provide access to existing SRA and Trace Archive data for the foreseeable future. NCBI is also continuing to discuss with NIH Institutes approaches for handling other next-generation sequencing data associated with specific large-scale studies.

2. Repositorios Públicos: NCBI



Sequence Read Archive (SRA),

Sequence Read Archive

[Main](#) [Browse](#) [Search](#) [Download](#) [Submit](#) [Documentation](#) [Software](#) [Trace Archive](#) [Trace Assembly](#) [Trace Home](#) [Trace BLAST](#)

SRA Objects [by Sequence/Name](#) [BLAST](#) [Entrez](#)

Announcements

2011-05-09: SRA Archive is still in service. [less...](#)

Recently, NCBI announced that due to budget constraints, it would be discontinuing its Sequence Read Archive (SRA) and Trace Archive repositories for high-throughput sequence data. However, NIH has since committed interim funding for SRA in its current form until October 1, 2011. In addition, NCBI has been working with staff from other NIH Institutes and NIH grantees to develop an approach to continue archiving a widely used subset of next generation sequencing data after October 1, 2011.

We now plan to continue handling sequencing data associated with:

1. RNA-Seq, ChIP-Seq, and epigenomic data that are submitted to GEO
2. Genomic and Transcriptomic assemblies that are submitted to GenBank
3. 16S ribosomal RNA data associated with metagenomics that are submitted to GenBank

In addition, NCBI will continue to provide access to existing SRA and Trace Archive data for the foreseeable future. NCBI is also continuing to discuss with NIH Institutes approaches for handling other next-generation sequencing data associated with specific large-scale studies.

Search for SRA related objects

Any Entrez request. Examples: "[human](#)", "[SRX006930](#)"



2. Repositorios Públicos: NCBI



Sequence Read Archive (SRA),

NCBI Site map | All databases | PubMed | Search

Sequence Read Archive

Main | Browse | **Search** | Download | Submit | Documentation | Software | Trace Archive

SRA Objects | by Sequence/Name | BLAST | Entrez

i Announcements

2011-05-09: SRA Archive is still in service. [more...](#)

Search for SRA related objects

Any Entrez request. Examples: "[human](#)", "[SRX006930](#)"

	Public access	Controlled access	All
SRA Experiments	57		57
SRA Studies	11		11
BioSamples	49		49
GEO Datasets	4		4

2. Repositorios Públicos: NCBI



Sequence Read Archive (SRA),

The screenshot shows the NCBI Sequence Read Archive interface. At the top, there are navigation tabs for various databases: All Databases, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, Journals, and Books. The search bar contains the text "SRA" and "(tomato) NOT cluster_dbgap[PROP]". Below the search bar, there are buttons for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The display settings are set to "Summary" and "Show 20". The search results show 57 items in total, with 0 controlled access and 57 public access. The first result is SRX036645, titled "tomato leaves after 144 hours of P. infestans infection". The submitter is Boyce Thompson Institute, Fei (SRA027389, released on Dec 26, 2010). The study is "Transcriptome profiling of tomato and Phytophthora infestans during tomato infection by P. infestans (SRP004932)". The sample is SRS150759 and the instrument is 454 GS FLX. A table shows the total runs, spots, and bases for the experiment, and a detailed table lists the individual runs.

NCBI Sequence Read Archive

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search SRA for (tomato) NOT cluster_dbgap[PROP] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 57 Controlled Access: 0 Public Access: 57

Items 1 - 20 of 57 Page 1 of 3 Next

1: [SRX036645](#) tomato leaves after 144 hours of P. infestans infection Links

Submitter: Boyce Thompson Institute, Fei (SRA027389, released on Dec 26, 2010) Download reads for this experiment in [sra](#) or [sra-lite](#) (70.0M) formats

Study: Transcriptome profiling of tomato and Phytophthora infestans during tomato infection by P. infestans (SRP004932) • [Summary](#) • [All experiments](#)

Sample: ([SRS150759](#))

Instrument: 454 GS FLX

Total: 2 runs, 206,671 spots, 51.3M bases			
#	Run	# of Spots	# of Bases
1.	SRR088779	187,459	44.4M
2.	SRR088823	19,212	6.9M

2. Repositorios Públicos: NCBI



Taxonomy, base de datos que contiene nombres y clasificaciones para todos los organismos contenidos en el NCBI

<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>

The screenshot shows the NCBI Taxonomy Browser homepage. At the top, there is a navigation bar with links to PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. Below this is a search bar with a dropdown menu set to 'AS' and a 'complete name' option. The main content area is titled 'The NCBI Taxonomy Homepage' and features a list of direct links to commonly used organisms in molecular research projects. A sidebar on the left contains links to various taxonomy-related resources.

NCBI Taxonomy Browser

Search for AS complete name lock

The NCBI Taxonomy Homepage

These are direct links to some of the organisms commonly used in molecular research projects:

Arabidopsis thaliana	Escherichia coli	Pneumocystis carinii
Bos taurus	Hepatitis C virus	Rattus norvegicus
Caenorhabditis elegans	Homo sapiens	Saccharomyces cerevisiae
Chlamydomonas reinhardtii	Mus musculus	Schizosaccharomyces pombe
Danio rerio (zebrafish)	Mycoplasma pneumoniae	Takifugu rubripes
Dictyostelium discoideum	Oryza sativa	Xenopus laevis
Drosophila melanogaster	Plasmodium falciparum	Zea mays

Taxonomy browser
Archaea
Bacteria
Eukaryota
Viroids
Viruses

Taxonomy common tree

Taxonomy information

Taxonomy resources

Taxonomic advisors

2. Repositorios Públicos: NCBI



Taxonomy, búsqueda de datos basados en grupos taxonómicos.

NCBI Taxonomy Browser interface. Search for 'tomato' as 'complete name'. The 'lock' checkbox is checked. The 'Display' is set to 3 levels using filter: none. The 'BLAST' checkbox is checked. The 'Lineage' is shown as: [root](#); [cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Streptophytina](#); [Embryophyta](#); [Tracheophyta](#); [Euphyllophyta](#); [Spermatophyta](#); [Magnoliophyta](#); [eudicotyledons](#); [core eudicotyledons](#); [asterids](#); [lamiids](#); [Solanales](#); [Solanaceae](#); [Solanoideae](#); [Solaneae](#); [Solanum](#); [Lycopersicon](#).

- [Solanum lycopersicum](#) (tomato) [MapView](#) [LinkOut](#) [BLAST page](#) *Click on organism name to get more information.*
 - [Solanum lycopersicum var. cerasiforme](#) (cherry tomato) [LinkOut](#)



NCBI Taxonomy Browser interface. Search for 'tomato' as 'complete name'. The 'lock' checkbox is checked. The 'Display' is set to 3 levels using filter: none. The 'BLAST' checkbox is checked. The 'Lineage' is shown as: [root](#); [cellular organisms](#); [Eukaryota](#); [Viridiplantae](#); [Streptophyta](#); [Streptophytina](#); [Embryophyta](#); [Tracheophyta](#); [Euphyllophyta](#); [Spermatophyta](#); [Magnoliophyta](#); [eudicotyledons](#); [core eudicotyledons](#); [asterids](#); [lamiids](#); [Solanales](#); [Solanaceae](#); [Solanoideae](#); [Solaneae](#); [Solanum](#); [Lycopersicon](#).

- [Solanum lycopersicum](#) (tomato) [1 298,295](#) [1 43,810](#) [44](#) [BLAST page](#) *Click on organism name to get more information.*
 - [Solanum lycopersicum var. cerasiforme](#) (cherry tomato) [287](#) [1,160](#)






2. Repositorios Públicos.



EBI (European Bioinformatics Institute)

<http://www.ebi.ac.uk/>

EMBL-EBI  European Bioinformatics Institute

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index  

Explore the EBI:

FIND

Examples: [ROA1_HUMAN](#), [tpi1](#), [Sulston...](#) [Help](#) | [Feedback](#)

Data Resources and Tools

- [ENA](#)
- [UniProt](#)
- [ArrayExpress](#)
- [Ensembl](#)
- [InterPro](#)
- [PDBe](#)
- [Genomes](#)
- [Nucleotide Sequences](#)
- [Protein Sequences](#)
- [Macromolecular Structures](#)
- [Small Molecules](#)
- [Gene Expression](#)
- [Molecular Interactions](#)
- [Reactions & Pathways](#)
- [Protein Families](#)
- [Enzymes](#)
- [Literature](#)
- [Taxonomy](#)
- [Ontologies](#)
- [Patent Resources](#)
- [Sequence Similarity & Analysis](#)
- [Pattern & Motif Searches](#)
- [Structure Analysis](#)
- [Text Mining](#)
- [Downloads](#)
- [Web Services](#)



2. Repositorios Públicos.



EBI (European Bioinformatics Institute)

<http://www.ebi.ac.uk/>

Partes a destacar:

- ENA (European Nucleotide Archive).
- UniProt
- ArrayExpress
- Ensembl
- InterPro



2. Repositorios Públicos: EBI



ENA (European Nucleotide Archive), base de datos que guarda el conjunto de todas las secuencias públicamente disponibles.

<http://www.ebi.ac.uk/ena/>

Comprende los datos de varias bases de datos como: EMBL-Bank, SRA y Trace Archive

The screenshot displays the EMBL-ENA website interface. At the top, there is a navigation bar with the EMBL-EBI logo and a search bar. Below this, a menu lists various sections: Databases, Tools, Research, Training, Industry, About Us, and Help. The main content area is titled "European Nucleotide Archive" and includes a description of the archive's purpose, a search bar for text queries, and a sequence search tool. A sidebar on the left contains a navigation menu for ENA, news and announcements, and a survey link.

EMBL-EBI

Databases Tools Research Training Industry About Us Help

ENA

- ENA Home
- Search & Browse
- Submit & Update
- About ENA
- Contact

NEWS AND ANNOUNCEMENTS

ArchiveBAM 1.0 specification 16 Mar 2011

The ArchiveBAM 1.0 specification has been published. SRA submitters are advised to submit their data using the BAM format.

ENA User Survey 2011 now available 11 Mar 2011

Have your say and help us to improve ENA in our brief survey at http://www.surveymonkey.com/s/ENA_User_Survey

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation ... [more](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text search

Enter search query, for example: BN000065 Search

Sequence Search

Enter or paste a nucleotide sequence

Advanced Search

Search

2. Repositorios Públicos: EBI



UniProt, base de datos de secuencias de proteínas e información funcional de las mismas (<http://www.uniprot.org/>).

Esta compuesta por 4 secciones:

- *UniProt Knowledgebase* (UniProtKB), punto central de acceso para la información de proteínas, incluida función, clasificación y referencias cruzadas.
- *Uniprot Archive* (UniParc), set de secuencias de proteínas.
- *UniProt Reference Clusters* (UniRef), set completo y redundante de datos compuesto por UniProtKB y UniParc.
- *Uniprot Metagenomic and Environmental Sequences* (UniMES), repositorio de datos específico de datos metagenómicos y ambientales.

2. Repositorios Públicos: EBI



- UniProt Knowledgebase (UniProtKB)*, se divide en dos secciones:
- UniProtKB/Swiss-Prot, manualmente anotadas y curadas.
 - UniProtKB/TrEMBL, automáticamente anotadas

WELCOME

The mission of [UniProt](#) is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: <ul style="list-style-type: none">★ Swiss-Prot, which is manually annotated and reviewed.★ TrEMBL, which is automatically annotated and is not reviewed. Includes Complete Proteome Sets .
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords and more .

NEWS

[UniProt release 2011_06 - May 31, 2011](#)

New biocuration pages on UniProt website

- › [Statistics for UniProtKB: Swiss-Prot · TrEMBL](#)
- › [Forthcoming changes](#)
- › [News archives](#)

SITE TOUR



2. Repositorios Públicos: EBI



ArrayExpress, base de datos que contiene experimentos de genómica funcional, concretamente datos de expresión. **Gene Expression Atlas**, es una subsección de ArrayExpress que contiene datos de expresión curados con acceso por gen, experimento o condición (<http://www.ebi.ac.uk/arrayexpress/>).

The screenshot shows the ArrayExpress website. At the top left is the ArrayExpress logo, a circular pattern of green and red dots. To its right is the text "ARRAYEXPRESS". In the top right corner is the Gene Expression Atlas logo, a stylized "AE" in red and green. Below the logo is a paragraph of text: "The **ArrayExpress Archive** is a database of functional genomics experiments including gene expression where you can query and download data collected to **MIAME** and **MINSEQE** standards. **Gene Expression Atlas** contains a subset of curated and re-annotated Archive data which can be queried for individual gene expression under different biological conditions across experiments."

The main content area is divided into two panels. The left panel is titled "Experiments Archive" and features a sub-header "22582 experiments, 638410 assays" and a sub-header "Experiment, citation, sample and factor annotations" above a search input field. It includes a "Query" button, a link to "Browse experiments | platform designs^{NEW}", and a "Submitter/reviewer login" link. The right panel is titled "Gene Expression Atlas" and features a sub-header "Information is unavailable at the moment". It has a search interface with "Genes" and "Conditions" input fields, a "up/down in" dropdown menu, and a "Query" button. Below the search fields is a link to "Gene Expression Atlas Home".

2. Repositorios Públicos: EBI



ArrayExpress. La página de resultados permite navegar entre las distintas opciones o enlaces con otras bases de datos como GEO.

Experiment, citation, sample and factor annotations [clear] Filter on [reset] All species All arrays All assays by molecule by All technologies Display options [reset] 25 experiments per page Detailed view Query

ArrayExpress data only [Advanced query syntax](#) [Submitter/reviewer login](#) [ArrayExpress Browser Help](#)

ID	Title	Assays	Species	Date	Processed	Raw	Atlas
E-GEOD-22304	Expression data in response to abiotic stresses in tomato at flowering stage	24	Solanum lycopersicum	2010-06-23			-
Description	A comparative study was made to know the abiotic stress tolerance mechanism between tolerant and susceptible plants at flowering stage. The tolerance in response to abiotic stresses are sum of expression of thousands of genes at a particular stage. Tomato plants were exposed to drought and heat stress for RNA extraction and hybridization on Affymetrix microarrays. To study the molecular mechanism of abiotic stress tolerance to increase the tolerance in tomato plants, transcripts of tolerant and susceptible plants at flowering stage were compared in response to heat and water stress.						
MIAME score	Array designs	Protocols	Factors	Processed data	Raw data		
	✓	-	✓	✓	✓		
Contacts	Ashutosh Rai <ashutoshraiivr@gmail.com>, K N Singh, Major Singh, Neha Jain, P H Pandey, Rajesh Kumar, Shankar R Dubey, Upama Mishra						
Links	GEO - GSE22304 Array design A-AFFY-87 - Affymetrix GeneChip Tomato Genome Array [Tomato] Experimental protocols (old interface) Experiment Page (old interface)						
Files	Data Archives	E-GEOD-22304.processed.1.zip, E-GEOD-22304.raw.1.zip					
	Investigation Description	E-GEOD-22304.idf.txt view ^{NEW}					
	Sample and Data Relationship	E-GEOD-22304.sdrf.txt view ^{NEW}					
	Experiment Design Images	E-GEOD-22304.biosamples.png, E-GEOD-22304.biosamples.svg					
	Array Design	A-AFFY-87.adf.txt					
	Browse all available files						
Experiment type	transcription profiling by array						
Experimental factors	Factor name	Factor values					
	GENOTYPE	Drought Susceptible Genotype, Drought Tolerant Genotype, Heat Susceptible Genotype, Heat Tolerant Genotype					

2. Repositorios Públicos: EBI



Ensembl, conjunto de bases de datos de genomas de eucariotas, mostrados a través de un navegador genómico (Genome Browser) (<http://www.ensembl.org/index.html>).

The screenshot shows the Ensembl website interface. At the top, there is a navigation bar with the Ensembl logo and links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a search box with a dropdown menu for species selection and a 'Go' button. The main content area is divided into several sections: 'Browse a Genome' with a description and links to popular genomes (Human, Mouse, Zebrafish); 'New to Ensembl?' with a list of helpful links; 'Did you know...?' with a CpG island search tip; and 'What's New in Release 62 (13 April 2011)' with a link to new species.

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Login · Register

Search: All species for Go

e.g. BRCA2 or rat X:100000..200000 or coronary heart disease

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online. Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize this list](#))

- Human** GRCh37
- Mouse** NCBI37
- Zebrafish** Zv9

All genomes

-- Select a species --

[View full list of all Ensembl species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

New to Ensembl?

Did you know you can:

- [Learn how to use Ensembl](#) with our video tutorials and walk-throughs
- [Add custom tracks](#) using our new Control Panel
- [Upload and analyse your data](#) and save it to your Ensembl account
- [Search for a DNA or protein sequence](#) using BLAST or BLAT
- [Fetch only the data you want](#) from our public database, using the Perl API
- [Download our databases via FTP](#) in FASTA, MySQL and other formats
- [Mine Ensembl with BioMart](#) and export sequences or tables in text, html, or Excel format

Still got questions? Try our [FAQs](#) or [glossary](#)

Did you know...?

CpG Find CpG islands in [region in detail](#). Turn them on using [configure this page](#).

What's New in Release 62 (13 April 2011)

- [New Species - Gibbon](#)

2. Repositorios Públicos: EBI



Plantas están contenidas en la sección **EnsemblGenomes**, (<http://plants.ensembl.org/index.html>).

The screenshot shows the Ensembl Plants website interface. At the top, there is a navigation bar with the Ensembl Plants logo, a search bar, and links for BLAST, BioMart, Tools, Downloads, and Help & Documentation. The search bar contains the text "All species" and a "Go" button. Below the search bar, there is a section for "Popular genomes" with a link to "Log in to customize this list". This section lists three genomes: Arabidopsis thaliana (TAIR10), Oryza sativa (MSU6), and Zea mays (AGPV2). Below this, there is a section for "All genomes" with a dropdown menu showing "Vitis vinifera" and a link to "View full list of all Ensembl Plants species". At the bottom, there is a section for "What's in Release 9 (April 2011)" with two bullet points: "New core database containing the genome for Oryza glaberrima" and "Updates to cross-references for most core databases". On the right side, there is a section for "Ensembl Genomes" with a description of the project and a list of access methods. The access methods include: "this web browser (go to http://bacteria.ensembl.org, http://metazoa.ensembl.org, etc., or to http://www.ensemblgenomes.org for the project homepage)", "through BioMarts (query optimised data warehouses) constructed for each of the Ensembl Genomes sites (Bacteria Metazoa Protists Fungi Plants)", "via FTP (ftp.ensemblgenomes.org/pub)", "via the Ensembl Genomes public mysql server (mysql.ebi.ac.uk:4157:anonymous).", and "using the Ensembl API". Below the access methods, there is a paragraph about the API and a link to "See the documentation for more information about how to use Ensembl Genomes." At the bottom right, there is a logo for "e!mpowered".

Ensembl Plants BLAST | BioMart | Tools | Downloads | Help & Documentation Login · Register

Search: All species for Go
e.g. Carboxy* or chx28

Popular genomes ([Log in to customize this list](#))

- Arabidopsis thaliana**
TAIR10
- Oryza sativa**
MSU6
- Zea mays**
AGPV2

All genomes
Vitis vinifera
[View full list of all Ensembl Plants species](#)

What's in Release 9 (April 2011)

- New core database containing the genome for *Oryza glaberrima*
- Updates to cross-references for most core databases

Ensembl Genomes

The Ensembl Genomes project produces genome databases for important species from across the taxonomic range, using the Ensembl software system. Five sites are now available: [Ensembl Bacteria](#), [Ensembl Protists](#), [Ensembl Metazoa](#), [Ensembl Plants](#) and [Ensembl Fungi](#). These new sites complement the existing [Ensembl](#) site, with its focus on vertebrate genomes. You can search all Ensembl and Ensembl Genomes databases from the search bar in the top right of this page.

Ensembl Genomes data is available through many of the same routes as Ensembl data. Data can be accessed via:

- this web browser (go to <http://bacteria.ensembl.org>, <http://metazoa.ensembl.org>, etc., or to <http://www.ensemblgenomes.org> for the project homepage).
- through BioMarts (query optimised data warehouses) constructed for each of the Ensembl Genomes sites ([Bacteria](#) [Metazoa](#) [Protists](#) [Fungi](#) [Plants](#))
- via FTP (<ftp.ensemblgenomes.org/pub>)
- via the Ensembl Genomes public mysql server (mysql.ebi.ac.uk:4157:anonymous).
- using the Ensembl API.

The API has been modified slightly to support the existence of "genome collections", i.e. the existence of many small genomes in a single Ensembl database (a model which has been adopted for Ensembl Bacteria). The API makes the use of multi-genome databases transparent to users interested in a single genome, while methods to access a traditional, single-genome database, are unchanged. We aim to keep Ensembl Genomes software in synch with software releases of Ensembl, to ensure that users can access databases from across the taxonomic range using the same software.

See the [documentation](#) for more information about how to use Ensembl Genomes.

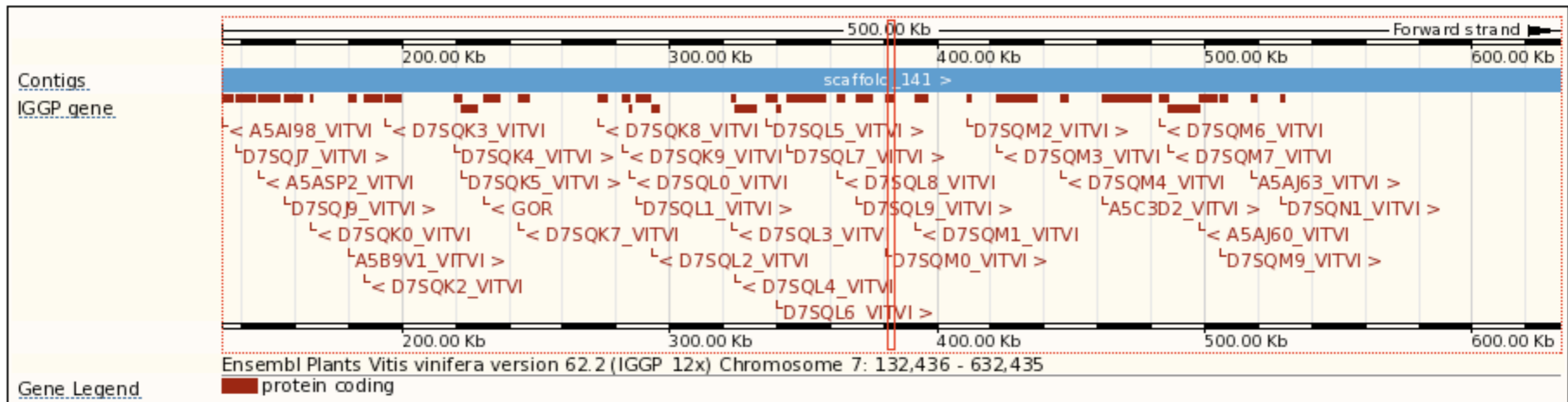
e!mpowered

2. Repositorios Públicos: EBI



EnsemblGenomes, los datos son presentados a traves de un Genome Browser.

Region in detail



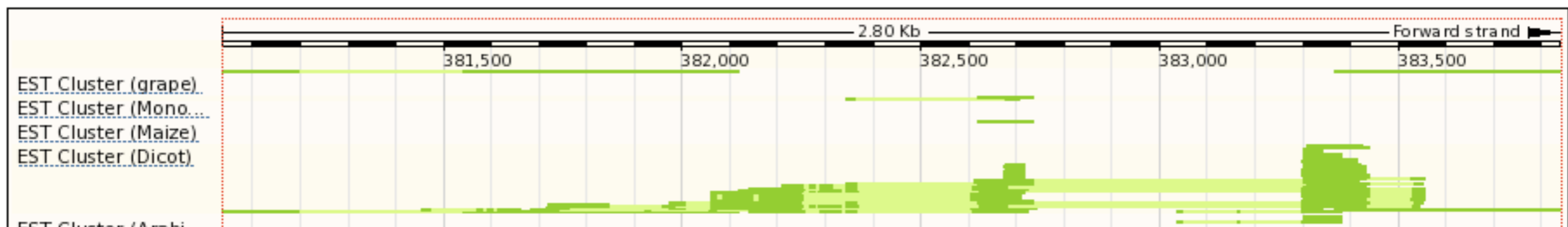
Export Image

Location: 7:381035-383836

Go

Gene:

Go



2. Repositorios Públicos: EBI



InterPro, es una base de datos integrada que contiene dominios de proteínas organizados en superfamilias, familias y subfamilias. Es frecuentemente usada para anotación de genomas y proteínas (<http://www.ebi.ac.uk/interpro/>).

EBI > Databases > InterPro

InterPro protein sequence analysis & classification

InterPro is an integrated database of predictive protein "signatures" used for the classification and automatic annotation of proteins and genomes. InterPro classifies sequences at superfamily, family and subfamily levels, predicting the occurrence of functional domains, repeats and important sites. InterPro adds in-depth annotation, including GO terms, to the protein signatures.

Current release: **32.0 18th April 2011** (see [Release Notes](#) for further details)

Search InterPro:

Do a sequence search of InterPro, via [InterProScan](#)

Extract large datasets by querying our [BioMart](#)

You can access our data programmatically, via [Web Services](#)

Use the updated [InterProScan Web Service](#)

If you have any questions or feedback please [contact us](#).



InterPro 32.0

2. Repositorios Públicos: EBI



InterPro: Una de las herramientas de búsqueda utilizada es InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>)

InterProScan Sequence Search

This form allows you to query your sequence against InterPro. For more detailed information see the documentation for the perl stand-alone InterProScan package ([Readme file](#) or [FAQ's](#)), or the [InterPro user manual](#) or [help pages](#).

Use this tool

STEP 1 - Enter your input sequence

Enter or paste a **PROTEIN** sequence in any supported format:

Or, upload a file: No file chosen

STEP 2 - Select the applications to run

Select All Clear All

<input checked="" type="checkbox"/> BlastProDom	<input checked="" type="checkbox"/> FPrintScan	<input checked="" type="checkbox"/> HMMPiR	<input checked="" type="checkbox"/> HMMPfam	<input checked="" type="checkbox"/> HMMSmart
<input checked="" type="checkbox"/> HMMTigr	<input checked="" type="checkbox"/> ProfileScan	<input checked="" type="checkbox"/> HAMAP	<input checked="" type="checkbox"/> PatternScan	<input checked="" type="checkbox"/> SuperFamily
<input checked="" type="checkbox"/> SignalPHMM	<input checked="" type="checkbox"/> TMHMM	<input checked="" type="checkbox"/> HMMPanther	<input checked="" type="checkbox"/> Gene3D	

STEP 3 - Submit your job

Be notified by email (*Tick this box if you want to be notified by email when the results are available*)



Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.



3. Bases de datos específicas de comunidades



Nombre	Especies	Datos	Enlace
The Arabidopsis Information Resource (TAIR)	Arabidopsis	Distintos Genomas, Marcadores, Polimorfismos, Genes, Expresión, Proteínas, Ontologías, Rutas Metabólicas, Publicaciones	http://www.arabidopsis.org/
Gramene	Monocotiledóneas (Uva y Arabidopsis)*	Distintos Genomas, Marcadores, Polimorfismos, Genes, Proteínas, Ontologías, Rutas Metabólicas, QTLs	http://www.gramene.org/
Sol Genomics Network (SGN)	Solanáceas, Rubicáceas y Plantagináceas	Distintos Genomas, Marcadores, Polimorfismos, Genes, Expresión*, Proteínas, Ontologías, Rutas Metabólicas, Publicaciones, QTLs, Fenotipos, Unigenes	http://solgenomics.net/
Genome Database for Rosaceae (GDR)	Rosáceas	Distintos Genomas, Marcadores, Polimorfismos, Genes, Proteínas, Ontologías, Unigenes	http://www.rosaceae.org/
Dendrome	Árboles	Distintos Genoma, Genes, Polimorfismos, Marcadores, Mapas, Fenotipos, Literatura, Expresion	http://dendrome.ucdavis.edu/
Curcubit Genomics Database (ICuGI)	Curcubitáceas	Distintos Genomas, Genes, Unigenes, Marcadores, Mapas, Rutas Metabólicas	http://www.icugi.org/

3. Bases de datos específicas de comunidades



Nombre	Especies	Datos	Enlace
The Compositae Genome Project	Compositaceas	Distintos Genomas, Marcadores, Mapas, Genes	http://compgenomics.ucdavis.edu/
MaizeGDB	Zea mays (Maiz)	Distintos Genomas, Marcadores, Polimorfismos, Genes, QTLs, citogenética, Fenotipos.	http://www.maizegdb.org/
CottonDB	Gossypium spp.	Marcadores, Mapas, Genes y Datos Taxonómicos.	http://cottondb.org/
SoyBase	Glycine Max (Soja)	Distintos Genomas, Marcadores, Polimorfismos, Genes, Proteínas, Ontologías, Rutas Metabólicas, QTLs	http://soybase.org/
Cacao Genome Database	Theobroma cacao	Genome, Genes, Maps	http://www.cacaogenomedb.org/
Plant Genome Database (PlantGDB)	Plantas	Distintos Genomas, Genes, Unigenes	http://www.plantgdb.org/



Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.



3. Bases de datos específicas de comunidades



Existen otras bases de datos, no basadas en una especie, sino en una area de conocimiento específica, por ejemplo:

Bases de datos metabólicas:

MetaCyc: <http://metacyc.org/>

KEGG: <http://www.genome.jp/kegg/>

Bases de datos de Ontologías:

Gene Ontology: <http://www.geneontology.org/>

Plant Ontology: <http://www.plantontology.org/>

Bases de datos de Factores de Transcripción:

TranscriptionFactorDB (DBD): www.transcriptionfactor.org



Bases de Datos Biológicas:

1- Tipos de Bases de Datos Biológicas.

2- Repositorios Públicos.

3- Bases de datos de comunidades científicas.

3.1- Especie específicas.

3.2- Otras bases de datos.

4- Navegadores Genómicos.



4. Navegadores Genómicos

Un Navegador Genómico (**Genome Browsers**) es una interfaz gráfica que muestra datos genómicos de forma alineada.

Generalmente cada tipo de dato se encuentra en una pista (**track**).

Las pistas se encuentran jerárquicamente organizadas por el tamaño de la pista. Así por ejemplo, la primera pista podría ser un cromosoma completo, la segunda una región de este, y la tercera.



4. Navegadores Genómicos

Los Navegadores Genómicos más conocidos son:

- GBrowse (GMOD).
- UCSC Genome Browser.
- Ensembl Genome Browser.
- Vista Genome Browser.
- JBrowse (GMOD).



4. Navegadores Genómicos

Los Navegadores Genómicos más conocidos son:

- GBrowse (GMOD).
- UCSC Genome Browser.
- Ensembl Genome Browser.
- Vista Genome Browser.
- JBrowse (GMOD).



GBrowse (Generic Genome Browser):

Es uno de los componentes más populares del proyecto GMOD, usado por bases de datos como FlyBase, Wormbase, SGN, Gramene GDR.

Es una combinación de base de datos y páginas web interactivas para manipular y mostrar anotaciones sobre genomas.

Está organizado en tres niveles de detalle: **Overview**, **Region** y **Detail**. Cada nivel de detalle posee pistas (**tracks**) muestran un tipo de información.





GBrowse (Generic Genome Browser):

Las características más importantes son:

- Tres vistas simultáneas con distinto nivel de detalle.
- Desplazamientos, ampliaciones y centrados.
- Capacidad de adjuntar URL arbitrarias a cualquier localización.
- Capacidad de ordenar pista por el usuario.
- Búsqueda por nombre, ID o anotación.
- Soporta anotaciones externas en formato .gff3.
- Volcado de datos en formato .fasta y .gff3.
- Conectividad entre distintas bases de datos como chado.

4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

The screenshot shows the GBrowse web interface. At the top, there are menu items 'File' and 'Help'. Below that, a title bar reads 'ITAG2 Release: genomic annotations: 4.549 kbp from SL2.31ch10:62,005,772..62,010,320'. The main interface has several sections: 'Browser' with tabs for 'Select Tracks', 'Upload and Share Tracks', and 'Preferences'; a 'Search' section with a 'Landmark or Region' input field containing 'SL2.31ch10:62005772..€' and a 'Search' button; a 'Data Source' dropdown menu set to 'ITAG2 Release: genomic annotations'; and a 'Scroll/Zoom' section with navigation arrows and a 'Show 4.549 kbp' dropdown. On the left, there are expandable sections for 'Overview', 'Region', and 'Details', along with a 'Select Tracks' button. Two blue callout boxes with arrows point to the 'Landmark or Region' field (labeled '2') and the 'Data Source' dropdown (labeled '1'). To the right of the interface, the text 'Región' and 'Fuente (Set de datos)' are written with arrows pointing to the respective callouts.

Región:

Secuencia de Referencia + “:” + Coordenada Inicial + “..” + Coordenada final.

Ejemplo:

SL2.40ch08:56810100..56812450



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 2.351 kbp from SL2.40ch08:56,810,100..56,812,450

Browser Select Tracks Upload and Share Tracks Preferences

Search

Landmark or Region: SL2.40ch08:56810100..5 Search

Examples: Solyc05g005010, SL2.40ch08:2626000..2648999, Sulfite oxidase, SGN-M100, SGN-M7588.

Data Source ITAG2.3 Release: genomic annotations

Annotate Restriction Sites Configure... Go

4 Scroll/Zoom: << < - + > >> Show 2.351 kbp Flip

1 Overview
SL2.40ch08
0M 10M 20M 30M 40M 50M 60M

2 Region
56800k 56900k

3 Details
500 bp
SL2.40ch08: 2.351 kbp
56811k 56812k

Select Tracks Clear highlighting

Select Tracks

Tres niveles de detalle: **Overview**, **Region** y **Details**



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Browser Select Tracks Upload and Share Tracks Preferences

Search

Landmark or Region:

SL2.40ch08:56806275..5 Search

Examples: Solyc05g005010, SL2.40ch08:2626000..2648999, Sulfite oxidase, SGN-M100, SGN-M7588.

Data Source

ITAG2.3 Release: genomic annotations

Annotate Restriction Sites

Configure...

Go

Nivel de Zoom sobre **Details**

4

Scroll/Zoom:



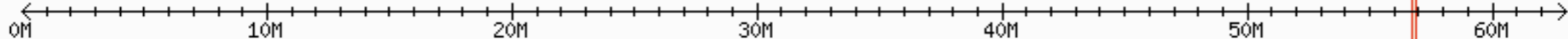
Show 10 kbp



Flip

Overview

SL2.40ch08



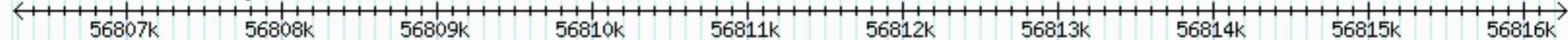
Region



5

Details

SL2.40ch08: 10 kbp



Select Tracks

Clear highlighting

Select Tracks

Botones de desplazamiento *Downstream* y *Upstream* region=**Details**



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [Solyc05g005010](#), [SL2.40ch08:2626000..2648999](#), [Sulfite oxidase](#), [SGN-M100](#), [SGN-M7588](#).

Data Source:

Scroll/Zoom: Flip

Overview: SL2.40ch08

Region: 56800k 56900k

Details: SL2.40ch08: 10 kbp

56807k 56808k 56809k 56810k 56811k 56812k 56813k 56814k 56815k 56816k

6

Selección de las pistas de información (**Tracks**)



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Browser [Select Tracks](#) [Upload and Share Tracks](#) [Preferences](#)

Search

Landmark or Region:

Examples: [Solyc05g005010](#), [SL2.40ch08:2626000..2648999](#), [Sulfite oxidase](#), [SGN-M100](#), [SGN-M7588](#).

Data Source:

Scroll/Zoom: Flip

Overview
SL2.40ch08
0M 10M 20M 30M 40M 50M 60M

Region
56800k 56900k

Details
SL2.40ch08: 10 kbp
5 kbp
56807k 56808k 56809k 56810k 56811k 56812k 56813k 56814k 56815k 56816k

6

Selección de las pistas de información (**Tracks**)



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

Browser **Select Tracks** Upload and Share Tracks Preferences

<< Back to Browser

Tracks

7 **Overview** All on All off

Infernal Markers tRNAscanSE

8 **Region** All on All off

GC content Infernal

Genes Markers

9 **DNA** All on All off

DNA/GC content 6-frame translation

Gene models All on All off

CDS - click to browse protein Gene models KD plot/peptides

Genetic loci All on All off

SGN Locus Sequences SGN markers

Genome data and reagents All on All off

SL2.40 assembly SL1.00 / ITAG1 scaffolds Tomato BACs

ESTs and cDNAs - Other Solanaceae MicroTom full-length cDNAs

ESTs and cDNAs - Tomato SGN unigenes

Other genomes All on All off

Potato BACs Tobacco genome contigs

Paired clone-end mapping All on All off

All clones All clones (linking)

Selección de las pistas de información (**Tracks**)



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

Browser **Select Tracks** Upload and Share Tracks Preferences

<< Back to Browser

Tracks

7 **Overview** All on All off

Infernal Markers tRNAscanSE

8 **Region** All on All off

GC content Infernal

Genes Markers

9 **DNA** All on All off

DNA/GC content 6-frame translation

Gene models All on All off

CDS - click to browse protein Gene models KD plot/peptides

Genetic loci All on All off

SGN Locus Sequences SGN markers

Genome data and reagents All on All off

SL2.40 assembly SL1.00 / ITAG1 scaffolds Tomato BACs

ESTs and cDNAs - Other Solanaceae MicroTom full-length cDNAs

ESTs and cDNAs - Tomato SGN unigenes

Other genomes All on All off

Potato BACs Tobacco genome contigs

Paired clone-end mapping All on All off

All clones All clones (linking)

Selección de las pistas de información (**Tracks**)



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Browser Select Tracks Upload and Share Tracks Preferences

Search

Landmark or Region:

SL2.40ch08:56806275..5 Search

Annotate Restriction Sites

Configure...

Go

Examples: Solyc05g005010, SL2.40ch08:2626000..2648999, Sulfite oxidase, SGN-M100, SGN-M7588.

Data Source

ITAG2.3 Release: genomic annotations

Scroll/Zoom: << < -

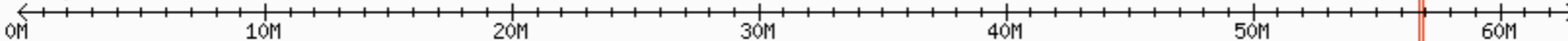
Show 10 kbp

+ > >>

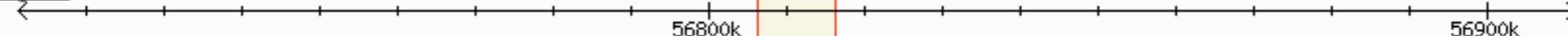
Flip

Overview

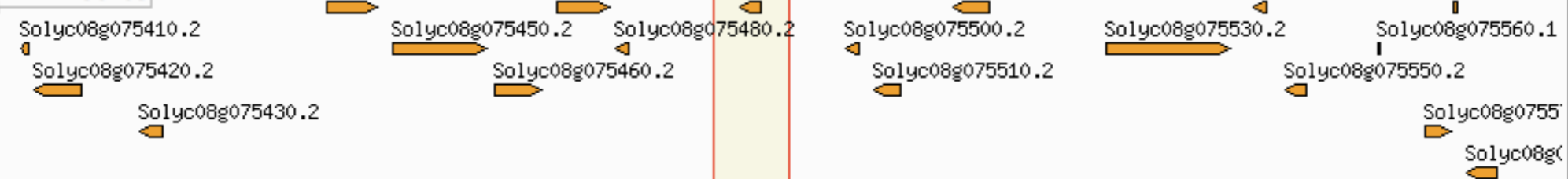
SL2.40ch08



Region

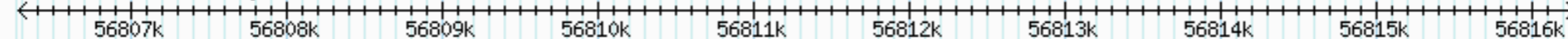


Genes

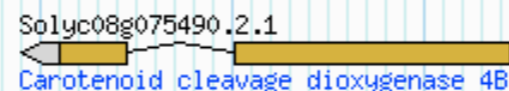


Details

SL2.40ch08: 10 kbp



Gene models



ESTs and cDNAs - Other Solanaceae

Select Tracks

Clear highlighting

7

8

9

4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

File Help

ITAG2.3 Release: genomic annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Browser Select Tracks Upload and Share Tracks Preferences

Search

Landmark or Region:

SL2.40ch08:56806275..5 Search

Annotate Restriction Sites

Configure...

Go

Examples: Solyc05g005010, SL2.40ch08:2626000..2648999, Sulfite oxidase, SGN-M100, SGN-M7588.

Data Source

ITAG2.3 Release: genomic annotations

Scroll/Zoom:



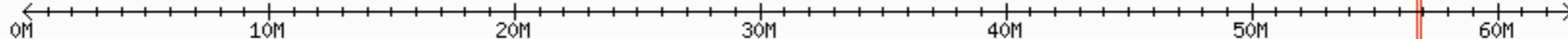
Show 10 kbp



Flip

Overview

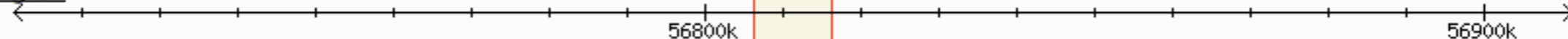
SL2.40ch08



Markers

Markers

Region

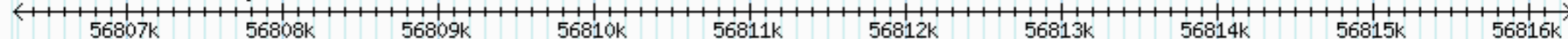


Markers

Markers

Details

SL2.40ch08: 10 kbp



SGN markers

SGN markers

C2_At4g19170



Select Tracks

Clear highlighting

Select Tracks

7

8

9



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

10

File Help

- Bookmark this
- Share these tracks
- Export as...
 - ...low-res PNG image
 - ...editable SVG image
 - ...GFF annotation table
 - ...FASTA sequence file
- Get chrom sizes
- Reset to defaults

Examples: Solyc05g005010, SL oxidase, SGN-M100, SGN-M75E

Data Source: ITAG2.3 Release: genomic annotations

Annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

Annotate Restriction Sites [Configure...] [Go]

Scroll/Zoom: [Navigation icons] [Show 10 kbp] [Flip]

Overview: SL2.40ch08

Region: 56800k - 56900k

Gene: C2_At4g19170

Marker: SSR41

Volcado de datos en distintos formatos.



4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

10

File Help ▾

- Bookmark this
- Share these tracks
- Export as...**
- Get chrom sizes
- Reset to defaults

Examples: [Solyc05g005010](#), [SL oxidase](#), [SGN-M100](#), [SGN-M758](#)

Data Source: ITAG2.3 Release: genomic annotations

Overview: SL2.40ch08

Region: 56800k

Annotations: 10 kbp from **SL2.40ch08:56,806,275..56,816,274**

- ...low-res PNG image
- ...editable SVG image
- ...GFF annotation table
- ..FASTA sequence file**

```
Markers_region+Genes_region+mrna+sgn_unigenes+sgn_loci+Markers_over...
>SL2.40ch08:56810100..56812450
TTATAGCTTCATAAGATCATTTCCTCCTCAGTCCAGTCCATGGAACCGTAAGGCACACG
ACGAGGCAATTTGACATTAGCCACAATGTCAAGATTAGGTGATGTTGCATCCATGACCAA
AAAAATTTGATTCCTCCTCTCATTGTGTACGTATGACACCACGTAGCCATCGTCTTC
ATCAATCGAAGGATTTTAGGCACGAAAAATGGTTCACCACCATAGCAATCTTTTCCAAA
TATACGACATGCCACGATGCAATCGCGACGATCAATTTCTGCTACGGATACGCTAATTT
TGCTATCCCTATTACTTTTGGCATAGGACCCCAATGGCTGCATATACGTACCTACACGA
AACAAAAAATTTTATATGTCTGTAAATTTTGGTTCGACACGAAATTTAAAAATGATT
TTTTGTTTGTATGGTCTCTGCTGAAATTAATCATGTGTTTCATAAGTCATAAAGGACCTA
CACATGTCCATATGCTGCACTATAGATGGTCCAAACAAAAATCAATAAATTTATAATAGTG
TTATATACTAGTACTAGTACTGTCAATGTTTATTTATTTGGTGCACACATTAATAAATGTTG
ATAGAATAAAGTTATAATTTTATATATTTTATTTTAAAAATTTGGTATATAAAAAAT
GCAAATTTATTTTGGATTTTCAAATCAAAGAGTAAAAATAGATACATATTTTTCAGTT
TAGTGGATTAGTAAAAAATGAGGAGAGTGAATTTGTTCTATAAAAGTACCGAGCTA
CTCCTTTGCTCAAATTAATGGCATAATTTGTTTTCGAGATTCGATTTAACTAATTTTA
CGATTAATATGTATGTATACTTATATTAATGAATATCAAGTAAGTTACTGAAGATAAAT
GGATATTTAAAAACGTAATTTGTTCTTTTCCAAACATAAGCTGGATTGATGACTCCAAAA
TCAAGATTTCTGTAGAAATTTGGATGTCTGCTCACCATTCTCTGCTTCAAATTTATTTTC
ACTTTCTCAACACTTGCATGTATCATGTCCATCTCTCTAGTGTATGTTCCACCGATAAT
ATATTCGGCGCGATCAACACTATCGTATCACCCTCCATCCCTCATCCACCGGTTATCGCG
TGTACAATATTAATCTGTATACATCGAACCCCTCATTTCGACTCATCTTTGGCGTAA
CGTGGTATCACACCAATTCGGGAATTTCCCGAGTCAAGTACCCACCGTGAACCCCG
GTGAGTAAATCAATTTGGATTCATTTCAATTTGTATGTCGAAAAATATCGCGTATTTTTT
GTGATTGCGAAGTCGTAAGAAATGAGGGGCGGGTATTGAGAATATGGGTACATCTGTTG
GTTTTGTACCATCCGGTTGACCCGAAAGTAAGTTATAAACGGAGGTATCGGGCCGTAA
CGAAAGCAAAGCCCTCGTTAGTTTCGTGGTTCGATTTTCGGATGTGCCGTCATGCTCATA
GAAAGCTTTCCGTTAAAAATCGTGACCGCGAGGGTAAAAATGTCACCATTTGGGGCTATT
TTACCTCATATGGTAAATCAGATTCACCGAGGGCGAAAAAGTTTACCTCCAAATAAAGCT
AAGCTTGTATTCGCAAGGCCTATACCGTTTCCGGGATTAAGTGTCCCGCAATAGCTCGG
GCCGCGGTGAGTGCACCACGCGCGCCGAGGCGTTAGACCGTTGAAACCGGAGAACAACA
TTAGGAATAATGGGTGACCCGGCTTCATTTTTCGATGTTGACTTGTAAAGTTTGCACAAAT
CGACTACATAACGTAGCTTTACCTTGGGAAATTTAATGGAGTGAAGCATTCCATCACCG
TCGAATAGATGGTAAGGTCCACGTGGAAGATATTGAGGGTTAGGTCCATTTCTAATATAC
GCCCATCTAGGCAAGGTGGAAGGAGCCTTCCACTACTTCGATTCGGTAGGAGGAAGT
TCGCTACTGGAGCGAAGTTGTTAGAGAGAACATACCTTGGATCAACAGACGATTTCCCTA
GGTGGATCAATAAAGTATTTACGAAATCATCGAATGCATTGAAGATAAAGAGGGAAAT
GATGGTTCAATAGGTTTTCTTGGGGTATTGGTTCTTTTTGGAGATGGTTTTGGTGGT
GAAGGGTTGATTTTTCTTGTGGTTTTGTTATAGTTGTGGTTGTTTGGTGGTCTTTCTCA
ATTCTAACAGAAAAATCTTTTAGAGTAGGAGAATAGTAATGATAATTTATTTATTTATTA
TTATAAGGAGAAAGAGATTTAGGGTTTTGTGATAATGTAGAAGGAAAGTTGAAGAC
```

4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

10

File Help

Bookmark this

Share these tracks

Export as...

Get chrom sizes

Reset to defaults

Examples: Solyc05g005010, SL oxidase, SGN-M100, SGN-M756

Data Source

ITAG2.3 Release: genomic annotat

Overview

SL2.40ch08

Markers

Region

Markers

Details

annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

...low-res PNG image

...editable SVG image

...GFF annotation table

...FASTA sequence file

Annotate Restriction Sites

Configure...

Go

```
##gff-version 3
##date Sun Jun 12 18:54:23 2011
##source gbrowse gbgff gff3 dumper
##genome-build unknown
##sequence-region SL2.40ch08:56810100..56812450
SL2.40ch08 ITAG_eugene gene 56809898 56812517 . . . Name=Solyc08g075490.2;ID=2441230;Alias=Solyc08g075490;from_BOGAS=1;length=2620
SL2.40ch08 ITAG_eugene mRNA 56809898 56812517 . . . Name=Solyc08g075490.2.1;Parent=2441230;ID=2441231;from_BOGAS=1;sifter_term=60:0005515;Note=Carotenoid%20c leavage%20di oxygenase%204B;length=2620;Ontology_term=G0:0005515;nb_exon=2
SL2.40ch08 ITAG_eugene exon 56809898 56810451 . . . Parent=2441231;ID=2441232;from_BOGAS=1
SL2.40ch08 ITAG_eugene three_prime_UTR 56809898 56810099 . . . Parent=2441231;ID=2441233;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56810100 56810451 . . 1 Parent=2441231;ID=2441234;from_BOGAS=1
SL2.40ch08 ITAG_eugene intron 56810452 56811017 . . . Parent=2441231;ID=2441235;from_BOGAS=1
SL2.40ch08 ITAG_eugene exon 56811018 56812517 . . 1 Parent=2441231;ID=2441236;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56811018 56812450 . . 0 Parent=2441231;ID=2441237;from_BOGAS=1
SL2.40ch08 ITAG_eugene five_prime_UTR 56812451 56812517 . . . Parent=2441231;ID=2441238;from_BOGAS=1
SL2.40ch08 ITAG_eugene mRNA 56809898 56812517 . . . Name=Solyc08g075490.2.1;ID=2441231;from_BOGAS=1;sifter_term=60:0005515;Note=Carotenoid%20c leavage
SL2.40ch08 ITAG_eugene exon 56809898 56810451 . . . Parent=2441231;ID=2441232;from_BOGAS=1
SL2.40ch08 ITAG_eugene three_prime_UTR 56809898 56810099 . . . Parent=2441231;ID=2441233;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56810100 56810451 . . 1 Parent=2441231;ID=2441234;from_BOGAS=1
SL2.40ch08 ITAG_eugene intron 56810452 56811017 . . . Parent=2441231;ID=2441235;from_BOGAS=1
SL2.40ch08 ITAG_eugene exon 56811018 56812517 . . 1 Parent=2441231;ID=2441236;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56811018 56812450 . . 0 Parent=2441231;ID=2441237;from_BOGAS=1
SL2.40ch08 ITAG_eugene five_prime_UTR 56812451 56812517 . . . Parent=2441231;ID=2441238;from_BOGAS=1
SL2.40ch08 ITAG_sgn_markers match 56809921 56811225 0.915 . . Name=C2_At4g19170%20C%20SGN-M8299
SL2.40ch08 ITAG_sgn_markers match_part 56809921 56810451 0.889 . . Parent=6137688;ID=6137689
SL2.40ch08 ITAG_sgn_markers match_part 56811018 56811225 0.981 . . Parent=6137688;ID=6137695
SL2.40ch08 ITAG_sgn_unigenes match 56809921 56812128 0.933 . . Name=SGN-U272543;ID=6137690;Target=SGN-U272543%201%201653%20+
SL2.40ch08 ITAG_sgn_unigenes match_part 56809921 56810451 0.886 . . Parent=6137690;ID=6137691
SL2.40ch08 ITAG_sgn_unigenes match_part 56811018 56812128 0.955 . . Parent=6137690;ID=6137696
SL2.40ch08 ITAG_sgn_unigenes match 56809937 56812326 0.998 . . Name=SGN-U570287;ID=6137692;Target=SGN-U570287%201%201824%20+
SL2.40ch08 ITAG_sgn_unigenes match_part 56809937 56810451 0.998 . . Parent=6137692;ID=6137693
SL2.40ch08 ITAG_sgn_unigenes match_part 56811018 56812326 0.998 . . Parent=6137692;ID=6137697
SL2.40ch08 ITAG_sgn_loci match 56810100 56812450 0.954545454545455 . . Name=carotenoid%20c leavage
%20di oxygenase;ID=6137694;Target=Potato_SGNlocusID_1962_DQ206633%201%20418%20+;blat_cdna_base_mismatches=19;cdna_seq_name=Solyc08g075490.1.1;blat_cdna_match_pct=95.45;blat_cdna_t_gaps=0;blat_cdn
a_t_gap_bases=0;Alias=cod,DQ206633,Potato_SGNlocusID_1962_DQ206633;sgn_locus_id=1962;blat_cdna_q_gaps=0;blat_cdna_q_gap_bases=0
```

4. Navegadores Genómicos



Ejemplo: GBrowse en el SGN.

10

File Help

Bookmark this

Share these tracks

Export as...

Get chrom sizes

Reset to defaults

Examples: Solyc05g005010, SL oxidase, SGN-M100, SGN-M756

Data Source

ITAG2.3 Release: genomic annotat

Overview

SL2.40ch08

Markers

Region

Markers

Details

annotations: 10 kbp from SL2.40ch08:56,806,275..56,816,274

...low-res PNG image

...editable SVG image

...GFF annotation table

...FASTA sequence file

Annotate Restriction Sites

Configure...

Go

```
##gff-version 3
##date Sun Jun 12 18:54:23 2011
##source gbrowse gbgff gff3 dumper
##genome-build unknown
##sequence-region SL2.40ch08:56810100..56812450
SL2.40ch08 ITAG_eugene gene 56809898 56812517 . . . Name=Solyc08g075490.2;ID=2441230;Alias=Solyc08g075490;from_BOGAS=1;length=2620
SL2.40ch08 ITAG_eugene mRNA 56809898 56812517 . . . Name=Solyc08g075490.2.1;Parent=2441230;ID=2441231;from_BOGAS=1;sifter_term=60:0005515;Note=Carotenoid%20c leavage%20di oxygenase%204B;length=2620;Ontology_term=G0:0005515;nb_exon=2
SL2.40ch08 ITAG_eugene exon 56809898 56810451 . . . Parent=2441231;ID=2441232;from_BOGAS=1
SL2.40ch08 ITAG_eugene three_prime_UTR 56809898 56810099 . . . Parent=2441231;ID=2441233;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56810100 56810451 . . 1 Parent=2441231;ID=2441234;from_BOGAS=1
SL2.40ch08 ITAG_eugene intron 56810452 56811017 . . . Parent=2441231;ID=2441235;from_BOGAS=1
SL2.40ch08 ITAG_eugene exon 56811018 56812517 . . 1 Parent=2441231;ID=2441236;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56811018 56812450 . . 0 Parent=2441231;ID=2441237;from_BOGAS=1
SL2.40ch08 ITAG_eugene five_prime_UTR 56812451 56812517 . . . Parent=2441231;ID=2441238;from_BOGAS=1
SL2.40ch08 ITAG_eugene mRNA 56809898 56812517 . . . Name=Solyc08g075490.2.1;ID=2441231;from_BOGAS=1;sifter_term=60:0005515;Note=Carotenoid%20c leavage
%20di oxygenase%204B;length=2620;Ontology_term=G0:0005515;nb_exon=2
SL2.40ch08 ITAG_eugene exon 56809898 56810451 . . . Parent=2441231;ID=2441232;from_BOGAS=1
SL2.40ch08 ITAG_eugene three_prime_UTR 56809898 56810099 . . . Parent=2441231;ID=2441233;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56810100 56810451 . . 1 Parent=2441231;ID=2441234;from_BOGAS=1
SL2.40ch08 ITAG_eugene intron 56810452 56811017 . . . Parent=2441231;ID=2441235;from_BOGAS=1
SL2.40ch08 ITAG_eugene exon 56811018 56812517 . . 1 Parent=2441231;ID=2441236;from_BOGAS=1
SL2.40ch08 ITAG_eugene CDS 56811018 56812450 . . 0 Parent=2441231;ID=2441237;from_BOGAS=1
SL2.40ch08 ITAG_eugene five_prime_UTR 56812451 56812517 . . . Parent=2441231;ID=2441238;from_BOGAS=1
SL2.40ch08 ITAG_sgn_markers match 56809921 56811225 0.915 . . Name=C2_At4g19170%20C%20SGN-M8299
%20name(s):%20C2_At4g19170%20C%20SGN-M8299
SL2.40ch08 ITAG_sgn_markers match_part 56809921 56810451 0.889 . . Parent=6137688;ID=6137689
SL2.40ch08 ITAG_sgn_markers match_part 56811018 56811225 0.981 . . Parent=6137688;ID=6137695
SL2.40ch08 ITAG_sgn_unigenes match 56809921 56812128 0.933 . . Name=SGN-U272543;ID=6137690;Target=SGN-U272543%201%201653%20+
SL2.40ch08 ITAG_sgn_unigenes match_part 56809921 56810451 0.886 . . Parent=6137690;ID=6137691
SL2.40ch08 ITAG_sgn_unigenes match_part 56811018 56812128 0.955 . . Parent=6137690;ID=6137696
SL2.40ch08 ITAG_sgn_unigenes match 56809937 56812326 0.998 . . Name=SGN-U570287;ID=6137692;Target=SGN-U570287%201%201824%20+
SL2.40ch08 ITAG_sgn_unigenes match_part 56809937 56810451 0.998 . . Parent=6137692;ID=6137693
SL2.40ch08 ITAG_sgn_unigenes match_part 56811018 56812326 0.998 . . Parent=6137692;ID=6137697
SL2.40ch08 ITAG_sgn_loci match 56810100 56812450 0.954545454545455 . . Name=carotenoid%20c leavage
%20di oxygenase;ID=6137694;Target=Potato_SGNlocusID_1962_DQ206633%201%20418%20+;blat_cdna_base_mismatches=19;cdna_seq_name=Solyc08g075490.1.1;blat_cdna_match_pct=95.45;blat_cdna_t_gaps=0;blat_cdn
a_t_gap_bases=0;Alias=cod,DQ206633,Potato_SGNlocusID_1962_DQ206633;sgn_locus_id=1962;blat_cdna_q_gaps=0;blat_cdna_q_gap_bases=0
```

4. Navegadores Genómicos



Formato: GFF3 (<http://www.sequenceontology.org/gff3.shtml>)

Columna 1: Identificación de la secuencia, "seqid", ejemplo: SL2.40ch08

Columna 2: Fuente, "source", ejemplo: ITAG_eugene

Columna 3: Tipo, "type", ejemplo: mRNA

Columnas 4 & 5: Comienzo y fin, "start" and "end", ejemplo: 56809898 y 56812517

Columna 6: Puntuación, "score", ejemplo: . (vacío)

Columna 7: Hebra, "strand", ejemplo: - (negativa)

Columna 8: Fase, "phase", ejemplo: . (vacío)

Columna 9: Atributos, "attributes", ejemplo:

Name=Solyc08g075490.2.1;length=2620;Ontology_term=GO:0005515;



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Unix y líneas de comandos.

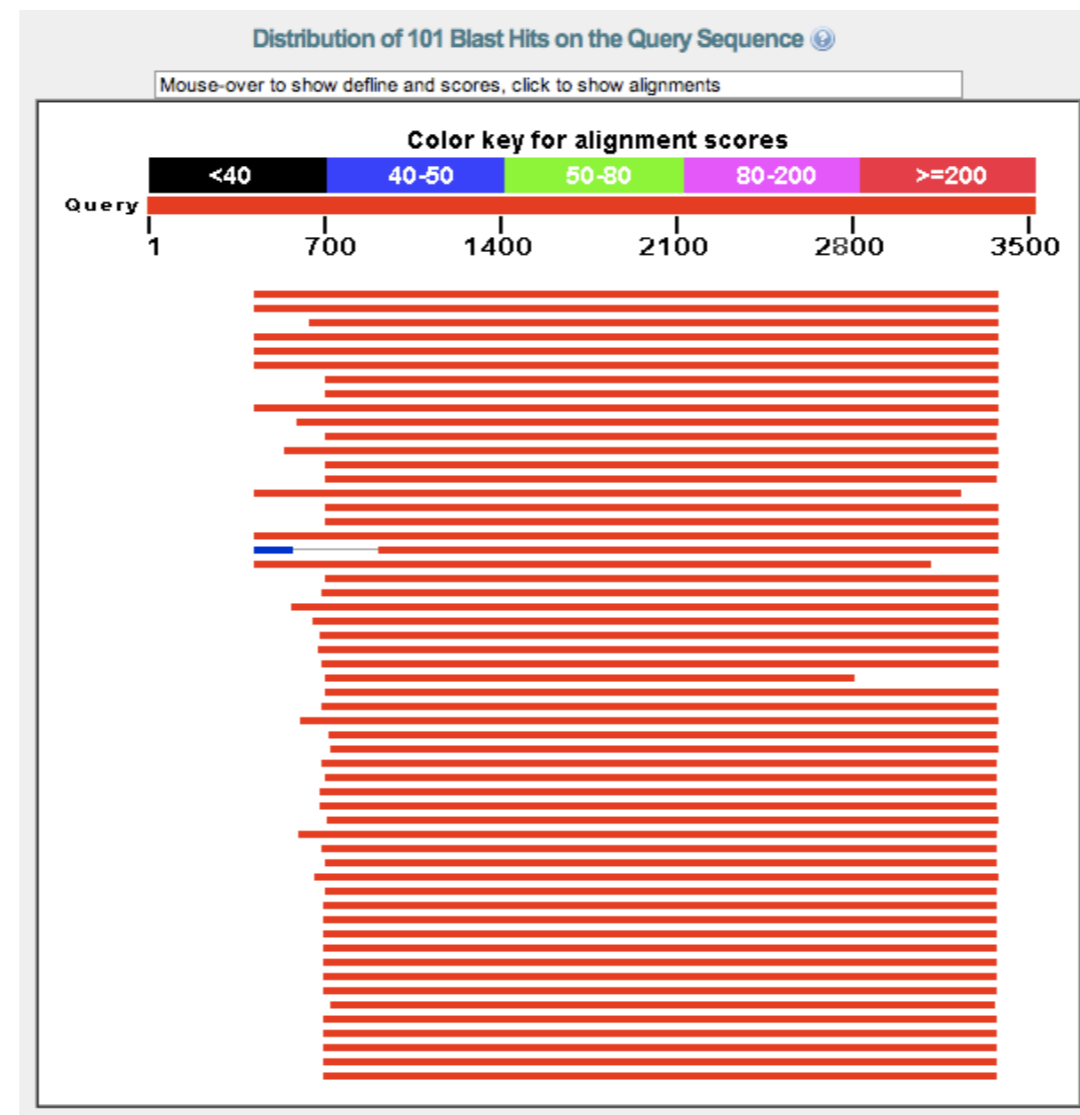
15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.





Herramientas de las Bases de Datos Biológicas.





Herramientas de las Bases de Datos Biológicas:

1- Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Herramientas de las Bases de Datos Biológicas:

1- Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



I. Herramientas de búsqueda

Las herramientas de búsqueda son una parte esencial de la base de datos.

Las herramientas de búsqueda han de ser intuitivas y fáciles de manejar, rápidas y reproducibles.

En las bases de datos biológicas existen varios tipos de búsqueda:

- Búsqueda de texto (palabras clave, anotaciones, ontologías)
- Búsqueda por homología de secuencias.
- Búsqueda por localizaciones (coordenadas)
- Búsqueda por valores numéricos de señal (ratios, correlaciones, perfiles, *clusters*)

I. Herramientas de búsqueda



Búsqueda de texto:

Las palabras introducidas en la sección (caja, *box*) de búsqueda serán utilizadas para buscar coincidencias en determinados campos, por ejemplo descripción de secuencias.

NCBI: <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI homepage with the following elements:

- Search Bar:** Located at the top, with a dropdown menu set to 'All Databases'. A red arrow points to the search input field.
- Navigation Menu (Left):** Includes links for 'NCBI Home', 'Site Map (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', and 'Homology'.
- Welcome to NCBI:** A central message stating: 'The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.' Below this are links for 'About the NCBI | Mission | Organization | Research | RSS Feeds'.
- Get Started:** A section with four bullet points: 'Tools: Analyze data using NCBI software', 'Downloads: Get NCBI data or software', 'How-To's: Learn how to accomplish specific tasks at NCBI', and 'Submissions: Submit data to GenBank or other NCBI databases'.
- Popular Resources (Right):** A list of resources including 'BLAST', 'Bookshelf', 'Gene', 'Genome', 'Nucleotide', 'OMIM', 'Protein', 'PubChem', 'PubMed', 'PubMed Central', and 'SNP'.



Búsqueda de texto:

EBI: <http://www.ebi.ac.uk/>

EMBL-EBI European Bioinformatics Institute

Databases Tools Research Training Industry About Us Help Site Index

Explore the EBI:

FIND

Examples: [ROA1_HUMAN](#), [tpi1](#), [Sulston...](#) [Help](#) [Feedback](#)

Data Resources and Tools

- [ENA](#)
- [UniProt](#)
- [ArrayExpress](#)
- [Ensembl](#)
- [InterPro](#)
- [PDBe](#)
- [Genomes](#)
- [Nucleotide Sequences](#)
- [Protein Sequences](#)
- [Macromolecular Structures](#)
- [Small Molecules](#)
- [Gene Expression](#)
- [Molecular Interactions](#)
- [Reactions & Pathways](#)
- [Protein Families](#)
- [Enzymes](#)
- [Literature](#)
- [Taxonomy](#)
- [Ontologies](#)
- [Patent Resources](#)
- [Sequence Similarity & Analysis](#)
- [Pattern & Motif Searches](#)
- [Structure Analysis](#)
- [Text Mining](#)
- [Downloads](#)
- [Web Services](#)

I. Herramientas de búsqueda



Búsqueda de texto:

TAIR: <http://www.arabidopsis.org/>

The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a [database of genetic and molecular biology data](#) for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every two weeks from the latest published research literature and community data submissions. Gene structures are updated 1-2 times per year using computational and manual methods as well as community submissions of new and updated genes. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The [Arabidopsis Biological Resource Center](#) at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of *Arabidopsis thaliana* and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

TAIR is located at the [Carnegie Institution for Science Department of Plant Biology](#) and funded by the National Science Foundation.

Updates on TAIR funding are available [here](#).

Breaking News

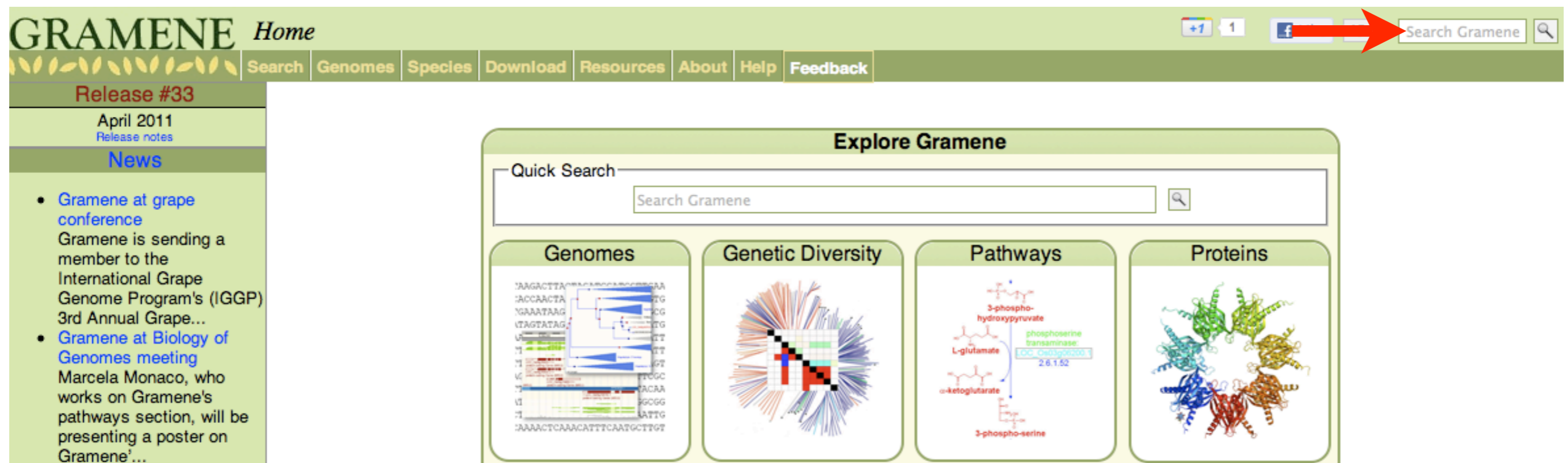
- [Subscribe to news feed](#)
- [Follow our Twitter feed](#)
- [Join our Facebook group](#)

GBrowse now available for eight plant species at TAIR [May 19, 2011]

GBrowse instances for the following plants have been added to TAIR: *Arabidopsis lyrata*, *Brachypodium distachyon*, *Oryza sativa japonica*, *Oryza sativa indica*, *Populus trichocarpa*, *Physcomitrella patens*, *Sorghum bicolor*, *Vitis vinifera*, *Zea mays*.

Búsqueda de texto:

GRAMENE: <http://www.gramene.org/>



The screenshot shows the GRAMENE website home page. At the top left, the logo "GRAMENE Home" is displayed. To the right of the logo is a navigation menu with links for "Search", "Genomes", "Species", "Download", "Resources", "About", "Help", and "Feedback". Further right, there are social media icons for Google+, Facebook, and Twitter, followed by a search bar labeled "Search Gramene" with a magnifying glass icon. A red arrow points to the search bar.

Below the navigation menu, there is a "Release #33" section dated "April 2011" with a link to "Release notes". Below that is a "News" section with two bullet points:

- [Gramene at grape conference](#)
Gramene is sending a member to the International Grape Genome Program's (IGGP) 3rd Annual Grape...
- [Gramene at Biology of Genomes meeting](#)
Marcela Monaco, who works on Gramene's pathways section, will be presenting a poster on Gramene's...

The main content area is titled "Explore Gramene" and features a "Quick Search" bar with the text "Search Gramene" and a magnifying glass icon. Below the search bar are four panels:

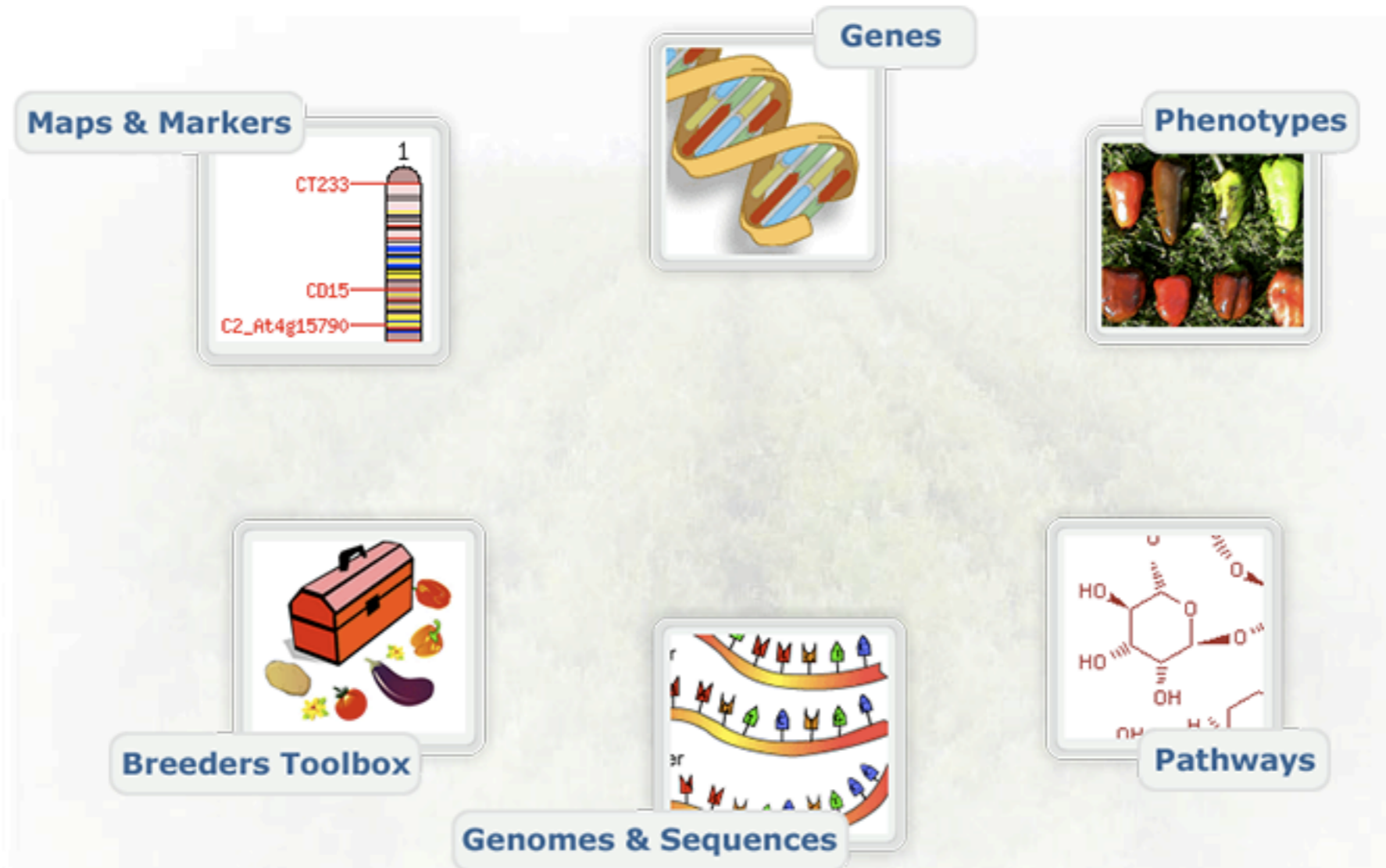
- Genomes:** A panel showing a sequence alignment or comparison of DNA sequences.
- Genetic Diversity:** A panel showing a circular phylogenetic tree or genetic diversity plot.
- Pathways:** A panel showing a metabolic pathway diagram. The central enzyme is "phosphoserine transaminase" (EC: 2.6.1.52). The pathway involves the conversion of L-glutamate and alpha-ketoglutarate to 3-phospho-hydroxypyruvate and 3-phospho-serine.
- Proteins:** A panel showing a 3D ribbon diagram of a protein structure.

I. Herramientas de búsqueda



Búsqueda de texto:

SGN: <http://solgenomics.net/>





Búsqueda por homología de secuencias:

Búsqueda basada en la comparación mediante alineamiento por pares de una o varias secuencias con un conjunto de secuencias usando diferentes algoritmos (blast utiliza una aproximación del algoritmo Smith-Waterman). Aquellas secuencias con cierto nivel de similitud son seleccionadas para la salida del programa.

Los programas más conocidos de búsqueda por homología de secuencias son:

- Blast: (más rápido) <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Fasta (más sensible): <http://www.ebi.ac.uk/Tools/sss/fasta/>

Más información: http://en.wikipedia.org/wiki/Sequence_alignment_software



Búsqueda por homología de secuencias:

NCBI: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST website. At the top, there is a navigation bar with the BLAST logo and the text "Basic Local Alignment Search Tool". Below this, there are tabs for "Home", "Recent Results", "Saved Strategies", and "Help". The main content area starts with a heading "NCBI/BLAST Home" and a description: "BLAST finds regions of similarity between biological sequences. [more...](#)". There is a promotional box for "New Aligned Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool." with a "Go" button. Below this, there is a section titled "BLAST Assembled RefSeq Genomes" with the instruction "Choose a species genome to search, or [list all genomic BLAST databases.](#)". A list of species is provided with checkboxes: Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, Microbes, and Apis mellifera. The next section is "Basic BLAST" with the instruction "Choose a BLAST program to run.". A table lists various BLAST programs with their descriptions and algorithms:

Program	Description	Algorithms
nucleotide blast	Search a nucleotide database using a nucleotide query	blastn, megablast, discontinuous megablast
protein blast	Search protein database using a protein query	blastp, psi-blast, phi-blast
blastx	Search protein database using a translated nucleotide query	
tblastn	Search translated nucleotide database using a protein query	
tblastx	Search translated nucleotide database using a translated nucleotide query	



Búsqueda por homología de secuencias:

EBI: <http://www.ebi.ac.uk/Tools/sss/>

The screenshot shows the EBI Sequence Similarity Searching (SSS) web interface. At the top, there is a search bar with the text "Enter Text Here" and a "Find" button. Below the search bar is a navigation menu with tabs for "Databases", "Tools", "Research", "Training", "Industry", "About Us", and "Help". The "Tools" tab is selected. On the left side, there is a sidebar menu with links for "Tools Home", "Tools A-Z", "Web Services", and "Download". The main content area is titled "Sequence Similarity Searching" and is divided into two sections: "BLAST" and "FASTA".

BLAST

- NCBI BLAST** ⓘ NCBI BLAST Sequence Similarity Search using the NCBI BLAST (blastall) program. This tool is available for the following databases:
 - Protein
 - Nucleotide
 - Vectors
- WU-BLAST** ⓘ Sequence Similarity Search using the Washington University (WU) BLAST2 program (BLAST 2.0 with gaps). This tool is available for the following databases:
 - Protein
 - Nucleotide
 - Parasites
- PSI-BLAST** ⓘ Position Specific Iterative **BLAST (PSI-BLAST)** refers to a feature of BLAST 2.0 in which a profile is automatically constructed from the first set of BLAST alignments.
 - Launch PSI-BLAST

FASTA

- FASTA** ⓘ Sequence Similarity Search using the FASTA program. This tool is available for the following databases:
 - Protein
 - Nucleotide
 - Proteomes
 - Genomes
 - Whole Genome Shotgun
 - ASD Protein
 - ASD Nucleotide
 - LGIC Protein
 - LGIC Nucleotide
- SSEARCH** ⓘ Sequence Similarity Search using the SSEARCH program. This tool is available for the following databases:
 - Protein
 - Nucleotide
 - Proteomes
 - Genomes
 - Whole Genome Shotgun
 - ASD Protein
 - ASD Nucleotide
 - LGIC Protein
 - LGIC Nucleotide
- PSI-Search** ⓘ PSI-Search combines the sensitivity of the Smith-Waterman search algorithm (SSEARCH) with the PSI-BLAST (blastpgp) iterative profile construction strategy to find distantly related protein sequences.
 - Launch PSI-Search



Búsqueda por homología de secuencias:

TAIR: <http://www.arabidopsis.org/Blast/index.jsp>
<http://www.arabidopsis.org/cgi-bin/fasta/nph-TAIRfasta.pl>

The screenshot shows the TAIR BLAST 2.2.8 search interface. It includes a navigation menu with options like Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area is titled "TAIR BLAST 2.2.8" and includes a note: "Please note that this form uses NCBI BLAST2.2.8, and NOT WU-BLAST2.0." The "Blast" section contains a "BLAST™ program" dropdown set to "BLASTN: NT query, NT db", a "Datasets" dropdown set to "TAIR10 Transcripts (-introns, +UTRs) (DNA)", and an "Input" section with radio buttons for "query sequence" (selected) and "locus name (At1g01030)". There is also an "Upload a file" section with a "Choose File" button and "No file chosen" text. A checkbox for "Filter query" is checked. At the bottom, there is an "Advanced BLAST™ Parameter Options" section with a plus sign.

The screenshot shows the TAIR FASTA search interface. It includes a navigation menu with options like Search, Browse, Tools, Portals, Download, Submit, News, and ABRC Stocks. The main content area is titled "FASTA" and includes a "Name of query: (optional)" text box, an "Enter a query sequence: (format: raw or fasta)" text area, and an "OR Upload a file containing query sequence: (format: raw or fasta)" section with a "Choose File" button and "No file chosen" text. There is a "Datasets: [Description]" dropdown set to "TAIR10 Transcripts (-introns, +UTRs) (DNA)" and "Submit" and "Reset" buttons. Below this, a note states: "The type of query sequence and target dataset determine the program used:". A table follows:

Query	Dataset	Program
DNA	DNA	fasta3, searches both strands
Protein	Protein	fasta3
Protein	DNA	tfasx3, searches all six frames
DNA	Protein	fastx3, forward 3 frames, see options for reverse

Options



Búsqueda por homología de secuencias:

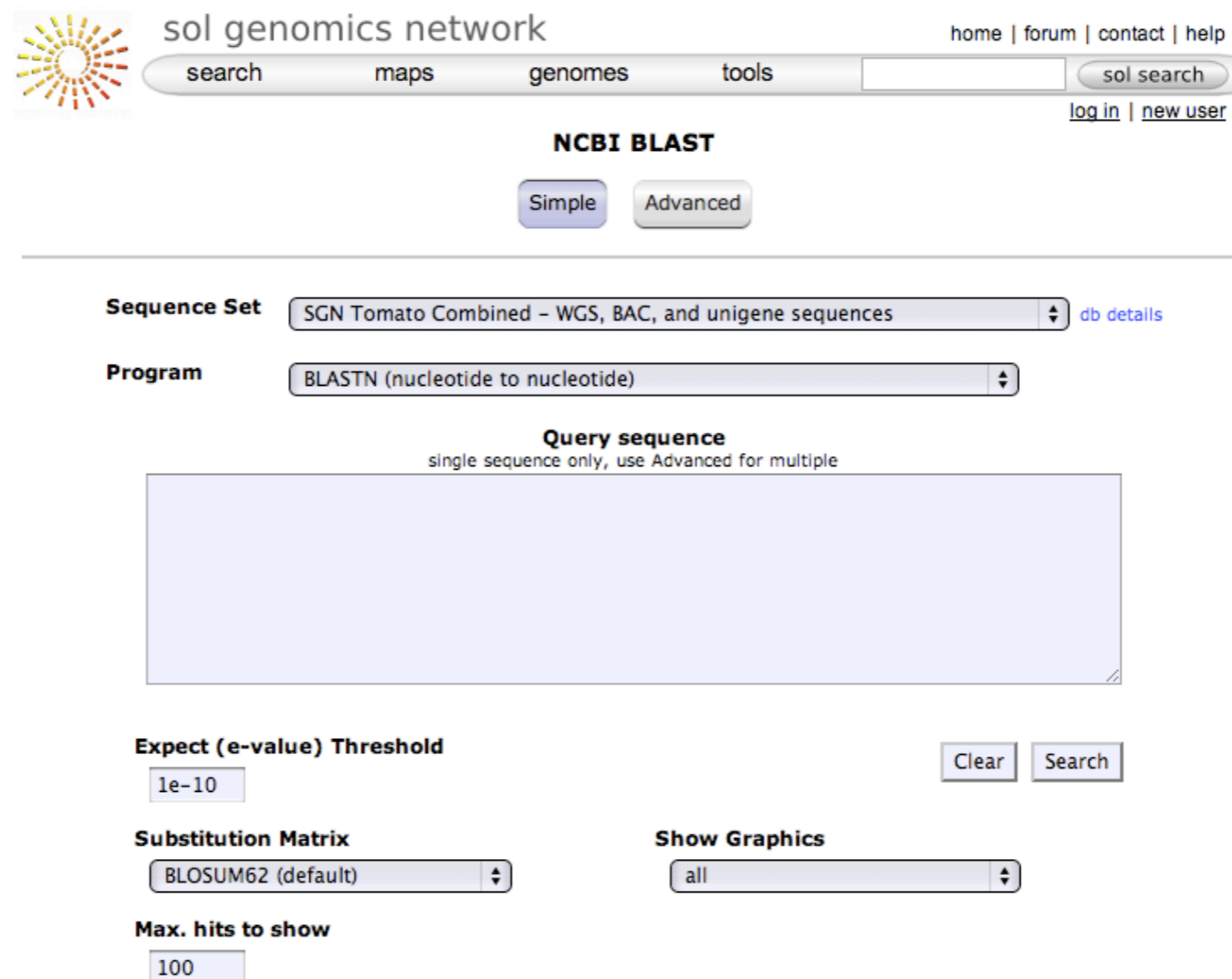
GRAMENE: <http://www.gramene.org/multi/blastview>

The screenshot shows the GRAMENE BLAST search interface. At the top, there is a navigation bar with the GRAMENE logo and links for BLAST, BioMart, Documentation, Help, and Feedback. Below this is a toolbar with buttons for 'new', 'SETUP', 'CONFIG', 'RESULTS', and 'DISPLAY'. On the right side, there are buttons for 'refresh' and 'Online Help'. The main content area is divided into several sections: 1. 'Important Notice' with a message about using Blat as the default DNA search. 2. 'Enter the Query Sequence' section with options to paste sequences, upload a file, or enter a sequence ID or accession. 3. 'Select the databases to search against' section with a dropdown menu for species selection. The interface is clean and functional, designed for easy navigation and data entry.



Búsqueda por homología de secuencias:

SGN: <http://solgenomics.net/tools/blast/index.pl>



The screenshot shows the SGN NCBI BLAST search interface. At the top, there is a navigation bar with the SGN logo, the text "sol genomics network", and links for "home", "forum", "contact", and "help". Below this is a search bar with "search", "maps", "genomes", and "tools" tabs, and a "sol search" button. There are also links for "log in" and "new user".

The main section is titled "NCBI BLAST" and has two buttons: "Simple" and "Advanced".

Below this, there are two dropdown menus: "Sequence Set" (set to "SGN Tomato Combined - WGS, BAC, and unigene sequences") and "Program" (set to "BLASTN (nucleotide to nucleotide)").

The "Query sequence" section has a text area for input, with a note: "single sequence only, use Advanced for multiple".

At the bottom, there are several settings: "Expect (e-value) Threshold" (set to "1e-10"), "Substitution Matrix" (set to "BLOSUM62 (default)"), "Max. hits to show" (set to "100"), and "Show Graphics" (set to "all"). There are "Clear" and "Search" buttons.

Búsqueda por localización (coordenadas):

Basada en la localización de una región cromosómica o de otro tipo de secuencia utilizando una coordenada de comienzo y una coordenada de finalización (*start-end coordinates pair*).

Este tipo de búsqueda utilizan como herramienta los Visualizadores Genómicos (*Genome Browser*)

Más información: http://en.wikipedia.org/wiki/Genome_browser



Búsqueda por valores numéricos asociados a una señal:

Es un tipo de búsqueda asociada a valores cuantitativos, generalmente valores de expresión donde se recogen aquellos resultados que poseen un valor de señal por encima de un valor de corte. Si se comparan experimentos este valor puede ser la ratio entre ambos experimentos.

TAIR: http://www.arabidopsis.org/servlets/Search?action=new_search&type=expression



Búsqueda por valores numéricos asociados a una señal:

Select Array Type/Design

Array Type: Affymetrix GeneChips®

Array Design: any

Limit Search by Expression Values

Data from Arrays with Replicates Data from All Arrays

Affymetrix GeneChip Parameters

Detection: present (P)

Signal: between 0 and 50000

Signal Percentile: between 0 and 100

cDNA Microarray Parameters

Absolute: n/a

Relative: n/a

Fold Change: n/a

Std Error: n/a



Herramientas de las Bases de Datos Biológicas:

1- Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Ontología: exhaustivo y riguroso esquema conceptual dentro de uno o varios dominios dados; con la finalidad de facilitar la comunicación y el intercambio de información entre diferentes sistemas y entidades (Wikipedia)

Ontologías en bioinformática:

Vocabulario controlado de términos para describir funciones de genes, localizaciones, partes de una planta que permite el fácil intercambio de información.

Ejemplo: *Size increase, Mass increase...*

Growth, The increase in size or mass of an entire organism, a part of an organism or a cell. (GO:0040007)





Tipo de Ontologías: Existen muchos tipos de Ontologías en bioinformática, aunque la más usada es *Gene Ontology* (<http://www.geneontology.org/>), para definir términos asociados a procesos biológicas, componentes celulares y y funciones moleculares.

En el caso de plantas, es muy usado *Plant Ontology* (<http://www.plantontology.org/>), para definir partes de la planta y procesos de la planta como maduración o floración.

Para un listado completo de ontologías bioinformáticas:

<http://www.obofoundry.org/>





Búsqueda por Ontologías: Algunas bases de datos poseen búsqueda por ontologías.

Ejemplo: SGN.

The screenshot shows the 'sol genomics network' search interface. At the top, there are navigation links: 'home | forum | contact | help'. Below that, a search bar with 'sol search' and 'log in | new user' links. A 'Search SGN' section contains buttons for 'Genes', 'QTLs & Phenotypes', 'Unigenes', 'Unigene Families', 'Markers', 'Genomic Clones', 'ESTs', 'Images', 'People', and 'Expression'. The 'Gene search' section includes a search type dropdown set to 'contains' and a search input field. The 'Advanced search options' section is expanded, showing fields for 'Organism', 'Chromosome / Linkage Group', 'Locus editor', 'Phenotype', 'Ontology term', and 'Genbank ID'. A red arrow points to the 'Ontology term' field, which has a placeholder text: '(Term name or ID: e.g. 'carotenoid' or 'PO:0007010')'. Below these fields are checkboxes for 'Show only genes with' sequence, markers, and GO/PO annotation. A 'Search' button is at the bottom.



Búsqueda por Ontologías: Una interfaz común para la búsqueda por ontologías, a parte de la casilla de búsqueda es el uso de un Navegador de Ontologías (*Ontology Browser*)

Ejemplo: SGN.

sol genomics network home | forum | contact | help

search maps genomes tools sol search

log in | new user

Browse Ontologies

Ontology browser

Find exact ID Find clear highlight | reset view

Search for text GO (gene ontology) Search

- GO:0003674 molecular_function
- GO:0005575 cellular_component
- GO:0008150 biological_process
 - is_a GO:0008283 cell proliferation
 - is_a GO:0007587 sugar utilization
 - is_a GO:0019740 nitrogen utilization
 - is_a GO:0009758 carbohydrate utilization
 - is_a GO:0006791 sulfur utilization
 - is_a GO:0015976 carbon utilization
 - is_a GO:0006794 phosphorus utilization
 - is_a GO:0071840 cellular component organization or biogenesis
 - is_a GO:0023052 signaling
 - is_a GO:0000003 reproduction
 - is_a GO:0009987 cellular process
 - is_a GO:0016032 viral reproduction
 - is_a GO:0040007 growth
 - is_a GO:0007117 budding cell bud growth
 - is_a GO:0055017 cardiac muscle tissue growth
 - is_a GO:0016049 cell growth



Búsqueda por Ontologías:

GO:0040007 'growth'

Cvterm details

Term id **0040007**
Term name **growth**
Term definition **The increase in size or mass of an entire organism, a part of an organism or a cell.**
Comment **See also the biological process term 'cell growth ; GO:0016049'.**

Synonyms

growth pattern RELATED
"growth pattern" RELATED []
"non-developmental growth" RELATED [GOC:mah]

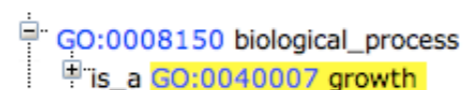
Definition dbxrefs

GOC:bf
GOC:ma

Definition dbxrefs

GO:0048590

Ontology browser



Annotated loci

102 loci

Organism	Symbol	Name
Tomato	gluii	alpha glucosidase II
Tomato	tub	beta-tubulin
Tomato	cry1b	cryptochrome 1b
Tomato	dwf1	DWARF1/DIMINUTO
Tomato	cel3	endo-1,4-beta-glucanase (Cel3)
Tomato	cel2	endo-1,4-beta-glucanase precursor (Cel2)
Tomato	erabp1	ER auxin binding 1



Herramientas de las Bases de Datos Biológicas:

1- Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.





Búsqueda por homología de secuencias:

Búsqueda basada en la comparación mediante alineamiento por pares de una o varias secuencias con un conjunto de secuencias usando diferentes algoritmos (blast utiliza una aproximación del algoritmo Smith-Waterman). Aquellas secuencias con cierto nivel de similitud son seleccionadas para la salida del programa.

Los programas más conocidos de búsqueda por homología de secuencias son:

- Blast: (más rápido) <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Fasta (más sensible): <http://www.ebi.ac.uk/Tools/sss/fasta/>

Más información: http://en.wikipedia.org/wiki/Sequence_alignment_software



Blast:

Es una herramienta para encontrar regiones de similitud local entre secuencias. Compara secuencias de nucleótidos o proteínas y calcula la significancia estadística del resultado.

Programas de Blast:

Existen distintos tipos de programas según la naturaleza de las secuencias a comparar.

		INPUT		
		Nucleotide	Translated Nucleotide	Protein
DATABASE	Nucleotide	BlastN	-	-
	Translated Nucleotide	-	TBlastX	TBlastN
	Protein	-	BlastX	BlastP





Usos de Blast:

- ▶ Búsqueda de genes homólogos:

BlastX (input=cDNA, database=proteins).

BlastP (input=protein, database=proteins).

TBlastN (input=cDNA, database=proteins)

- ▶ Alineamiento para búsqueda de intrones-exones:

BlastN (input=cDNA, database=genomic DNA).

(mejor usar Blat o GeneWise)

- ▶ Búsqueda de polimorfismos.

BlastN (input=cDNA,gDNA, database=cDNA,gDNA).



Términos usados en la salida de Blast:

Query: Secuencia usada en la búsqueda (*input*).

Subject: Secuencia de la base de datos que muestra homología con la secuencia usada en la búsqueda (*database*).

Query Coverage: Porcentaje de la secuencia usada en la búsqueda cubierta por la secuencia de la base de datos que muestra homología.

E-value (expect value): Número de resultados (*hits*) que cabría esperar al azar para la base de datos utilizada, dependiendo del tamaño de la misma. Decrece exponencialmente con el *score* de dos secuencias.

% identity: Porcentaje de identidad compartidos entre ambas secuencias.



Evaluación de los parámetros estadísticos de un alineamiento usando Blast:

%identity vs E-value vs bit score.

E-value es un parámetro estadístico que estima la probabilidad de que la secuencia utilizada haya obtenido un resultado por azar en la base de datos. No tiene significado biológico.

Score, es un valor de puntuación del alineamiento basado en parámetros estadísticos (<http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-I.html>). Está normalizado, luego puede usarse para comparar resultados entre distintas bases de datos, pero sigue sin tener un significado biológico.

% identity, es el porcentaje de nucleótidos idénticos entre ambas secuencias. Dos secuencias se encuentran más cercanas evolutivamente entre sí mientras más parecidas sean a nivel de nucleótidos.



Herramientas de las Bases de Datos Biológicas:

1- Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Búsqueda por localización (coordenadas):

Basada en la localización de una región cromosómica o de otro tipo de secuencia utilizando una coordenada de comienzo y una coordenada de finalización (*start-end coordinates pair*).

Este tipo de búsqueda utilizan como herramienta los Visualizadores Genómicos (*Genome Browser*)

Más información: http://en.wikipedia.org/wiki/Genome_browser



Los Navegadores Genómicos más conocidos son:

- GBrowse (GMOD).
- UCSC Genome Browser.
- Ensembl Genome Browser.
- Vista Genome Browser.
- JBrowse (GMOD).





Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Existen decenas de herramientas de manipulación de secuencias. Muchas de ellas son paquetes comerciales como *LaserGene* de *DNAStar* (<http://www.dnastar.com/t-products-lasergene.aspx>). Otros son paquetes gratuitos como *BioEdit* (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) para Windows o *eBioTools* (<http://www.ebioinformatics.org/>) para MacOX.

Muchas bases de datos poseen programas con funciones similares, a veces integrados con la base de datos y otras veces como interfaces gráficas independientes.





Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



2.1 - Traductores y *Gene Predictors*.

Existen dos tipos de herramientas que pueden utilizarse para encontrar el ORF correcto, y por lo tanto la secuencia de una proteína que corresponde a un gen.

- Alineamiento con una secuencia de mRNA o cDNA, posterior traducción y selección del ORF más largo.
- Predicción del gen basado en su estructura.

Basadas en estas aproximaciones existen dos tipos de herramientas que pueden aplicarse comunmente a un conjunto de secuencias.

- ▶ Traductores (DNA a Proteínas sin tener en cuenta exones, basados en los posibles ORFs).
- ▶ *Gene Predictors* (DNA a CDS teniendo en cuenta la estructura de exones). Generalmente requieren del entrenamiento del programa basado en secuencias y distribución de exones conocidas.



Existen varios programas de traducción de secuencias via web:

- Translate Tool (ExPASy): <http://expasy.org/tools/dna.html>
- ORF Finder (NCBI): <http://www.ncbi.nlm.nih.gov/projects/gorf/>
- Transeq (EBI): <http://www.ebi.ac.uk/Tools/emboss/transeq/>
- RevTrans 1.4 Server (CBS): <http://www.cbs.dtu.dk/services/RevTrans/>
- Transeq (UMass): <http://biotools.umassmed.edu/cgi-bin/biobin/transeq>
- Dnatoprotein (JHI): <http://www.dnatoprotein.com/>
- EstScan (embnet): <http://www.ch.embnet.org/software/ESTScan2.html>

2.1 - Traductores y Gene Predictors.



- **Transeq (EBI):** <http://www.ebi.ac.uk/Tools/emboss/transeq/>

EMBL-EBI [Help](#) | [Feedback](#)

[Databases](#) [Tools](#) [Research](#) [Training](#) [Industry](#) [About Us](#) [Help](#) [Site Index](#)

EBI > Tools > Sequence Analysis > EMBOSStanseq

EMBOSS Transeq

Transeq translates nucleic acid sequences to the corresponding peptide sequence. It can translate in any of the 3 forward or three reverse sense frames, or in all three forward or reverse frames, or in all six frames.

Frame: Table:

Trim: Reverse: Colour:

Enter or upload a nucleic acid Sequence in any format:

Upload a file: No file chosen



- EstScan (embnet): <http://www.ch.embnet.org/software/ESTScan2.html>

+ ch.EMBnet.org

Home Services Courses Links Contacts

ESTScan 2

(old version [ESTScan 1](#))

Usage: Paste a DNA sequence, ID or AC (ex. X91349) or a Unigene cluster ID (ex. Hs.9194), select the format, the species and the options and click the "Run ESTScan" button.

- Please cite:
- Lottaz C, Iseli C, Jongeneel CV, Bucher P. (2003)
Modeling sequencing errors by combining Hidden Markov models
Bioinformatics **19**, ii103-ii112.
- Iseli C, Jongeneel CV, Bucher P. (1999)
ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.
*Proc Int Conf Intell Syst Mol Biol.*138-48.

ESTScan is available on SourceForge: <http://estscan.sourceforge.net>
See ESTScan [help page](#) for explanations or subscribe to our [mailing list](#).
Make sure that the format button shows the correct format.

Select format:	Plain Text
Select species:	Human (or any mammalian)
	<input type="text"/> Query title (option)
Paste your sequence here: (or ID or accession number)	<input type="text"/>
Options:	
Insertion/Deletion penalty	<input type="text"/> (default: -50)
Output	<input checked="" type="radio"/> DNA (CDS) <input type="radio"/> full DNA <input type="radio"/> Protein



Existen varios programas de predicción de genes via web:

- **FGENESH (ULondon):**

<http://mendel.cs.rhul.ac.uk/mendel.php?topic=fgen-file>

- **GENESCAN (MIT):**

<http://genes.mit.edu/GENSCAN.html>

- **GeneMark.hmm (GaTech):**

<http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi>

- **Augustus:**

<http://augustus.gobics.de/submission>



Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.

2.2 - Alineamiento de múltiples secuencias



Al contrario que programas como Blast, Blat o Fasta, programas como ClustalW o Muscle permiten el alineamiento de más de dos secuencias (nucleótidos o proteínas) al mismo tiempo.

Alguno de ellos, como ClustalW, permiten la representación de los alineamientos como un árbol, generalmente mediante algoritmos de cálculo simple como *Neighbor-Joining*.

- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>
- Kalign (EBI): <http://www.ebi.ac.uk/Tools/msa/kalign>
- MAFFT (EBI): <http://www.ebi.ac.uk/Tools/msa/mafft>
- MUSCLE (EBI): <http://www.ebi.ac.uk/Tools/msa/muscle>
- T-Coffee (EBI): <http://www.ebi.ac.uk/Tools/msa/tcoffee>

2.2 - Alineamiento de múltiples secuencias



- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

```
sequence.fasta
>gil15225089|refINP_180705.1| F-box protein DOR [Arabidopsis thaliana]
MKSRRQNVSVARQITLGRDENFEPIPIDLVIEIFSRSPVKSIARCRCVSKLWASILRLPYFTELYLTKSC
ARPRLLFACQKHRELFSTPQPHNPNESSSPLAASFHMKIPFDGRFNIIISPIGGLVFRYEQILKGRKT
PEFVSAICNPSTGQSLTLPKPKTRKRIWGTSHFGYDPIEKQFKVLSMNIGDGVYKEHYVLTGTEENLSWR
RIECSIPHVHGSKGICINGVLYYRAKADMFSGTLMIVCDFVRFKFSYIKILKPTTTLISYNGKLASLWV
EGPSYICGKRFEMWVLDPEKHEWLKHTYELRPRWQNVLDGEDLLIFAGMTGTNEIVLSPKYPSHPFYVY
YNLERNTIRRVEIQGMGAFKVNEDIYFLDHVEDVKLI

>gil297837229|refIXP_002886496.1| hypothetical protein ARALYDRAFT_893290 [Arabidopsis lyrata subsp. lyrata]
MKTERQNVSEDEVVVVTERNKRAKTSNNGGEPDFDLTVEICSRPAKSISRFRVCLKLGWSILRLPYFTE
LFLTRSLARPQLLFACHKDNHVFVSSPQPNIDNNASSLLAANYHMKIPFYASSFERCSSVRGLVFFG
DERYSNGKEHKVSVICNPSTRQSLTLPKPKTRKRIWGTSHFGYDPIEKQFKVLSMTWGIYTRDMDSEEH
QVLTGTRKPSWRMIECWIPHSLYHTYNNVCINGVLYYPAVNTSSKGFIIVSFDFRSEEFRFVEDTDTSI
SSYYGPHLINYNGKLGSLGSGGFGGIGASCTSLTRVLEDAEKHEWSEHIYVLPAAWKNIFGGECTVLSV
VGVTRTNEIVLSLRFSTPFYVYFYNTERNAIRRVEIQGQEAQKDHVSVYTFLDHVENVMKLLLEGF

>gil15239182|refINP_201386.1| F-box protein [Arabidopsis thaliana]
MRTLRRNVTENRLTISRRTKTKTSNPKTEKSVQIPVDIIIEILLRLPAKSIATCRVSKLWISVICRQD
FTELFLTRSLHRPQLLFCKKDGNLFFFSSPQLQNPYENSSAISLKNFSLCYKISRPNGLICFKRKEMN
ETVTVICNPSTGHTLSLPMKTSIGPSRFFVYEQKQFKVLLSYKSDHQVLTGTEGELSWRIIECSM
PHILGMSEICINGVLYYPAINLSSGDYIIVCFDVRSEKFRFITVMEEFKAAHDGTLINYNGKLASLVSE
RYCFVDGRSKEIWLWVLDQAEKKEWSKHTYVLPAAWQHRIGTLNLRVGVTRTNEIMLSPCYQTVPFVY
YFNIERKTMMVAIQGMEAFQGHVFTYLDHVENVKLLHNMF

>gil15229553|refINP_189038.1| putative F-box protein [Arabidopsis thaliana]
MRSRQLHNVEDRETLRNRKRSKTSLNHGIPIIDLLIEIFLKLVPKSIATCRVSKFWTYVLRQDFTEL
FLTKSSSRPQLLFACANDNGYFFFSSNPQNLDENSSPIAAYPLTHVPKSRDLGPPINGLVSLRGERILK
GRIRPVDVSIINPSTGESLTLPKTNMTRKKIYTVTSFLGYDPIEKQFKVLSMNMSYEKHPKCEGYQVLT
LGTGKLSWRMIKCLNYQHPLKNSEICINGVLYLAMVNGSSWPTRAVVCFDIRSEMFMFMEVYRELSYT
TTLINYNNGKLGMLMGQEAHKTISGICRSFELWVLEDTVKHEWSKHVYLLPPLWKDAVANTRLYFAGMIG
TSEIVLFRPDEPLCVFYINIDRNTIKRVGIRGLEAFKYFRIFLNHVENVKLF

>gil297819588|refIXP_002877677.1| hypothetical protein ARALYDRAFT_906230 [Arabidopsis lyrata subsp. lyrata]
MSTMMKKRKRHVSKEDVALTISSSLGEYGENSGTLPMDLMVEILSRVPAKSAAKFHCVSKNWNLSLLRSSY|
FTNLYLTRSPTRPRLLITFQAEKGWSFFSSPEYLISDQNSNLVVVDNHMDVDPKDYSGVCEPVCGLLCTR
DEWVLSRKKDARMMICNPSTRQFQSLPKVRSRRNKVITYIGYDPIEKQFKVLCMTICERPVMFKAEHQV
LTLGTGKLRWMLKCFVEHFPKHKEICINGVLYLAVKDETRDIIVCFHVKEKQFQFILNKAPLSTLIN
YNGKLGIRHGFMEGGVAGYELWDLIEKEDWTRHIHILPPMVKQVVGTRVYVVGMIQTSEIVFSPFVK
SNPFYIFHLNIERNISITRVEIQGTGPLEGQVYTFINHIENVKLIM
```

Input: Set de 5 proteínas



2.2 - Alineamiento de múltiples secuencias



- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

EBI > Tools > Multiple Sequence Alignment > ClustalW2

ClustalW2 - Multiple Sequence Alignment

ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins.

Use this tool

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

```
subsp. lyrata]
MSTMMKKRKRHVSKEDVALTISSSLGEYGENSGTLPMDLMVEILSRVPAKSAAKFHCVSKNWNLSLLRSSY
FTNLYLTRSPTRPRLLITFQAEGKWSFFSSPEYLISDQNSNLVVVDNHMDVDPKDYSGVCEPVCGLLCTR
DEWVLSRKKDARMMICNPSTRQFQSLPKVRSRRNKVITYIGYDPIEKEYKVLCTICERPVMFKAEEHQV
LTLGTGKLRMLKCFVEHFPKHKEICINGVLYYLAVKDETREDIIVCFHVKHEKFQFILNKAPLSTLIN
YNGKLGIRHGFMEGGVAGYELWDLIEKEDWTRHIHILPPMWKQVVGETRVYVVGMIGTSEIVFSPFVK
SNPFYIFHLNIERNISITRVEIQGTGPLEGQQVYTFINHIENVKLIM|
```

Or, upload a file: No file chosen

STEP 2 - Set your Pairwise Alignment Options

Alignment Type: Slow Fast

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 3 - Set your Multiple Sequence Alignment Options

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 4 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

2.2 - Alineamiento de múltiples secuencias



- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

```
EBI > Tools > Multiple Sequence Alignment > ClustalW2
ClustalW2 Results
Alignments Result Summary Guide Tree Submission Details Submit Another Job
Alignment
Download Alignment File Show Colors
CLUSTAL 2.1 multiple sequence alignment

gi|15225089|ref|NP_180705.1|      ---MKSR-RQNVSVARQTLGRDE-----NFEPIPIDLVIEIFSR 36
gi|297837229|ref|XP_002886496.  ---MKTE-RQNVSEdVVVVTERNKRAKTSNN---GGEPIPFDLTVEICSR 43
gi|15239182|ref|NP_201386.1|    ---MRTL-RRNVTENRLTISRRTTEKKTSPNKTEKSVQIPVDIIIEILLR 46
gi|15229553|ref|NP_189038.1|    ---MRSRQLHNVSEdRETLSRRNKRSKTSLN-----GHIPIDLLIEIFLK 42
gi|297819588|ref|XP_002877677.  MSTMMKKRKRHVSKEDVALTISSSLGEYGEN----SGTLPMDLMVEILSR 46
      * .   :::   .:                               :*.*: :**  :

gi|15225089|ref|NP_180705.1|    SPVKSIAARCRCVSKLWASILRLPYFTELYLTKSCARPRLLFACQKHREL 86
gi|297837229|ref|XP_002886496.  LPAKSISRFRCVLKLWGSILRLPYFTEFLTRSLARPQLLFACHKDNHVF 93
gi|15239182|ref|NP_201386.1|    LPAKSIATCRCVSKLWISVICRQDFTEFLTRSLHRPQLLFCKKDGRLF 96
gi|15229553|ref|NP_189038.1|    LPVKSIAATCRSVSKFWTYVLGRQDFTEFLTKSSSRPQLLFACANDNGYF 92
gi|297819588|ref|XP_002877677.  VPAKSAAKFHCVSKNWNLSLRSSYFTNLYLTRSPTRPRLLITFQAEKWS 96
      *.** :  :.* * *  ::   **:*:**:* **:*:**  .

gi|15225089|ref|NP_180705.1|    FFSTPQPHNP--NESSPLAASFHMKIPFDG-RFNIISPIGGLVFRVYEQ 133
gi|297837229|ref|XP_002886496.  VFSSPQPNIDDNNASSLLAANYHMKIPFYASSFERCSSLVGLVFFGDER 143
gi|15239182|ref|NP_201386.1|    FFSSPQLQNPYENSSAISLKN-----FSLCYKISRPNGLICFK--- 135
gi|15229553|ref|NP_189038.1|    FFSSNQPNLDENSSPIAAYP-----LTHVPKSRDLGPPINGLVSLRGER 137
gi|297819588|ref|XP_002877677.  FFSSPEYLISDQNSNLVVVDN---HMDVPKDYSFGVCEPVCGLLCTRDEW 143
      .**: :      * .                               .: **

gi|15225089|ref|NP_180705.1|    ILKGRKTPEFVSAICNPSTGQSLTLPKPK-TRKRIW-GTSHFGYDPIEQ 181
gi|297837229|ref|XP_002886496.  YSNGKE--HKVSVICNPSTRQSLTLPKPK-TRKRIG-VRSYFGFEPIEQ 189
gi|15239182|ref|NP_201386.1|    ---RKEMNETVTVICNPSTGHTLSLPKP--MKTSIG-PSRFFVYEPIQK 179
gi|15229553|ref|NP_189038.1|    ILKGRIRPVDVSIYNPSTGESLTLPKTNMTRKKIYTVTSFLGYDPIEQ 187
gi|297819588|ref|XP_002877677.  VLSRKK--DARMMICNPSTRQFQSLPKVRSRRNKVI---TYIGYDPIEKE 188
      :      * **** .  :***  :. :      .: :****:
```

2.2 - Alineamiento de múltiples secuencias



- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

EBI > Tools > Multiple Sequence Alignment > ClustalW2

ClustalW2 Results

Alignments Result Summary **Guide Tree** Submission Details Submit Another Job

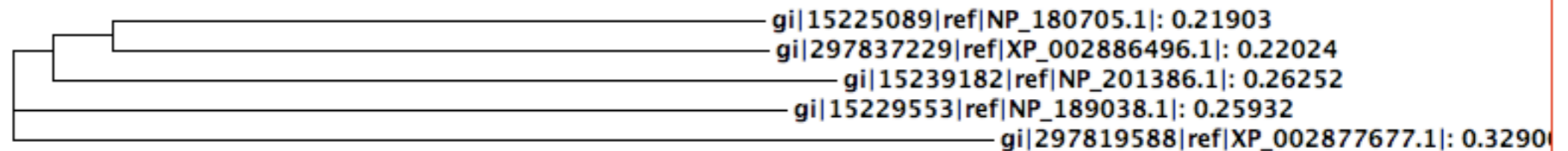
Guide Tree

Download Guide Tree File

```
(
(
(
gi|15225089|ref|NP_180705.1|:0.21903,
gi|297837229|ref|XP_002886496.1|:0.22024)
:0.01981,
gi|15239182|ref|NP_201386.1|:0.26252)
:0.01344,
gi|15229553|ref|NP_189038.1|:0.25932,
gi|297819588|ref|XP_002877677.1|:0.32906);
```

Phylogram

Show as Cladogram Tree Hide Distances



Right-click on the above tree to see display options.

2.2 - Alineamiento de múltiples secuencias



- ClustalW (EBI): <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

EBI > Tools > Multiple Sequence Alignment > ClustalW2

ClustalW2 Results

Alignments | Result Summary | Guide Tree | Submission Details | Submit Another Job

Alignment

Download Alignment File | Show Colors

CLUSTAL 2.1 multiple sequence alignment

Uso del alinenamiento por otros programas de análisis filogenético como Phylip.

phylip 3.67: protpars

Protein Sequence Parsimony Method ?

* Alignment File ? [use example data]

Run Reset

paste upload EDIT CLEAR

Enter your data below:

```
.....
EIQGMGAEKV NEDYJELDHV EDVKLI----
EIQGQAEAKD HSYVTELDHV ENVMKLLLEG F
AIQGMEAFQG HLYVETLDHV ENVKLLHNME -
GIRGLEAEKY FR--IFLNHV ENVKLF-----
EIQGTGPLEG QQVYTEINHI ENVKLIM---|
```

Parcimony options

Use Threshold parsimony (T)

* Threshold parsimony value

Genetic code for 'categories' model (C)

Protein parsimony algorithm, version 3.

One most parsimonious tree found:

```

          +--ArFbox2
         +-----3
        !     +--AtFbox1
    +--2
    ! !     +--ARALY_9062
    1 +-----4
    !     +--ARALY_8932
    !
    +-----AtDOR
```

remember: this is an unrooted tree!

requires a total of 1060.000



Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.

2.3 - Análisis de dominios funcionales



A menudo la anotación de secuencias se basa en la homología de dicha secuencias con secuencias de función conocida. No obstante algunas secuencias de proteínas solo poseen algunas regiones con cierta homología a otras proteínas. Estas regiones son **dominios** y en muchos casos están asociados a funciones.

InterPro (<http://www.ebi.ac.uk/interpro/>) agrupa varias bases de datos de dominios de proteínas como son ProSite, Pfam o Superfamily.



Una herramienta de búsqueda habitual para dominios funcionales es InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

2.3 - Análisis de dominios funcionales



InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

EBI > Tools > Protein Functional Analysis > InterProScan

InterProScan Sequence Search

This form allows you to query your sequence against InterPro. For more detailed information see the documentation for the perl stand-alone InterProScan package ([Readme file](#) or [FAQ's](#)), or the [InterPro user manual](#) or [help pages](#).

Use this tool

STEP 1 - Enter your input sequence

Enter or paste a **PROTEIN** sequence in any supported format:

```
>gi|15225089|ref|NP_180705.1| F-box protein DOR [Arabidopsis thaliana]
MKSRRQNVSVARQTILGRDENFEPIIDLVEIFSRSPVKSIARCRCVSKLWASILRLPYFTELYLTKSC
ARPRLLFACQKHRELFSTPQPHNPNESSPLAASFHMKIPFDGRFNIISPIGGLVFRYEQILKGRKT
PEFVSAICNPSTGQSLTLPKPKTRKRIWGTSHFGYDPIEQFKVLSMNIGDGVYKEHYVLTGTENLSWR
RIECSIPHVHGSKGICINGVLYYRAKADMFSGTLMIVCFDVRFEKFSYIKLKPTTTLISYNGKLASLW
EGPSYICGKRFEMWVLGDPEKHEWLKHTYELRPRWQNVLGEDLLIFAGMTGTNEIVLSPKYSPHPFYFY
YNLERNTIRRVEIQGMGAFKVNEDYIFLDHVEDVKLI
```

Or, upload a file: No file chosen

STEP 2 - Select the applications to run

Select All Clear All

<input checked="" type="checkbox"/> BlastProDom	<input checked="" type="checkbox"/> FPrintScan	<input checked="" type="checkbox"/> HMMPiR	<input checked="" type="checkbox"/> HMMPfam	<input checked="" type="checkbox"/> HMMSmart
<input checked="" type="checkbox"/> HMMTigr	<input checked="" type="checkbox"/> ProfileScan	<input checked="" type="checkbox"/> HAMAP	<input checked="" type="checkbox"/> PatternScan	<input checked="" type="checkbox"/> SuperFamily
<input checked="" type="checkbox"/> SignalPHMM	<input checked="" type="checkbox"/> TMHMM	<input checked="" type="checkbox"/> HMMPanther	<input checked="" type="checkbox"/> Gene3D	

STEP 3 - Submit your job

Be notified by email (*Tick this box if you want to be notified by email when the results are available*)

2.3 - Análisis de dominios funcionales



InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

EBI > Tools > Protein Functional Analysis > InterProScan Sequence Search

InterProScan Results

Summary Table Tool Output **Visual Output** Submission Details Submit Another Job

InterProScan Visual Output

Download in SVG format

InterProScan (version: 4.8)

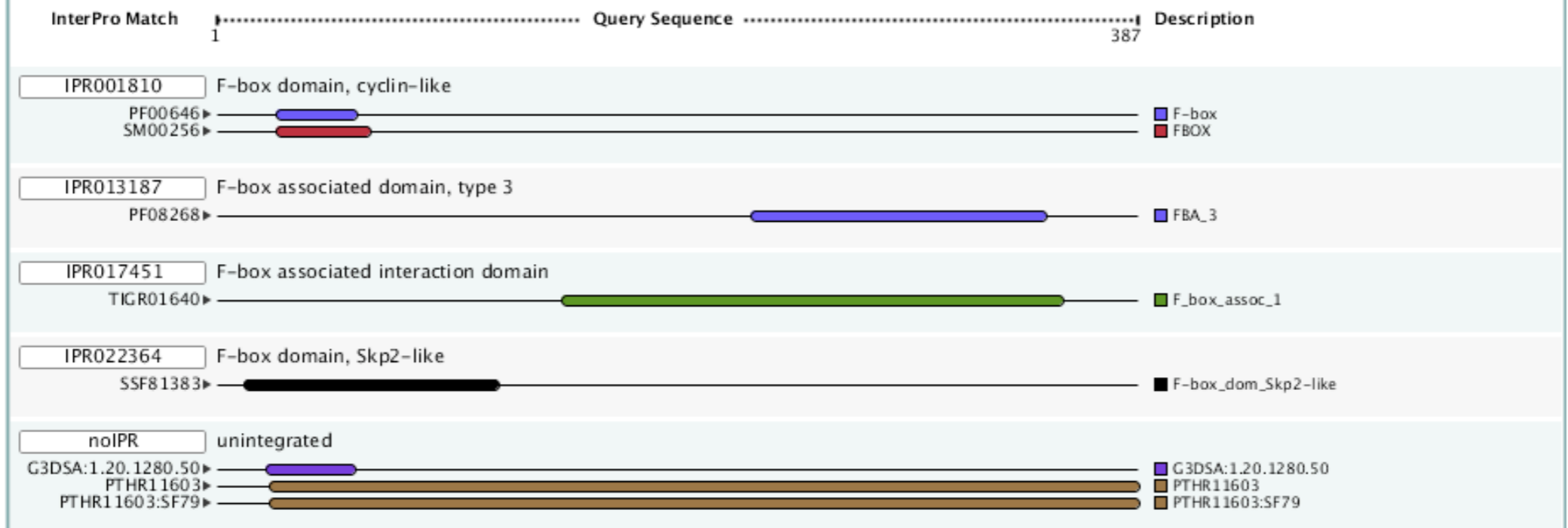
Sequence: NP_180705.1

Length: 387

CRC64: ABA57A2679C00184

Launched Mon, Jun 13, 2011 at 18:04:25

Finished Mon, Jun 13, 2011 at 18:04:53



PRODOM PRINTS PIR PFAM SMART TIGRFAMs PROFILE
HAMAP PROSITE SUPERFAMILY SIGNALP TMHMM PANTHER GENE3D

© European Bioinformatics Institute 2006-2011. EBI is an Outstation of the European Molecular Biology Laboratory.

2.3 - Análisis de dominios funcionales



InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>).

EBI > Tools > Protein Functional Analysis > InterProScan Sequence Search

InterProScan Results

Summary Table | Tool Output | Visual Output | Submission Details | Submit Another Job

InterProScan Summary

IPR001810 F-box domain, cyclin-like

Method	Identifier	Description	Matches
PFAM	PF00646	F-box	0.011 [25-59] T
SMART	SM00256	FBOX	1.1000002989614844E-5 [25-65] T

Parent IPR022364
Children No children
Found in No entries
Contains No entries
GO terms GO:0005515 protein binding

IPR013187 F-box associated domain, type 3

Method	Identifier	Description	Matches
PFAM	PF08268	FBA_3	7.499999999999999E-30 [224-348] T

Parent IPR017451
Children No children
Found in No entries
Contains No entries
GO terms None



Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Los péptidos señal son pequeñas cadenas de aminoácidos (3-60 aminoácidos) que dirigen el transporte de una proteína a lugares como el núcleo, el nucleolo, el peroxisoma o la matriz mitocondrial (wikipedia).

Ejemplos:

Transport to the nucleus (NLS)	-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-
Transport to the endoplasmic reticulum	H ₂ N-Met-Met-Ser-Phe-Val-Ser-Leu- Leu-Leu-Val-Gly-Ile-Leu-Phe- Trp-Ala-Thr-Glu-Ala-Glu-Gln- Leu-Thr-Lys-Cys-Glu-Val-Phe- Gln-
Retention to the endoplasmic reticulum	-Lys-Asp-Glu-Leu-COOH
Transport to the mitochondrial matrix	H ₂ N-Met-Leu-Ser-Leu-Arg-Gln-Ser- Ile-Arg-Phe-Phe-Lys-Pro-Ala- Thr-Arg-Thr-Leu-Cys-Ser-Ser- Arg-Tyr-Leu-Leu-
Transport to the peroxisome (PTS1)	-Ser-Lys-Leu-COOH
Transport to the peroxisome (PTS2)	H ₂ N-----Arg-Leu-X ₅ -His-Leu-

SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) es un programa de predicción de péptidos señales.

2.4 - Análisis de péptidos señal



SignalP (<http://www.cbs.dtu.dk/services/SignalP/>)

CBS >> CBS Prediction Servers >> SignalP

SignalP 3.0 Server

SignalP 3.0 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks and hidden Markov models.

View the [version history](#) of this server. All the previous versions are available on line, for comparison and reference.

New: the paper about using SignalP and other protein subcellular localization prediction methods:

Locating proteins in the cell using TargetP, SignalP, and related tools
Olof Emanuelsson, Søren Brunak, Gunnar von Heijne, Henrik Nielsen
Nature Protocols 2, 953-971 (2007).

is now again available for download - please click [here](#).

Background	Article abstracts	Instructions	Output format
----------------------------	-----------------------------------	------------------------------	-------------------------------

SUBMISSION

Paste a single sequence or several sequences in **FASTA** format into the field below:

Submit a file in **FASTA** format directly from your local disk:

No file chosen

Organism group <input checked="" type="radio"/> Eukaryotes <input type="radio"/> Gram-negative bacteria <input type="radio"/> Gram-positive bacteria	Method <input type="radio"/> Neural networks <input type="radio"/> Hidden Markov models <input checked="" type="radio"/> Both	Graphics <input type="radio"/> No graphics <input checked="" type="radio"/> GIF (inline) <input type="radio"/> GIF (inline) and EPS (as links)
Output format <input checked="" type="radio"/> Standard <input type="radio"/> Full	Truncation Truncate each sequence to max. <input type="text" value="70"/> residues.	

2.4 - Análisis de péptidos señal

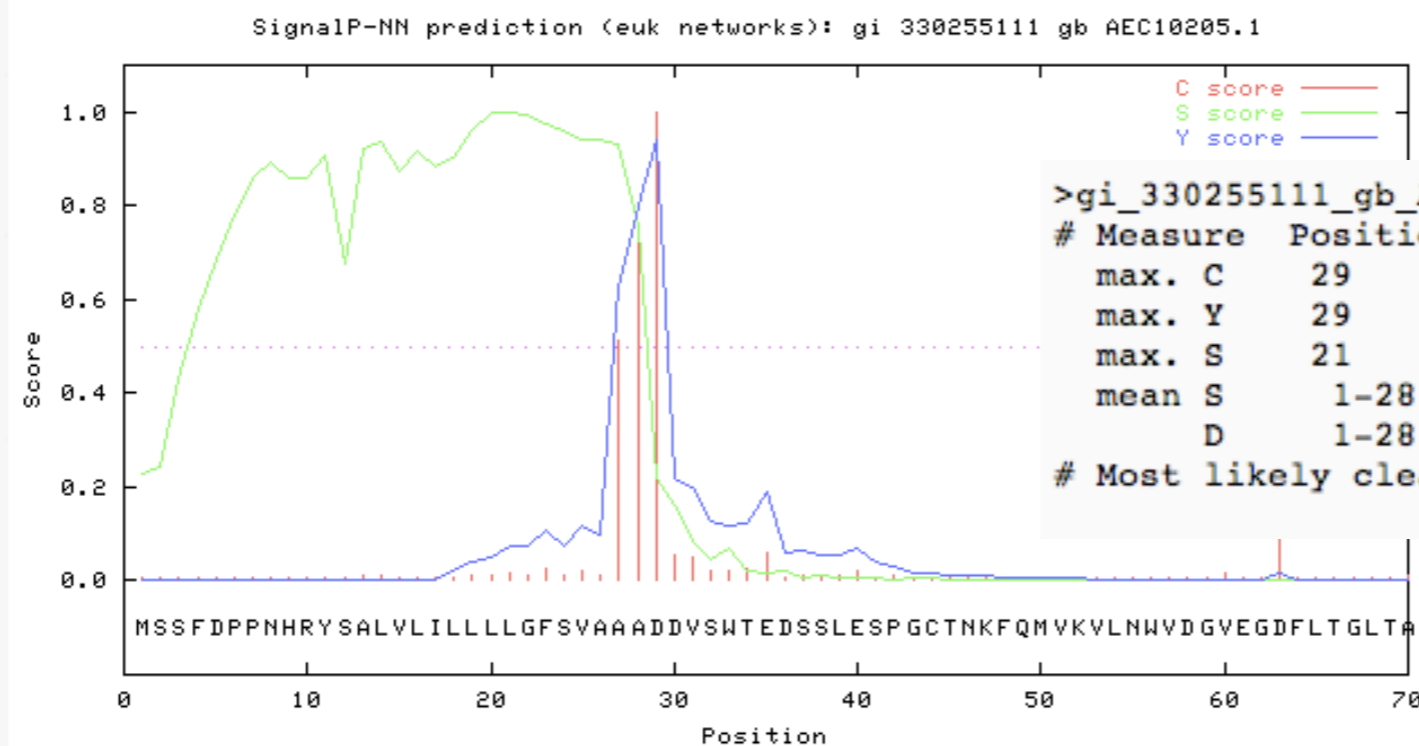


SignalP (<http://www.cbs.dtu.dk/services/SignalP/>)

Using neural networks (NN) and hidden Markov models (HMM) trained on eukaryotes

```
>gi_330255111_gb_AEC10205.1_ signal peptide peptidase-like 3 _Arabidopsis thaliana_
```

SignalP-NN result:



```
>gi_330255111_gb_AEC1 length = 70
# Measure Position Value Cutoff signal peptide?
max. C      29      1.000  0.32  YES
max. Y      29      0.939  0.33  YES
max. S      21      0.998  0.87  YES
mean S      1-28     0.818  0.48  YES
           D      1-28     0.878  0.43  YES
# Most likely cleavage site between pos. 28 and 29: AAA-DD
```



Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



3.1 - Visualizadores de mapas de ligamiento

Un mapa de ligamiento es un mapa genético de una población que muestra la posición conocida de genes o marcadores genéticos relativa en terminos de distancia de recombinación.

Los visualizadores de mapas de ligamiento (*Map Viewer, Chromosome Viewer*), son representaciones gráficas de estos mapas de ligamiento.

Existen distintos visualizadores:

- MapViewer (NCBI): <http://www.ncbi.nlm.nih.gov/projects/mapview/>
- MapViewer (TAIR): <http://www.arabidopsis.org/servlets/mapper>
- CMap (Gramene): <http://www.gramene.org/cmap/>
- CView (SGN): <http://solgenomics.net/cview/>

3.1 - Visualizadores de mapas de ligamiento



- CView (SGN): <http://solgenomics.net/cview/>

Selección del mapa.

Interactive maps

[\[Submit New Map\]](#)

Arabidopsis

[Arabidopsis COSII](#): Arabidopsis thaliana sequenced-based COSII map

Potato

[Potato meta-QTL consensus map](#): Potato meta-QTL consensus map

[Potato-TXB 1992](#): S. tuberosum (2x=24) x S. berthaultii, backcrossed to S. berthaultii, type BC1, 1992

Eggplant

[Eggplant-LXM 2002](#): S. linnaeanum MM195 x S. melongena MM738 type F2, 2002

[Eggplant-COSII](#): S. linnaeanum MM195 x S. melongena MM738

Tobacco

[Tobacco SSR 2007](#): N. tabacum var. Red Russian x N. tabacum var. Hicks Broadleaf

[Tobacco N. acuminata](#): N. acuminata TA3460 Å— N. acuminata var. multiflora TA3461

[Tobacco N. tomentosiformis](#): N. tomentosiformis TA3385 Å— N. otophora TA3353

Pepper

[Pepper-FAO3](#): Capsicum annuum cv. NuMex RNaky x Capsicum frutescens var. BG 2814-6

[Pepper-AC99](#): Capsicum annuum cv. NuMex RNaky x Capsicum chinense var PI159234

[Pepper-COSII](#): Capsicum annuum cv. NuMex RNaky x Capsicum frutescens var. BG 2814-6

[Pepper Capsicum](#): Capsicum annuum A44750157 x Capsicum chinense PI 152225

Tomato

[Kazusa F2-2000 genetic map](#): Kazusa F2-2000 genetic map

[Kazusa F2-2000 pseudomolecule map](#): Kazusa F2-2000 pseudomolecule map

[Tomato-EXPIMP 2009](#): Tomato-EXPIMP 2009: S. lycopersicum (NCEBR-1) x S. pimpinellifolium (LA2093) type RIL-F7, 2009

[Tomato FISH map](#): Solanum lycopersicum (Tomato) Fluorescence In-Situ Hybridization (FISH) Map

[Tomato-EXPIMP 2001](#): S. lycopersicum TA209 x S. pimpinellifolium LA1589 type RI, 2001

[Tomato-EXPEN 1992](#): S. lycopersicum cv. VF36 x S. pennellii LA716 type F2, 1992

[Tomato-EXHIR 1997](#): S. lycopersicum TA209 x S. habrochaites LA1777 type BC1, 1997

[Tomato-EXPEN 2000](#): S. lycopersicum LA925 x S. pennellii LA716 type F2.2000

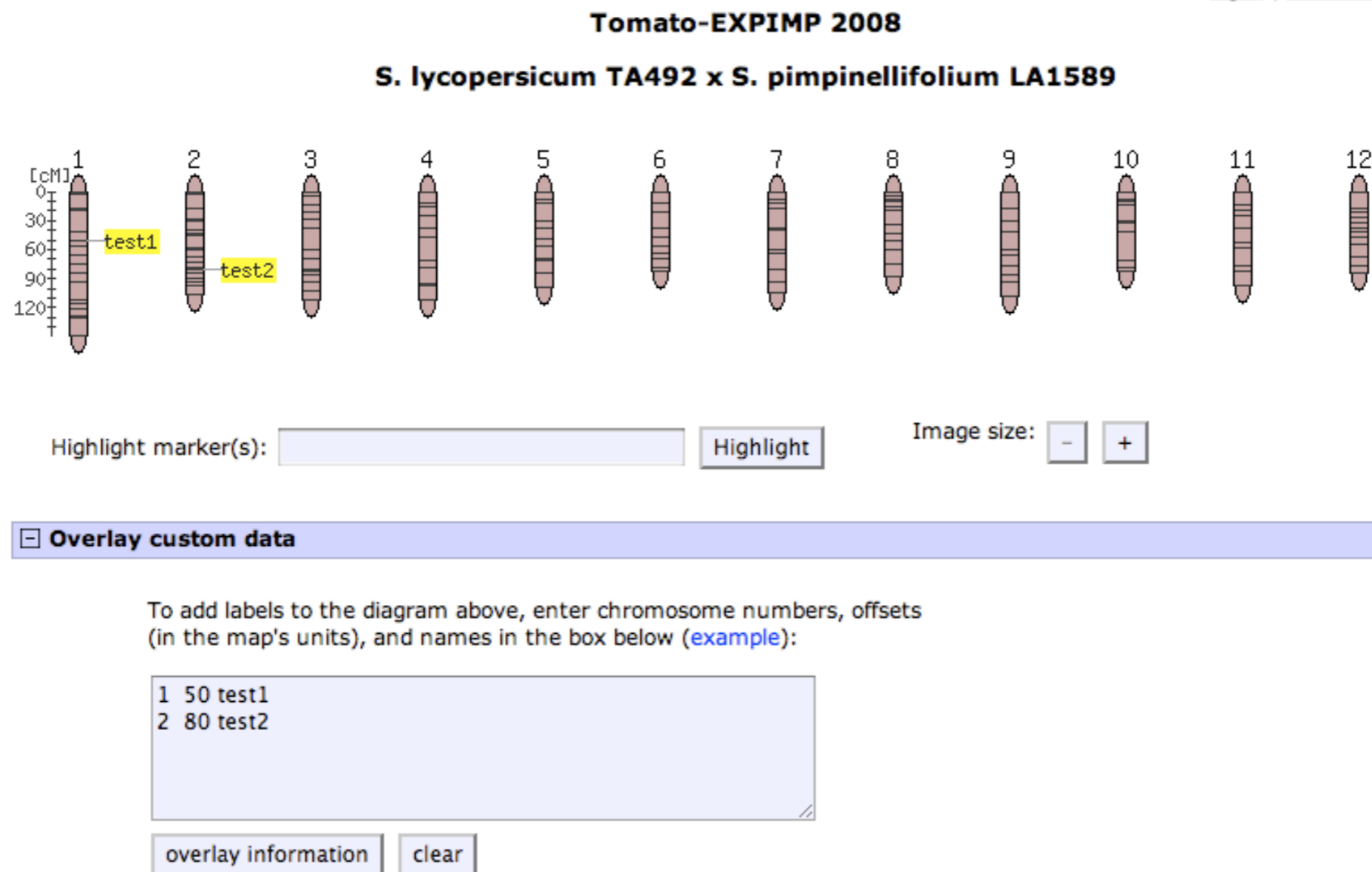
[Tomato-EXPIMP 2008](#): S. lycopersicum TA492 x S. pimpinellifolium LA1589

3.1 - Visualizadores de mapas de ligamiento



- CView (SGN): <http://solgenomics.net/cview/>

Representación de marcadores en la vista general del mapa.



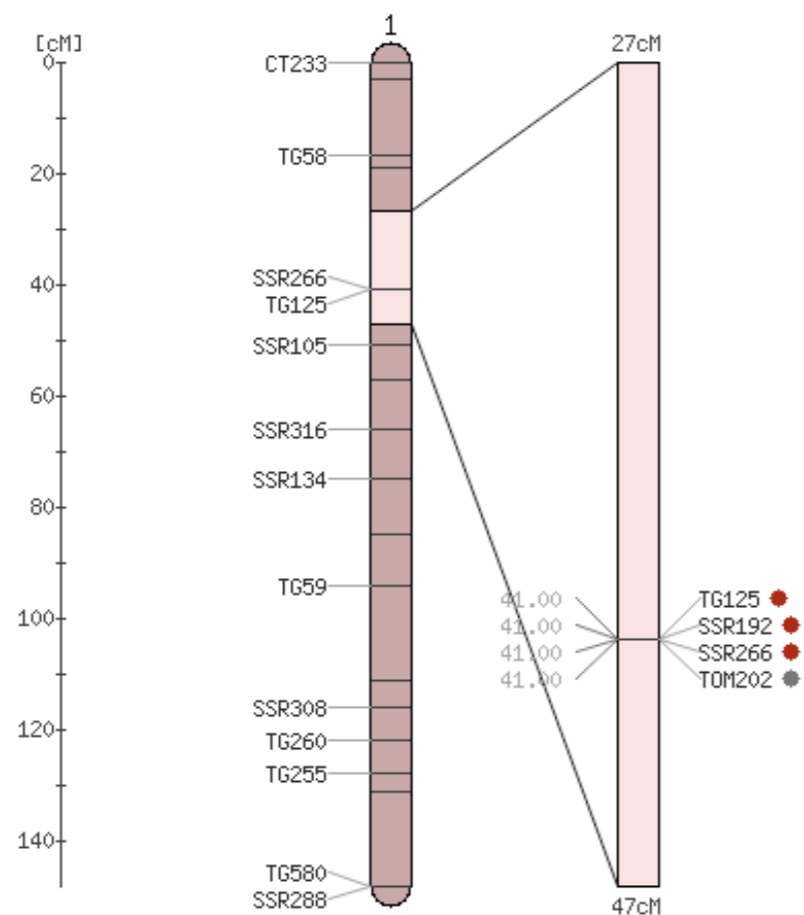


3.1 - Visualizadores de mapas de ligamiento

- CView (SGN): <http://solgenomics.net/cview/>

Selección de un cromosoma.

Viewing chr 1 of map **Tomato-EXPIMP 2008** [Help]



Marker color by LOD score: F(LOD3) | CF(LOD>=3) | I(LOD2) | I(LOD<2) | uncalculated

zoom range: to

Compare map to:

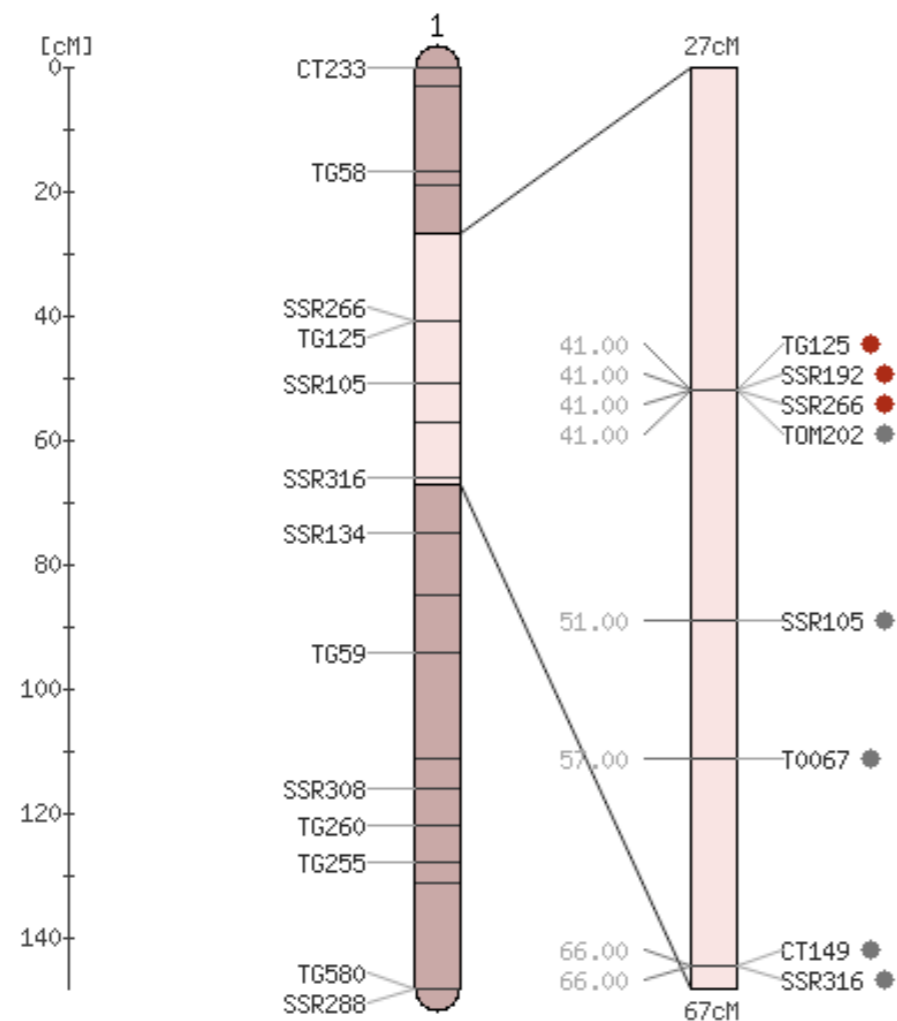


3.1 - Visualizadores de mapas de ligamiento

- CView (SGN): <http://solgenomics.net/cview/>

Cambio de coordenadas.

Viewing chr 1 of map **Tomato-EXPIMP 2008** [Help]



Marker color by LOD score: F(LOD3) | CF(LOD>=3) | I(LOD2) | I(LOD<2) | uncalculated

zoom range: to



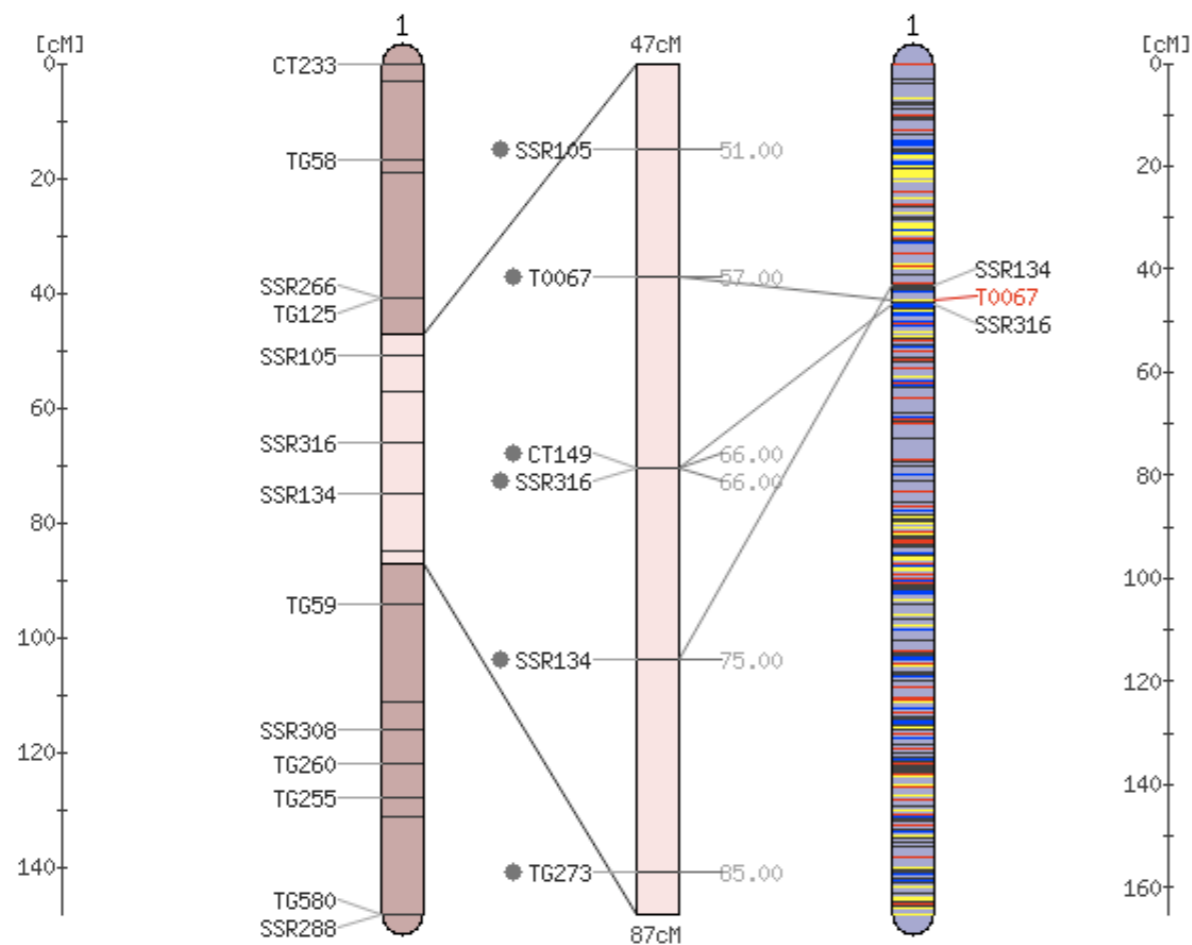


3.1 - Visualizadores de mapas de ligamiento

- CView (SGN): <http://solgenomics.net/cview/>

Comparación con otros cromosomas.

Viewing chr 1 of map **Tomato-EXPIMP 2008** Comparing to chr 1 of map **Tomato-EXPEN 2000** [Help]



Marker color by LOD score: F(LOD3) | CF(LOD>=3) | I(LOD2) | I(LOD<2) | uncalculated

| **zoom range:** to

Compare map to:





Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



3.2 - Diseño de cebadores (Primer3).

Algunas bases de datos ofrecen como herramienta el diseño de cebadores para PCR o análisis del fragmento amplificado.

- Primer-Blast (NCBI) (diseño):

<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>

- Primer3 (diseño):

<http://frodo.wi.mit.edu/primer3/>

- In-Silico PCR (SGN) (análisis de fragmento amplificado):

<http://solgenomics.net/tools/insilicopcr/index.pl>

3.2 - Diseño de cebadores (Primer3).



- Primer3 (diseño): <http://frodo.wi.mit.edu/primer3/>

Primer3 (v. 0.4.0) Pick primers from a DNA sequence.	Checks for mispriming in template.	disclaimer	Primer3 Home
	Primer3plus interface	cautions	FAQ/WIKI

Paste source sequence below (5'→3', string of ACGTNaagtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#):

<input checked="" type="checkbox"/> Pick left primer, or use left primer below:	<input type="checkbox"/> Pick hybridization probe (internal oligo), or use oligo below:	<input checked="" type="checkbox"/> Pick right primer, or use right primer below (5' to 3' on opposite strand):
<input type="text"/>	<input type="text"/>	<input type="text"/>

[Sequence Id:](#) A string to identify your output.

[Targets:](#) E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.

[Excluded Regions:](#) E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

[Product Size Ranges](#)

[Number To Return](#) [Max 3' Stability](#)

[Max Repeat Mispriming](#) [Pair Max Repeat Mispriming](#)

[Max Template Mispriming](#) [Pair Max Template Mispriming](#)



3.2 - Diseño de cebadores (Primer3).

- Primer3 (diseño): <http://frodo.wi.mit.edu/primer3/>

Gene Solyc04g078840.2

Gene

Name	Type	Organism
Solyc04g078840.2	Gene	<i>Solanum lycopersicum</i>

Synonyms / aliases	Length
Solyc04g078840	3,308

Description

BZIP transcription factor (AHRD V1 **-* Q0PN11_9FABA); contains Interpro domain(s) [IPR011616](#) bZIP transcription factor, bZIP-1

Location(s)

[SL2.40ch04:61097408..61094101](#)

GFF source

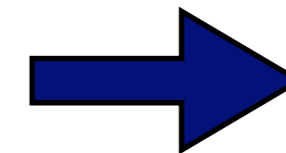
ITAG_eugene

From BOGAS

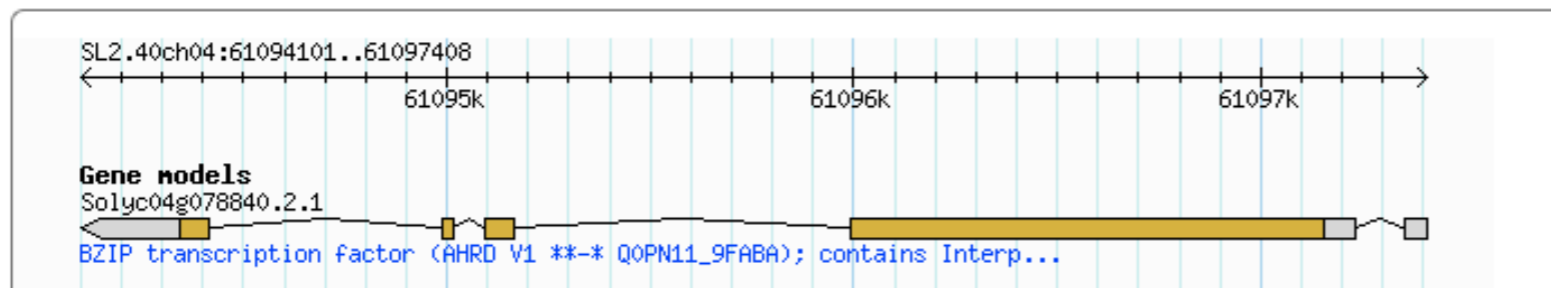
1

Length

3308



Diseño de
cebadores para
amplificar el primer
intron del gen
ARAB
(Sol04g078840.2)

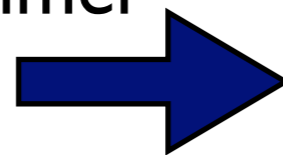




3.2 - Diseño de cebadores (Primer3).

- Primer3 (diseño): <http://frodo.wi.mit.edu/primer3/>

Diseño de
cebadores para
amplificar el primer
intron del gen
ARAB
(Sol04g078840.2)



features that are part of this mRNA

Summary

5	exon
4	intron
9	Total

Type	Name	Location(s)	Length	Strand	Phase
exon	exon:Solyc04g078840.2.1.1	SL2.40ch04:61097354..61097408	55	-	n/a
exon	exon:Solyc04g078840.2.1.2	SL2.40ch04:61095991..61097229	1,239	-	n/a
exon	exon:Solyc04g078840.2.1.3	SL2.40ch04:61095093..61095164	72	-	n/a
exon	exon:Solyc04g078840.2.1.4	SL2.40ch04:61094986..61095015	30	-	n/a
exon	exon:Solyc04g078840.2.1.5	SL2.40ch04:61094101..61094416	316	-	n/a
intron	intron:Solyc04g078840.2.1.2	SL2.40ch04:61097230..61097353	n/a	-	n/a
intron	intron:Solyc04g078840.2.1.3	SL2.40ch04:61095165..61095990	n/a	-	n/a
intron	intron:Solyc04g078840.2.1.4	SL2.40ch04:61095016..61095092	n/a	-	n/a
intron	intron:Solyc04g078840.2.1.5	SL2.40ch04:61094417..61094985	n/a	-	n/a

Calculamos las coordenadas de la region a volcar (+/- 200 pb intron):

$$\text{inicio: } 61097230 - 200 = 61097030$$

$$\text{final: } 61097353 + 200 = 61097553$$



3.2 - Diseño de cebadores (Primer3).



- Primer3 (diseño): <http://frodo.wi.mit.edu/primer3/>

Volcado de la secuencia usando GBrowse cambiando las coordenadas a: 61097030..61097553

2 → File Help

1 →

ITAG2.3 Release: genomic annotations: 524 bp from SL2.40ch04:61,097,030..61,097,553

Browser Select Tracks Upload and Share Tracks Preferences

Search

Landmark or Region: SL2.40ch04:61097030.. Search

Examples: Solyc05g005010, SL2.40ch08:2626000..2648999, Sulfite oxidase, SGN-M100, SGN-M7588.

Data Source ITAG2.3 Release: genomic annotations

Annotate Restriction Sites Configure... Go

Scroll/Zoom: << < - + >> >> Flip

Show 524 bp

Markers_region+Genes_region+mrna+sgn_unigenes+sgn_loci+Markers_overview+

>SL2.40ch04:61097030..61097553

```
AACGTCAATGAATAGATAGATGGCTGCCGCGGCAATCCAAAGTTCCTCCGGCTGCTTCCCG
CCACCACCACTTCCACCCTCACCTGGTGGCTCATTACCAAAGTTCCTGAAATGATAATTA
CTCCCCATTTCACTAAAACCTCCTCAGTCTCACACAATCCGCTTCAAACCTCAGCTCTG
TTATTCAAGAATCACAAAACCTACATATCAGATCAACAAGTTAATTCCTTCCCTTTGAA
CCTTTTTCTTATCATACTGTTCAACCCTTACATAAATGTACATCTATTTACAAACACA
CAGTTAATTAAGCAAAATATACCTGGAAAGAGATCAAAAATCAATTTACAGCTAAAAC
TCACCAACCCCTTATCAATAAAATCATCAAAAACAAATCCTATTTGAAATTCATTCTT
CAACTAAATTGACTGCATTTTCAGTTCACACCAAGAACCCCCAAAACACAACCTTCCCCAC
CCACCAATCCAATAAAGAACACACCTTTTTCACCTTCAATACAC
```

Overview

SL2.40ch04

0M 10M 20M 30M 40M

Markers

Region

61000k 61100k

Markers

CT50 C2_At3g49180 C2_At1g45976

Details

SL2.40ch04: 524 bp

200 bp

61097100 61097200 61097300 61097400 61097500

Gene models

Solyc04g078840.2.1

BZIP transcription factor (AHRD V1 *-* Q0PN11_9FABA); contains Interp...

3.2 - Diseño de cebadores (Primer3).



- Primer3 (diseño): <http://frodo.wi.mit.edu/primer3/>

Copiado de la secuencia volcada en el programa Primer3.

El tamaño mínimo debe ser 123 pb (el tamaño del intrón), y tendrá como región objetivo (*target*): 200 (comienzo del intrón), 123 (longitud del intrón)

Primer3 (v. 0.4.0) Pick primers from a DNA sequence.	Checks for mispriming in template.	disclaimer	Primer3 Home
	Primer3plus interface	cautions	FAQ/WIKI

Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#):

```
AACGTCAATGAATAGATAGATGGCTGCCGCGGCAATCCAAAGTTCCCCGGCTGCTTCCCG
CCACCACCACTTCCACCCTCACCTGGTGGCTCATTACCAAAGTTCTTGAAATGATAATTA
CTCCCCATTTCACTAAAACCTCCTCAGTCCTCACACAATCCGCCTTCAAACCTCAGCTCTG
TTATTCAAGAATCACAAAACCTACATATCAGATCAACAAGTTAATTCCTTCCCTTTGAA
CCTTTTTCCTTATCATACTGTTCAACCCTTACATAAATGTACATCTATTTACAAACACA
CAGTTAATTAAGCAAAATATACCTGGAAAGAGATCAAAAATCAATTTACAGCTAAAAC
```

<input checked="" type="checkbox"/> Pick left primer, or use left primer below:	<input type="checkbox"/> Pick hybridization probe (internal oligo), or use oligo below:	<input checked="" type="checkbox"/> Pick right primer, or use right primer below (5' to 3' on opposite strand):
<input type="text"/>	<input type="text"/>	<input type="text"/>

Sequence Id: A string to identify your output.

Targets: E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [and]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.

Excluded Regions: E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

Product Size Ranges

Number To Return <input type="text" value="5"/>	Max 3' Stability <input type="text" value="9.0"/>
Max Repeat Mispriming <input type="text" value="12.00"/>	Pair Max Repeat Mispriming <input type="text" value="24.00"/>
Max Template Mispriming <input type="text" value="12.00"/>	Pair Max Template Mispriming <input type="text" value="24.00"/>

3.2 - Diseño de cebadores (Primer3).



- In-Silico PCR (SGN): <http://solgenomics.net/tools/insilicopcr/index.pl>

La ampliación inespecífica en el genoma de tomate puede comprobarse utilizando la herramienta In-silico PCR del SGN.

In Silico PCR

PCR Primers:

Forward Primer

Reverse Complement Forward Primer

Reverse Primer

Reverse Complement Reverse Primer

Product Maximum Length

Allowed Mismatches

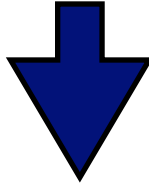
BLAST Attributes:

Database (-d)
[db details](#)

BLAST Program (-p)

Substitution Matrix (-M)

Expectation value (-e)



FORWARD-PRIMER	SL2.40ch04	100.00	19	0	0	1	19	61097186	61097204	0.008	38.2
REVERSE-PRIMER	SL2.40ch04	100.00	21	0	0	1	21	61097382	61097402	8e-04	42.1



Herramientas de las Bases de Datos Biológicas:

1 - Herramientas de búsqueda:

1.1 - Búsqueda por ontologías.

1.2 - Búsqueda por homología de secuencias (Blast).

1.3 - Búsqueda por localizaciones cromosómicas (GBrowse).

2 - Herramientas de manipulación y análisis de secuencias:

2.1 - Traductores y *Gene Predictors*.

2.2 - Alineamiento de múltiples secuencias (Clustalw).

2.3 - Análisis de dominios funcionales (InterProScan).

2.4 - Análisis de péptidos señal (SignalP).

3 - Otras herramientas:

3.1 - Visualizadores de mapas de ligamiento (CViewer).

3.2 - Diseño de cebadores (Primer3).

4 - Portales con listado de herramientas.



Algunos de los portales con herramientas bioinformáticas *on-line*:

- EBI (European Bioinformatic Institute): Análisis de secuencias.

<http://www.ebi.ac.uk/Tools/>

- Mobyly (Instituto Pasteur): Análisis filogenético.

<http://mobyly.pasteur.fr/cgi-bin/portal.py?#welcome>

- ExPASy (SwissProt): Análisis proteómico y de secuencias.

<http://expasy.org/tools/>

- CBS (Center For Biological Sequence Analysis).

<http://www.cbs.dtu.dk/biotools/>



PROGRAMA:

11.00 – 11.15 am. Introducción a la Bioinformática.

11.15 – 11.30 am. Caso de estudio: Estrés por sequía en tomate.

11.30 - 12.15 pm. Bases de datos biológicas.

12.15 – 13.00 pm. Herramientas comunes de las bases de datos.

13.00 – 14.30 pm. Almuerzo y descanso.

14.30 – 15.00 pm. Casos prácticos del bloque de base de datos y anotación.

15.00 – 15.45 pm. Terminal Linux-Ux y líneas de comandos.

15.45 – 16.30 pm. Bases de datos.

16.30 – 17.00 pm. Conclusiones y preguntas.

