

Motivos de ADN (Ejercicio 5)

5.1 Identificar motivos específicos de ADN en genomas.

Nota: Para este ejercicio usar <http://microsporidiadb.org>

- Encontrar todos los sitios de restricción de BamHI, en todas las secuencias genómicas de microsporidios disponibles en MicrosporidiaDB. (Pista: BamHI = GGATCC; buscar por “DNA Motif PAttern” que está bajo el título “Genomic Segments” en “Identify Other Data Types”).

The screenshot shows the MicrosporidiaDB website interface. At the top, there is a search bar for Gene ID (ECU03_0820) and Gene Text Search (synth*). Below the search bar, there are navigation tabs: Home, New Search, My Strategies, My Basket (0), Tools, Data Summary, Downloads, Community, and My Favorites. The main content area is divided into several sections:

- Data Summary**: Contains a News section with updates from May 2011 and November 2010, and a Community Resources section.
- Identify Genes by:** A list of search criteria including Text, IDs, Species, Genomic Position, Gene Attributes, Protein Attributes, SimilarityPattern, Transcript Expression, Cellular Location, Putative Function, and Evolution.
- Identify Other Data Types:** A list of search criteria including Genomic Sequences, Genomic Segments (DNA Motif), DNA Motif Pattern (highlighted with a red arrow), Genomic Location, ESTs, and ORFs.
- Tools**: Includes BLAST, Identify Sequence Similarities, Sequence Retrieval, PubMed and Entrez, Genome Browser, and Searches via Web Services.

A red box highlights the "Identify Genomic Segments based on DNA Motif Pattern" search form. It contains the following fields and options:

- Pattern:** A text input field containing "GGATCC".
- Organism:** A list of checkboxes for different microsporidian species: Encephalitozoon cuniculi GB-M1, Encephalitozoon intestinalis, Enterocytozoon bieneusi H348, Nematocida parisii ERTm1, Nematocida parisii ERTm3, and Octospora bayeri CER-3-3. There are also "select all" and "clear all" options.
- Give this search a weight:** A checkbox.
- Give this search a name:** A checkbox.
- Get Answer:** A button to execute the search.

- ¿Cuántas veces encontraste el sitio BamHI en esos genomas? Vuelve a mirar los resultados poniendo atención a las columnas que indican la posición genómica (Genomic location) y los motivos (Motifs).

My Strategies: [New](#) [Opened \(1\)](#) [All \(4\)](#) [Basket](#) [Examples](#) [Help](#)

(Sequence Segments)

DNA Motif(2) Backbone Copy Save As Share Delete

DNA Motif(2) Step 1 Add Step

DNA Motif(2) - step 1 - 5778 Genomic Segments Download 5778 Genomic Segments

First 1 2 3 4 5 Next Last Advanced Paging Select Columns Open Columns

Segment ID	Organism	Genomic Location	Heidi
NC_003229:103001-103007	Encephalitozoon cuniculi QS-M1	NC_003229:103001 - 103007 (+)	..TTCCATCTTTGGGCTTGATGGATCGGAGATATCAAGCTTTTGT..
NC_003229:103008-103014	Encephalitozoon cuniculi QS-M1	NC_003229:103008 - 103014 (+)	..GCCATGATGAATGGGGGAAGGGTGTCTAGGAGACAGCTTTATGA..
NC_003229:103702-103708	Encephalitozoon cuniculi QS-M1	NC_003229:103702 - 103708 (+)	..CATGGTCCGAATGAGCTTCGGATGCAAGTTCTTGTAAATGAGC..
NC_003229:111898-111898	Encephalitozoon cuniculi QS-M1	NC_003229:111898 - 111898 (+)	..GTGACTATGCGGTGACCAATGGATCGAGATGCAAAAGCCCTGTAA..
NC_003229:118373-118379	Encephalitozoon cuniculi QS-M1	NC_003229:118373 - 118379 (+)	..TATAGTAACCTCTCCGAGCCGGATCCAGGGTGGCAATCTTCTCA..
NC_003229:126287-126293	Encephalitozoon cuniculi QS-M1	NC_003229:126287 - 126293 (+)	..ATACAGCATGGGCGACTTGGGATGCTGTGTGAGGATTCCTTGGG..
NC_003229:127134-127140	Encephalitozoon cuniculi QS-M1	NC_003229:127134 - 127140 (+)	..TCCGAAATGATGATCTCGAGATATCTGGGCTCGATCCAGGATCT..
NC_003229:128216-128222	Encephalitozoon cuniculi QS-M1	NC_003229:128216 - 128222 (+)	..AATGCTTGTATGAGCCGGGGATCTGTGGCTTGGCTTGGCTTGG..
NC_003229:130478-130484	Encephalitozoon cuniculi QS-M1	NC_003229:130478 - 130484 (+)	..CCTGTTAAGCGGCGCTGGGATCCAGAACACACAAAGCTTAAG..
NC_003229:132036-132042	Encephalitozoon cuniculi QS-M1	NC_003229:132036 - 132042 (+)	..GAGAAATATCTGTTGAGATGGATCCGGGAAAGCAGGATCTCACT..
NC_003229:143457-143463	Encephalitozoon cuniculi QS-M1	NC_003229:143457 - 143463 (+)	..AGGTTCCGCTGCTCCTCTGGATCTGGATGAACTCCTCACTA..
NC_003229:143776-143782	Encephalitozoon cuniculi QS-M1	NC_003229:143776 - 143782 (+)	..ATGCAATGGAATGACTGGGATCTCCGCGCCGCTTGGGATGCT..
NC_003229:144898-144904	Encephalitozoon cuniculi QS-M1	NC_003229:144898 - 144904 (+)	..ATATACCAAAAAGAGAGGATCTTCCGTTCCGAAATCTCACT..
NC_003229:148857-148863	Encephalitozoon cuniculi QS-M1	NC_003229:148857 - 148863 (+)	..ATAGCTTGGAAATCTTGGATCTGGATCCGAGCAGGATGATAGCTGT..
NC_003229:148870-148876	Encephalitozoon cuniculi QS-M1	NC_003229:148870 - 148876 (+)	..AATATGGGCTGGATCTGGATCTTAGCCAGCGGAGCTGAGAT..
NC_003229:147200-147206	Encephalitozoon cuniculi QS-M1	NC_003229:147200 - 147206 (+)	..AGATCATGTATAGCGCTGTGGATCCATATGCAATAGCGCATCAC..

5.2 En la sección 6.1 encontre los sitios BamHI. ¿Cómo harías para encontrar genes que estén 500 nucleótidos alrededor de los sitios BamHI? Por ejemplo: ¿podrías encontrar todos los genes que estén posicionados dentro de los 500 nucleótidos río abajo (downstream) de un sitio BamHI?

- a. Esto lo puedes hacer mediante una búsqueda de colocación, definiendo la relación entre los sitios BamHI y los genes en contexto genómico (pista: agregar un paso a la estrategia de motivos, buscando genes por especie (genes/species) y seleccionando la opción “1 relative to 2, using “genomic locations”).

My Strategies: New | Opened (1) | All (4) | Basket | Examples | Help

(Genomic Segments)

DNA Motif(2)*
 Remain | Copy | Save As | Share | Delete

Step 1: DNA Motif
 5778 Segments

Add Step (1)

Run a new Search for: Genes
 Add contents of Basket: Genomic Segments (DNA Motif)
 Add existing Strategy: ORFs
 Filter by Weight:

Download 5778 Genomic Segments

Segment ID	Organism	Genomic Position
NC_003229:103601-103601	Encephalitozoon cuniculi	103601-103601
NC_003229:105065-105065	Encephalitozoon cuniculi	105065-105065
NC_003229:106762-106762	Encephalitozoon cuniculi	106762-106762
NC_003229:111938-111938	Encephalitozoon cuniculi	111938-111938
NC_003229:118373-118373	Encephalitozoon cuniculi	118373-118373
NC_003229:126087-126087	Encephalitozoon cuniculi	126087-126087
NC_003229:127134-127140	Encephalitozoon cuniculi	127134-127140
NC_003229:129014-129020	Encephalitozoon cuniculi	129014-129020
NC_003229:130478-130484	Encephalitozoon cuniculi	130478-130484
NC_003229:132534-132540	Encephalitozoon cuniculi	132534-132540
NC_003229:143457-143463	Encephalitozoon cuniculi	143457-143463
NC_003229:143776-143782	Encephalitozoon cuniculi	143776-143782
NC_003229:144888-144894	Encephalitozoon cuniculi	144888-144894
NC_003229:14547-14553	Encephalitozoon cuniculi	14547-14553
NC_003229:146669-146675	Encephalitozoon cuniculi	146669-146675

Add Step (2)

Text, IDs, Species: Text product name, notes
 Genomic Position: n/c
 Gene Accession: Gene ID
 Protein Abbrev: Species

Add Step 2: Species (3)

Organism:
 Encephalitozoon cuniculi
 Encephalitozoon intestinalis
 Encephalitozoon binuclei
 select all | clear all

Give this search a weight:
 Give this search a name:

Combine Genomic Segments in Step 1 with Genes in Step 2:

1 Intersect 2 | 1 Minus 2
 2 Union 1 | 2 Minus 1
 3 Relative to 2, using genomic locations

Add Step (4)

Add Step

Combine Step 1 and Step 2 using relative locations in the genome

You had 5778 Genomic Segments in your Strategy (Step 1). Your new Genes search (Step 2) returned 7748 Genes.

Return each [Genomic Segment] whose exact region overlaps the exact region of a [Gene] in Step 2 and is on either strand

5

Exact:
 Upstream: 1000 bp
 Downstream: 1000 bp

Custom:
 begin at: start (-) | end (+) bp
 end at: start (-) | end (+) bp

¿Cómo modificarías la ubicación en relación a los genes?

(5778 Genomic Segments in Step 1)

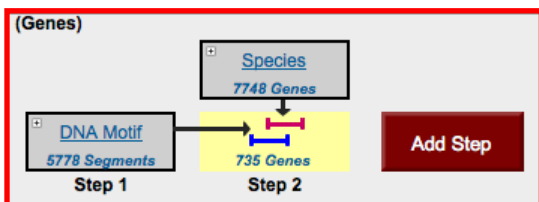
Region

Genomic Segment

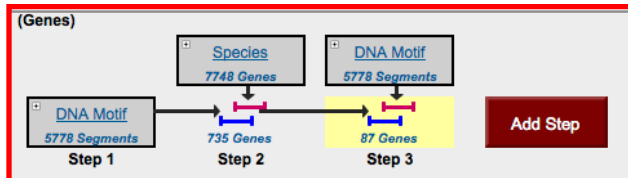
Upstream: 500 bp

Downstream: 1000 bp

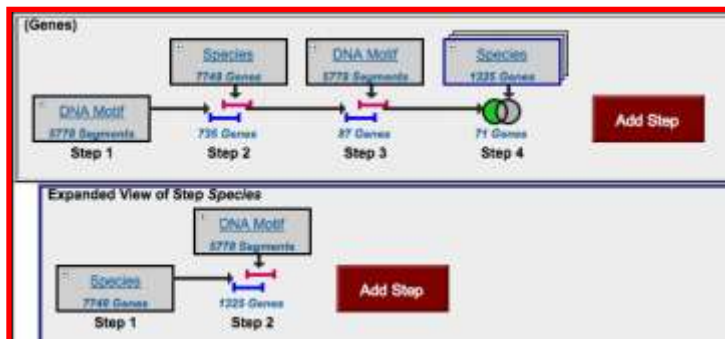
¿Cuántos genes obtuviste?



5.3 Empleando una secuencia de pasos similar a la del punto 5.2, ¿podrías definir cuáles de esos genes además tienen un sitio BamHI en la región comprendida dentro de las 500 bases río abajo (downstream)? (Pista: después de *clickear* en “Add Step” deberás seleccionar la búsqueda de motivos de DNA y la opción “genomic collocation”).



5.4 Llevando esto un paso adelante, ¿podrías definir cuáles de éstos genes **NO** contienen un sitio BamHI dentro de ellos? (Pista: podrás usar una estrategia anidada [nested strategy] para hacer esto).



Mirá los resultados... ¿tienen sentido? Confirma que obtuviste lo que buscabas mirando alguno de los genes en el “gbrowse”, configurándolo para ver los sitios de restricción BamHI.

Nota: Puedes agregar columnas a tus resultados haciendo click en el botón “Select Columns” (ver imagen abajo). Para el ejemplo anterior, se puede agregar la columna GBrowse.

Filter results by species (results removed by the filter will not be combined into the next step.)

All Results	Ortholog Groups	Encephalitozoon cuniculi	Encephalitozoon intestinalis	Enterocytozoon blauschi
71	69	34	25	12

DNA Motif - step 4 - 71 Genes Add 71 Genes to Basket | Download 71 Genes

First 1 2 3 4 Next Last Advanced Paging Select Columns | Reset Columns

Gene	Genomic Location	Product Description	Weight	Chromosome
EBI_24411	ABGB01000099: 438 - 728 (+)	hypothetical protein	10	N/A
EBI_25438	ABGB01000276: 1,036 - 1,248 (-)	hypothetical protein	10	N/A
EBI_25638	ABGB01000541: 218 - 430 (-)	hypothetical protein		
EBI_25857	ABGB01001308: 181 - 393 (+)	hypothetical protein		
EBI_25706	ABGB01000890: 191 - 403 (+)	hypothetical protein		
EBI_26324	ABGB01000351: 1,323 - 1,454 (+)	hypothetical protein		
EBI_25491	ABGB01000853: 329 - 541 (-)	hypothetical protein		
EBI_26096	ABGB01000982: 532 - 744 (+)	hypothetical protein		
EBI_26621	ABGB01000486: 358 - 558 (+)	hypothetical protein		
EBI_27056	ABGB01001170: 475 - 687 (+)	hypothetical protein		
EBI_27581	ABGB01000203: 975 - 1,491 (-)	hypothetical protein		

Select Columns close X

Update Columns

select all | clear all

<input checked="" type="checkbox"/> Gene	<input type="checkbox"/> Transcript Length	<input type="checkbox"/> Ortholog Group
<input checked="" type="checkbox"/> GBrowse ←	<input type="checkbox"/> CDS Length	<input type="checkbox"/> SignalP Scores
<input type="checkbox"/> Genomic Sequence ID	<input checked="" type="checkbox"/> Product Description	<input type="checkbox"/> SignalP Peptide
<input checked="" type="checkbox"/> Chromosome	<input type="checkbox"/> Protein Length	<input type="checkbox"/> Predicted GO Function
<input type="checkbox"/> Chromosome Order	<input type="checkbox"/> # TM Domains	<input type="checkbox"/> Predicted GO Process
<input checked="" type="checkbox"/> Genomic Location	<input type="checkbox"/> Molecular Weight	<input type="checkbox"/> Predicted GO Component
<input type="checkbox"/> Gene Strand	<input type="checkbox"/> Isoelectric point	<input type="checkbox"/> Organism
<input type="checkbox"/> Gene Type	<input type="checkbox"/> Ortholog count	<input type="checkbox"/> Is Pseudo
<input type="checkbox"/> # Exons	<input type="checkbox"/> Paralog count	<input checked="" type="checkbox"/> Weight

select all | clear all

Update Columns

Nota: Puedes configurar los sitios de restricción haciendo click en el botón de configuración en el Gbrowse y seleccionando los sitios de restricción de interés (ver imagen abajo).

Search

NOTE: If you load tracks and they appear empty, you can try two things to resolve this issue:
 1. Make sure you are viewing the correct species/strain to which the data was mapped.
 2. Reset gbrowse by clicking on the red **Reset** link, then try again.

Landmark or Region: NC_003235.104447..124448 Reports & Analysis: Annotate Restriction Sites Configure... Go

Data Source: MicrosporidiaDB
 Select the Data Source of your choice (above)

Overview

Region

Details

Associated Genes (with HTOs in gray when available)

SeeK1 restriction site

Configure... Restriction Sites Cancel | Configure

Select Restriction Sites To Annotate

Restriction Site Display off on

<input type="checkbox"/> AatII	<input type="checkbox"/> BspDI	<input type="checkbox"/> HpaII	<input type="checkbox"/> PspGI
<input type="checkbox"/> Acc65I	<input type="checkbox"/> BspEI	<input type="checkbox"/> Hpy188II	<input type="checkbox"/> PspOMI
<input type="checkbox"/> AccI	<input type="checkbox"/> BspHI	<input type="checkbox"/> Hpy188III	<input type="checkbox"/> PstI
<input type="checkbox"/> AclI	<input type="checkbox"/> BsrFI	<input type="checkbox"/> Hpy99II	<input type="checkbox"/> PvuI
<input type="checkbox"/> AfeI	<input type="checkbox"/> BsrGI	<input type="checkbox"/> HpyCH4II	<input type="checkbox"/> PvuII
<input type="checkbox"/> AflII	<input type="checkbox"/> BsaHI	<input type="checkbox"/> HpyCH4IV	<input type="checkbox"/> RsaI
<input type="checkbox"/> AflIII	<input type="checkbox"/> BsaKI	<input type="checkbox"/> HpyCH4V	<input type="checkbox"/> RsrII
<input type="checkbox"/> AgoI	<input type="checkbox"/> BstAPI	<input type="checkbox"/> KasI	<input type="checkbox"/> SacI
<input type="checkbox"/> AhdI	<input type="checkbox"/> BstBI	<input type="checkbox"/> KpnI	<input type="checkbox"/> SacII
<input type="checkbox"/> AluI	<input type="checkbox"/> BstEII	<input type="checkbox"/> MboI	<input type="checkbox"/> SalI
<input type="checkbox"/> AlwHI	<input type="checkbox"/> BstNI	<input type="checkbox"/> MfeI	<input type="checkbox"/> Sau3AI
<input type="checkbox"/> ApaI	<input type="checkbox"/> BstUI	<input type="checkbox"/> MluI	<input type="checkbox"/> Sau96I
<input type="checkbox"/> ApaLI	<input type="checkbox"/> BstXI	<input type="checkbox"/> MscI	<input type="checkbox"/> SbfI
<input type="checkbox"/> Apol	<input type="checkbox"/> BstYI	<input type="checkbox"/> MseI	<input type="checkbox"/> ScaI
<input type="checkbox"/> AscI	<input type="checkbox"/> BstZ17I	<input type="checkbox"/> MslI	<input type="checkbox"/> ScrFI
<input type="checkbox"/> AseI	<input type="checkbox"/> Bsu36I	<input type="checkbox"/> MspA1I	<input type="checkbox"/> SexAI
<input type="checkbox"/> AscI	<input type="checkbox"/> BglI	<input type="checkbox"/> MspI	<input type="checkbox"/> SfiI
<input type="checkbox"/> AvaI	<input type="checkbox"/> Cac8I	<input type="checkbox"/> MwoI	<input type="checkbox"/> SfiII
<input type="checkbox"/> AwaII	<input type="checkbox"/> ClaI	<input type="checkbox"/> NaeI	<input type="checkbox"/> SfiIII
<input type="checkbox"/> AvrII	<input type="checkbox"/> DdeI	<input type="checkbox"/> NarI	<input type="checkbox"/> SgrAI
<input checked="" type="checkbox"/> BamHI	<input type="checkbox"/> DpnI	<input type="checkbox"/> NciI	<input type="checkbox"/> SmaI
<input type="checkbox"/> BclI	<input type="checkbox"/> DpnII	<input type="checkbox"/> NcoI	<input type="checkbox"/> SmaII