

DNA Motifs Exercise 6

6.1 Identification of specific DNA motifs.

Note: For this exercise use <http://microsporidiadb.org>

- Find all BamHI restriction sites in all microsporidia genomic sequences available in MicrosporidiaDB. Note: you can use this search to find complex motifs like transcription factor binding sites using regular expressions.

Hint: BamHI = GGATCC and the DNA motif search is under the heading "Genomic Segments".

- How many times does the BamHI site occur in the genomes you searched? Take a look at your results; notice the Genomic location and the Motif columns.

Segment ID	Organism	Genomic Location	Motif
EcEC1_supercont1.1:100913-100919.f	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 100913 - 100919 (+)	..AGAAGTCGAAGCTCCACTCCGGATCCATCGACAGTATCTCCOCTC...
EcEC1_supercont1.1:100913-100919.r	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 100913 - 100919 (-)	..GAGGGGAAGATACTGTGATGGATCCGGAGTGGAGCTTCGACTTCT...
EcEC1_supercont1.1:105820-105826.f	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 105820 - 105826 (+)	..GAGAAACGAGGAGCTTTCGTGGATCCCTTGGAGAGATACTCGGACC...
EcEC1_supercont1.1:105820-105826.r	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 105820 - 105826 (-)	..GGTCCGCGATATCTCTCCAAGGATCCACGAAAGCTCCTCTTTCTC...
EcEC1_supercont1.1:107855-107861.f	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 107855 - 107861 (+)	..GGACTGGTCGGCGTGTATAGGGATCCCATGAAGCGCTCAGCAAAG...
EcEC1_supercont1.1:107855-107861.r	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 107855 - 107861 (-)	..CTTTGCTGACGCCCTTCATGGGATCCCTATACACGCCGACCAGTCC...
EcEC1_supercont1.1:108534-108540.f	Encephalitozoon cuniculi EC1	EcEC1_supercont1.1: 108534 - 108540 (+)	..GACACAAAAACAAGAGCAGGATCCAGAGCCATCATATGGAGGCG...

6.2 Find genes that have one of these BamHI sites within 500 nucleotides upstream of their start.

In the section 6.1 you found BamHI sites, but now you are looking for genes that have one of these sites located within 500 nucleotides upstream of their start.

Hint: You can achieve this by running a genomic collocation search defining the genomic relationship between the BamHI sites and genes. Add a “Genes by Organism” step to the motif search and select the “1 relative to 2, using genomic locations” option.

The screenshot displays a bioinformatics search interface with the following components:

- My Strategies:** New, Opened (1), All (2), Basket, Examples, Help
- (Genomic Segments):** DNA Motif(2) 13856 Segments, Step 1
- 1:** Points to the 'Add Step' button in the 'DNA Motif(2)' search step.
- 2:** Points to the 'Add Step' button in the 'Add Step' dialog box.
- 3:** Points to the 'Organism' field in the 'Add Step' dialog box, which is set to 'Encephalitozoon', 'Enterocytozoon', and 'Nosema'.
- 4:** Points to the '1 Relative to 2, using genomic collocation' option in the 'Combine Genomic Segments in Step 1 with Genes in Step 2:' section.

The 'Add Step' dialog box is titled 'Add Step 2 : Organism' and contains the following configuration:

- Organism:** select all | clear all | expand all | collapse all | reset to default
 - Encephalitozoon
 - Enterocytozoon
 - Nosema
- select all | clear all | expand all | collapse all | reset to default
- Give this search a weight
- Give this search a name

The 'Combine Genomic Segments in Step 1 with Genes in Step 2:' section shows the following options:

- 1 Intersect 2
- 1 Union 2
- 1 Minus 2
- 2 Minus 1
- 1 Relative to 2, using genomic collocation

The 'Continue...' button is located at the bottom right of the dialog box.

5

Add Step

Genomic Colocation ?

Combine Step 1 and Step 2 using relative locations in the genome
 You had **13856 Genomic Segments** in your Strategy (Step 1). Your new **Genes** search (Step 2) returned **12339 Genes**.

"Return each **select...** whose **exact region** overlaps the **exact region** of a ... in Step ... and is on **either strand**"

(13856 Genomic Segments in Step)

Exact

Upstream: 1000 bp

Downstream: 1000 bp

Custom:

begin at: start + 0 bp

end at: stop + 0 bp

(12339 Genes in Step)

Exact

Upstream: 1000 bp

Downstream: 1000 bp

Custom:

begin at: start + 0 bp

end at: stop + 0 bp

Submit

How did you modify the location relative to genes?

"Return each **Gene from Step 2** whose **upstream region**

(12339 Genes in Step)

Exact

Upstream: 500 bp

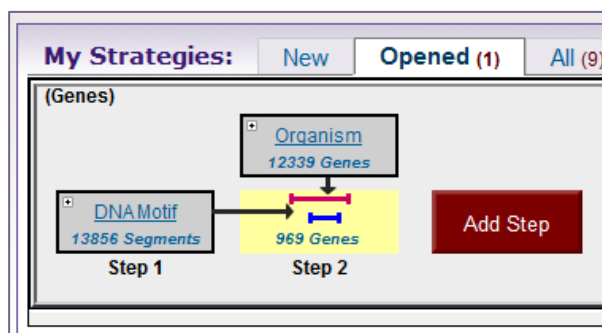
Downstream: 1000 bp

Custom:

begin at: start - 500 bp

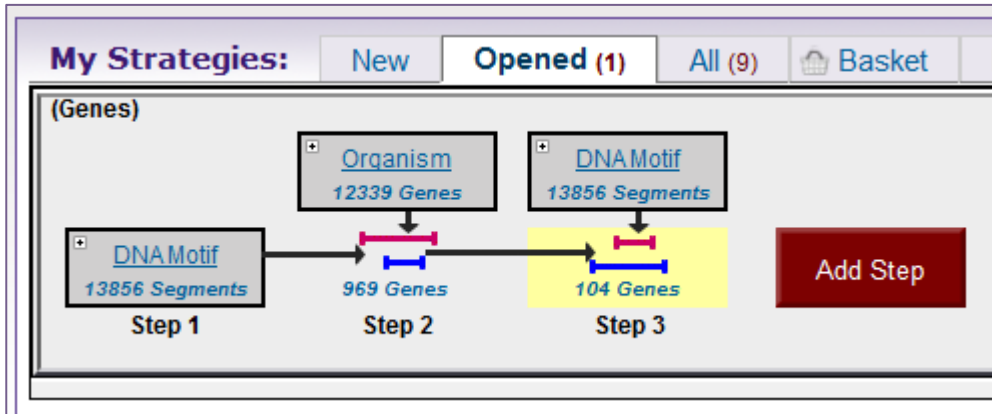
end at: start - 1 bp

How many genes did you get?



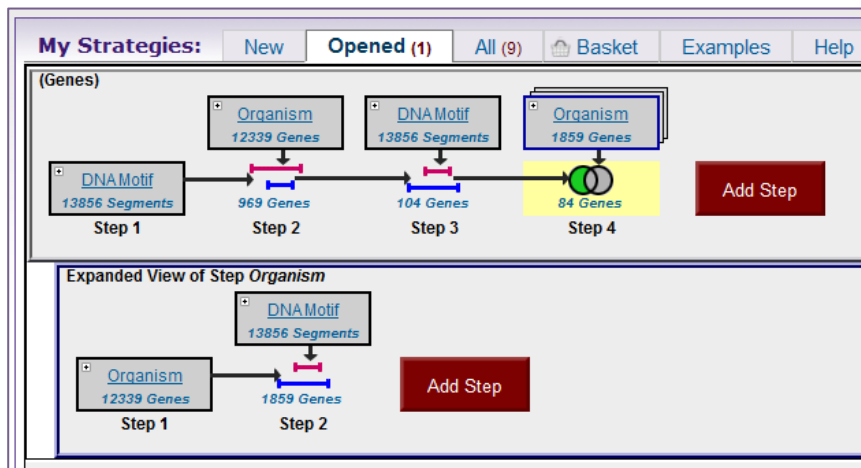
6.3 Using a similar sequence of steps as in part 6.2, define which of these genes also have a BamHI site in their 500 nucleotide downstream region.

Hint: after you click on add step you will have to select DNA motif search and select the genomic collocation option.



6.4 Taking this a step further, define which of these genes does NOT contain a BamHI site within them.

Hint: you will have to use a nested strategy.



Look at your results. Do they make sense? Confirm your results by looking at one of the genes in Gbrowse and showing BamHI restriction sites.

Note: you can add columns to your results by clicking on the Select Columns button (see image below). For the example above, you can add a GBrowse column that allows you to go directly to GBrowse at the genomic coordinates of any ID in your result list.

My Step Result:

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

All Results	Ortholog Groups	<i>Encephalitozoon cuniculi</i>	<i>Encephalitozoon hellem</i>	<i>Encephalitozoon intestinalis</i>	<i>Enterocytozoon bieneusi</i>	<i>Nosema ceranae</i>
84	70	31	18	23	12	0

DNA Motif - step 4 - 84 Genes [Add 84 Genes to Basket](#) | [Download 84 Genes](#)

Genes | Genome View (beta)

Advanced Paging

Select Columns

Gene ID	Genomic Location	Product Description
EBI_24411	ABGB01000099: 438 - 728 (+)	hypothetical protein
EBI_27581	ABGB01000203: 976 - 1,491 (-)	hypothetical protein
EBI_25435	ABGB01000276: 1,036 - 1,248 (-)	hypothetical protein
EBI_26304	ABGB01000351: 1,323 - 1,454 (+)	hypothetical protein
EBI_26621	ABGB01000486: 358 - 558 (+)	hypothetical protein
EBI_25638	ABGB01000541: 218 - 430 (-)	hypothetical protein
EBI_25705	ABGB01000850: 191 - 403 (+)	hypothetical protein
EBI_26491	ABGB01000853: 329 - 541 (-)	hypothetical protein
EBI_26598	ABGB01000992: 532 - 744 (+)	hypothetical protein
EBI_27558	ABGB01001170: 475 - 687 (+)	hypothetical protein
EBI_27632	ABGB01001257: 59 - 238 (+)	aspartate-ammonia ligase

Select Columns

Update Columns

clear all | expand all | collapse all
reset to current | reset to default

- Text, IDs, Species
- Genomic Position
 - Chromosome
 - Genomic Location
 - Gene Strand
- Gene Attributes
- Protein Attributes
 - Product Description
 - Molecular Weight
 - Isoelectric Point
- Protein Features
- Putative Function
- Evolution
- GBrowse
- Weight

clear all | expand all | collapse all
reset to current | reset to default

Update Columns

Note: you can configure restriction sites by clicking on the configure button in GBrowse and selecting the restriction sites you would like to display (see image below). To view restriction sites, the “Restriction Sites” data track must be turned on. Go to the “Select Tracks” page and click “Restriction Sites” under the “Analysis” section.

The image shows the GBrowse web interface. At the top, there are navigation tabs: "Browser", "Select Tracks", "Snapshots", "Custom Tracks", and "Preferences". Below these is a search bar with the text "NC_003229:162,593..182,592". To the right of the search bar is a dropdown menu labeled "Annotate Restriction Sites" with a "Configure..." button next to it. Below the search bar are buttons for "Save Snapshot" and "Load Snapshot". Further right are scroll/zoom controls including "Show 20 kbp" and "Flip".

The main content area is divided into several tracks:

- Overview:** Shows a genomic map for NC_003229 with a scale from 0k to 190k.
- Region:** A zoomed-in view of the region from 0k to 190k.
- Details:** A zoomed-in view of a 20 kbp region (NC_003229: 20 kbp) with a 5 kbp scale. It shows annotated genes: ECU02_1360, ECU02_1370, ECU02_1380, ECU02_1390, ECU02_1400, ECU02_1410, ECU02_1420, and ECU02_1430. UTRs are shown in gray.

A red circle highlights the "Configure..." button in the "Annotate Restriction Sites" dropdown. A red arrow points from this button to a configuration dialog box that is open. The dialog box contains the following text and options:

The restriction site plugin generates a restriction map on the current view. This plugin was written Elizabeth Nickerson & Lincoln Stein.

Buttons: "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"/> Hpy188I	<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"/> Hpy99I	<input type="checkbox"/> PvuI
<input type="checkbox"/> AfeI	<input type="checkbox"/> BsrGI	<input type="checkbox"/> HpyCH4III	<input checked="" type="checkbox"/> PvuII
<input type="checkbox"/> AflI	<input type="checkbox"/> BssHII	<input type="checkbox"/> HpyCH4IV	<input type="checkbox"/> RsaI
<input type="checkbox"/> AflIII	<input type="checkbox"/> BssKI	<input type="checkbox"/> HpyCH4V	<input type="checkbox"/> RsrII
<input type="checkbox"/> AgeI	<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"/> AlwNI	<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"/> ApeI	<input type="checkbox"/> BstYI	<input type="checkbox"/> MseI	<input type="checkbox"/> Scal
<input type="checkbox"/> AscI	<input type="checkbox"/> BstZ17I	<input type="checkbox"/> MspI	<input type="checkbox"/> ScrFI
<input type="checkbox"/> AseI	<input type="checkbox"/> Bsu36I	<input type="checkbox"/> MspA1I	<input type="checkbox"/> SexAI