

7.1 Identification of specific DNA motifs.

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

- a. Find all BamHI restriction sites in all *Pythium ultimum* genomic sequences available in FungiDB. Note: you can use the DNA motif 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”.

The screenshot shows the FungiDB website interface. At the top, there is a search bar with "Gene ID: NCU05768" and "Gene Text Search: kinases". Below the search bar, there are navigation tabs: Home, New Search, My Strategies, My Basket (0), Tools, Data Summary, Downloads, and Community. The main content area is divided into three columns: "Identify Genes by:", "Identify Other Data Types:", and "Tools:". A red arrow points from the "Genomic Segments (DNA Motif)" option in the "Identify Other Data Types:" column to the "Tools:" column. Below the main content area, there is a section titled "Identify Genomic Segments based on DNA Motif Pattern". This section includes a tree view of organisms, with "Pythium ultimum" selected. The tree view shows the following structure:

- Organism select all | clear all | expand all | collapse all | reset to default
 - Oomycetes
 - Albugo
 - Aphanomyces
 - Hyaloperonospora
 - Phytophthora
 - Pythium
 - Pythium aphanidermatum
 - Pythium arrhenomanes
 - Pythium irregulare
 - Pythium iwayamai
 - Pythium ultimum
 - Pythium ultimum DAOM BR144
 - Pythium ultimum var. sporangiferum BR650
 - Pythium vexans
 - Saprolegnia
 - Fungi

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

- b. 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.

My Strategies: [New](#) [Opened \(1\)](#) [All \(5\)](#) [Basket](#) [Public Strategies \(1\)](#) [Help](#)

(Segments) Strategy: DNA Motif *

[DNA Motif](#) [Add Step](#)

18800 Segments from Step 1 Add 18800 Genomic Segments to Basket | Download 18800 Genomic Segments

Strategy: DNA Motif

Genomic Segment Results [Genomic Locations](#)

First 1 2 3 4 5 Next Last [Advanced Paging](#) [Add Columns](#)

Segment ID	Organism	Genomic Location	Motif
PuitDAOMBR144_SC1231:148-154.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1231: 148 - 154 (+)	...TGGGTGTCGTTTACGTGTGA GGATCC CTCAGGAGCGGCCATCCGC...
PuitDAOMBR144_SC1231:148-154.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1231: 148 - 154 (-)	...GCGGAATGGCCGCTCCTGAG GGATCC TACACGTAAACGACACCCA...
PuitDAOMBR144_SC1237:1395-1401.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 1395 - 1401 (+)	...GAGAAGAGGAGTGAGTCT GGATCC AACGGCGTGTTCAGATCGCC...
PuitDAOMBR144_SC1237:1395-1401.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 1395 - 1401 (-)	...GGCGATCTGAACACGCCGT GGATCC CAGACTCGACTCTTCTC...
PuitDAOMBR144_SC1237:2140-2146.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 2140 - 2146 (+)	...CAGACGGGTGCAGTGAGCT GGATCC AGGGCACCGGCATACCCGTC...
PuitDAOMBR144_SC1237:2140-2146.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 2140 - 2146 (-)	...GACGGGTATGCCGGTGCCT GGATCC CAGCTCACTGCACCCGTCGT...
PuitDAOMBR144_SC1237:2173-2179.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 2173 - 2179 (+)	...CGGCATACCCGTCAGGTA CTGGATCC AGTGTGCGTTTGCAGAAATT...
PuitDAOMBR144_SC1237:2173-2179.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 2173 - 2179 (-)	...AATCTTCGCAACGACCA CTGGATCC AGTACCTGACGGGTATGCC...
PuitDAOMBR144_SC1237:266-272.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 266 - 272 (+)	...TAGTTGCCTGGCTCGGCT CTGGATCC CTTCTGTAAAGAAAGCAG...
PuitDAOMBR144_SC1237:266-272.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 266 - 272 (-)	...TCGTCTCTTTTACAAGAA GGATCC AGAGCCGAGCCAGGCAACTA...
PuitDAOMBR144_SC1237:602-608.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 602 - 608 (+)	...CTATCGGGATCTCCCA CTGGATCC CCAAGTCTCCAGGAAGCGAA...
PuitDAOMBR144_SC1237:602-608.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1237: 602 - 608 (-)	...TTCGCTTCTGGAGCA CTGGATCC GAGATGGGAGATCCCGATAG...
PuitDAOMBR144_SC1239:100825-100831.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1239: 100825 - 100831 (+)	...TCTTTGCTTTTTCTTT GGATCC ACTTGCGGTCTTCAGAGCCT...
PuitDAOMBR144_SC1239:100825-100831.r	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1239: 100825 - 100831 (-)	...AGGCTCTGAAGACCGCA AGTGGATCC GCAAAAGAAAAGACGAAAGA...
PuitDAOMBR144_SC1239:102460-102466.f	Pythium ultimum DAOM BR144	PuitDAOMBR144_SC1239: 102460 - 102466 (+)	...GCCAAGGTGCGCTGT GGATCC GCATCGGCACACAGTTCTT...

7.2 Find genes that have one of these BamHI sites within 250 nucleotides upstream of their start.

In the section 7.1 you found BamHI sites, but now you are looking for genes that have one of these sites located within 250 nucleotides upstream of their start.

Hint: You can achieve this by running a genomic collocation search that defines 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.

Add Step 2 : Organism

Advanced Parameters

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

Continue...

Genomic Colocation

Combine Step 1 and Step 2 using relative locations in the genome
 You had 18800 Genomic Segments in your Strategy (Step 1). Your new Genes search (Step 2) returned 15986 Genes.

"Return each Gene from Step 2 whose upstream region overlaps the exact region of a Genomic Segment in Step 1 and is on either strand"

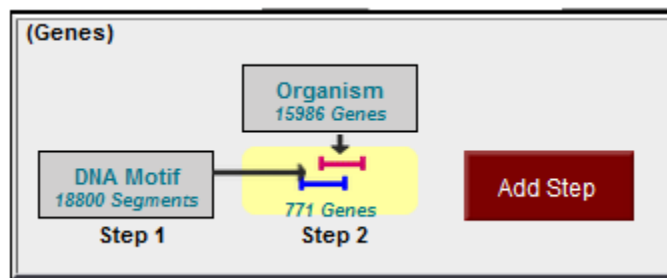
Submit

Close

How did you modify the location relative to genes?

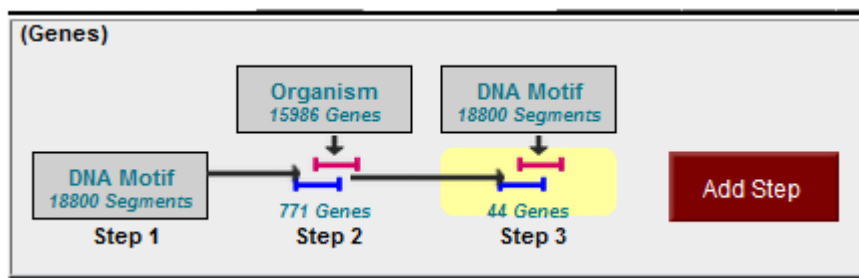


How many genes did you get?



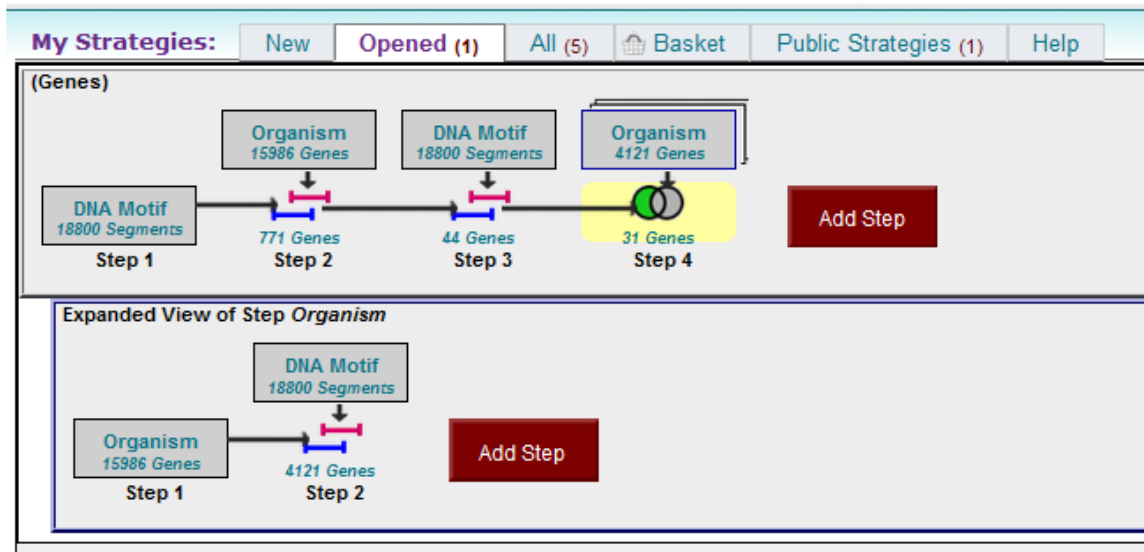
7.3 Using a similar sequence of steps as in part 7.2, define which of these genes also have a BamHI site in their 250 nucleotide downstream region.

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



7.4 Taking this a step further, define which of these genes do **NOT** contain a BamHI site within them.

Hint: you will have to use a nested strategy.



31 Genes from Step 4
Strategy: DNA Motif

Click on a number in this table to limit/filter your results

All Results	Ortholog Groups	<i>Ajellomyces</i>		<i>Allomyces</i>		<i>Aspergillus</i>					
		<i>A.capsulatus</i> (nr Genes: 0)		<i>A.macrogynus</i>		<i>A.aculeatus</i>	<i>A.carbonarius</i>	<i>A.clavatus</i>	<i>A.flavus</i>	<i>A.fumigatus</i>	<i>A.nidulans</i>
		G186AR	NAm1	ATCC 38327	ATCC 16872	ITEM 5010	NRRL 1	NRRL3357	Af293	FGSC A4	
31	31	0	0	0	0	0	0	0	0	0	0

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 a column to any result table that allows you to go directly to GBrowse at the genomic coordinates of any ID in your result list. Click on the Add Columns button.

The screenshot shows a web interface for gene results. At the top, there are tabs for 'Gene Results' and 'Genome View'. Below the tabs, there are navigation links: 'First 1 2 Next Last' and 'Advanced Paging'. On the right side, there is a button labeled 'Add Columns' which is circled in red. Below the navigation is a table with three columns: 'Gene ID', 'Genomic Location', and 'Product Description'. The table contains several rows of data, including gene IDs like Pytul_G007170 and genomic coordinates like PytulBR144_SC1242: 96,660 - 97,169 (-). A dialog box titled 'Select Columns' is open over the table. It has a close button (X) in the top right corner. At the top of the dialog, there is an 'Update Columns' button. Below that, there are links: 'clear all | expand all | collapse all' and 'reset to current | reset to default'. The dialog contains a list of categories with expand/collapse icons and checkboxes for various columns. The 'GBrowse' checkbox is highlighted with a red arrow. At the bottom of the dialog, there are more links: 'clear all | expand all | collapse all' and 'reset to current | reset to default', and an 'Update Columns' button.

Gene ID	Genomic Location	Product Description
Pytul_G007170	PytulBR144_SC1242: 96,660 - 97,169 (-)	Similar to TPRXL: Putative protein TPRXL (Homo sapiens);
Pytul_G007107	PytulBR144_SC1242: 242,551 - 242,829 (-)	Similar to TIM9: Mitochondrial import inner membrane translocase subunit Tim9 (Mesembryanthemum crys...
Pytul_G005845	PytulBR144_SC1402: 231,053 - 231,591 (+)	Protein of unknown function;
Pytul_G005579	PytulBR144_SC1402: 940,179 - 940,864 (-)	Similar to Tbc1d10c: Carabin (Mus musculu...
Pytul_G012872	PytulBR144_SC1722: 11,333 - 12,344 (+)	Protein of unknown function;
Pytul_G014631	PytulBR144_SC1811: 81,730 - 83,030 (+)	Protein of unknown function;
Pytul_G014158	PytulBR144_SC1895: 66,174 - 67,031 (+)	Similar to GARP: Glutamic acid-rich protein...
Pytul_G013849	PytulBR144_SC1980: 106,849 - 108,519 (+)	Protein of unknown function;
Pytul_G012238	PytulBR144_SC1991: 33,753 - 34,190 (+)	Similar to Cd36: Platelet glycoprotein 4 (Ra...
Pytul_G003273	PytulBR144_SC2006: 695,146 - 696,223 (+)	Similar to loc1: 60S ribosomal subunit asse...
Pytul_G002517	PytulBR144_SC2008: 374,121 - 376,923 (-)	Similar to 4CLL4: 4-coumarate-CoA ligase...
Pytul_G002216	PytulBR144_SC2008: 1,284,832 - 1,285,680 (-)	Similar to xecD: 2-(R)-hydroxypropyl-CoM c...
Pytul_G011671	PytulBR144_SC2014: 139,259 - 139,786 (+)	Protein of unknown function;
Pytul_G007849	PytulBR144_SC2020: 536,762 - 537,941 (-)	Similar to WDR77: Methylosome protein 50...
Pytul_G008131	PytulBR144_SC2022: 527,213 - 527,683 (+)	Similar to Rv2228c: Uncharacterized protei...
Pytul_G000785	PytulBR144_SC2023: 1,108,433 - 1,109,577 (-)	Protein of unknown function;
Pytul_G010215	PytulBR144_SC2026: 83,496 - 83,770 (-)	Protein of unknown function;

Note: you can configure restriction sites by clicking on the configure button in GBrowse and selecting the restriction sites you would like to display. 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 screenshot shows the GBrowse web interface. At the top, there are navigation tabs: **Browser**, **Select Tracks**, **Snapshots**, **Custom Tracks**, and **Preferences**. Below these, there is a search bar with the text "Pytu1BR144_SC1980:91,849..123,519". To the right, there are buttons for "Annotate Restriction Sites" (circled in red), "Configure...", "Save Snapshot", and "Load Snapshot". Below the buttons, there are scroll and zoom controls: "Scroll/Zoom: << < > >>" and "Show 31.67 kbp". The main part of the image is a genomic track for "Pytu1BR144_SC1980" showing a 31.67 kbp region. The track includes several data layers: "upload_5535", "Syntenic Sequences and Genes (Shaded by Orthology) (Showing 11 of 103 subtracks)", and "Annotated Genes (with UTRs in gray when available)". Genes shown include Pytu1_G013856, Pytu1_G013854, Pytu1_G013852, Pytu1_G013849, Pytu1_G013847, Pytu1_G013855, Pytu1_G013853, Pytu1_G013851, and Pytu1_G013850. At the bottom, there are "Select Tracks" and "Clear highlighting" buttons.

The screenshot shows the "Configure" dialog box for the restriction site plugin. The title is "The restriction site plugin generates a restriction map on the current view. This plugin was written Elizabeth Nickerson & Lincoln Stein." Below the title are "Cancel" and "Configure" buttons. The main section is "Select Restriction Sites To Annotate". There is a "Restriction Site Display" option with radio buttons for "off" and "on" (selected). Below this is a list of restriction enzymes with checkboxes:

- AatII
- Acc65I
- AccI
- AclI
- AfeI
- AflII
- AflIII
- AgeI
- AhdI
- AluI
- AlwNI
- ApaI
- ApaLI
- Apol
- AscI
- AseI
- BspDI
- BspEI
- BspHI
- BsrFI
- BsrGI
- BssHI
- BssKI
- BstAPI
- BstBI
- BstEII
- BstNI
- BstUI
- BstXI
- BstYI
- BstZ17I
- Bsu36I
- HpaII
- Hpy188I
- Hpy188III
- Hpy99I
- HpyCH4III
- HpyCH4IV
- HpyCH4V
- KpnI
- MboI
- MfeI
- MluI
- MscI
- MseI
- MslI
- MspA1I
- PspGI
- PspOMI
- PstI
- PvuI
- PvuII
- RsaI
- RsrII
- SacI
- SacII
- Sall
- Sau3AI
- Sau96I
- SbfI
- Scal
- ScrFI
- SexAI

The screenshot shows the GBrowse interface with the "Restriction Sites" data track selected. The track displays "BamHI restriction site" as vertical red lines. The genes Pytu1_G013849 and Pytu1_G013848 are highlighted with yellow and blue bars. The region shown is 5 kbp. The track also includes "Frame Usage", "Annotated Genes (with UTRs in gray when available)", "upload_5535", and "Syntenic Sequences and Genes (Shaded by Orthology) (Showing 11 of 103 subtracks)".