Welcome to

Introduction to Bioinformatics

Wednesday, 7 March

Sample Research Project (Part I)

It's difficult to introduce you to your research project, since each one will be different, and it is the differences that make them interesting. What I try to do here is to present an ***example*** of a research project, one that might be quite different from yours but at least will almost surely share with yours the elements of confusion and surprise.

As always, you'll gain the most from this tour if you follow along, doing what's shown with your own fingers.

Viral Genomes Project

Phage Analysis Groups

Each group is defined by a core focus, which comes with at least one suggested article. Group members may use this article as inspiration, a starting point to find other pertinent articles, and an aid in defining and subdividing the project. A group is under no obligation to stick to any predetermined agenda, so long as it goes in a direction that is even more interesting than the original.

See the group discusion forums on Blackboard for articles, and feel free to add to the list of articles and to the discussion.

A. **Lysogeny Group**
Do the new phages have protein inc
Can the sites of integration and rep

B. **Lysis Group**
What functions do the new phages
How is the moment of lysis control

C. **Sequence Bias Group**
Do the new genomes exhibit oligon
Do the phage sequence biases mat
What tRNAs do the phage have? Is

Our story begins…

I'm in the Mobile Element Group. The first step was for us collectively to decide what subtopic each of us would focus on. There was talk about transposases, introns, but I missed out on the discussion…

D. **Mobile Element Group**
Do the new phages have sequences that appear capable of transposition? Other mycobacteriophage?
What are the extents of the elements?

E. **DNA replication group**
What proteins do the new phages have to support the replication of phage DNA? Other mycobacteriophage?
What signals on the phage DNA serve to initiate DNA replication?

F. **Gene regulation group**
What proteins do the new phages use to regulate gene expression? Other mycobacteriophage?
To what DNA sites do these proteins bind?

Viral Genomes Project

I had a lot of things to do that week…

Biological interest of topic

3742–3756 *Nucleic Acids Research, 2001, Vol. 29, No. 18*

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

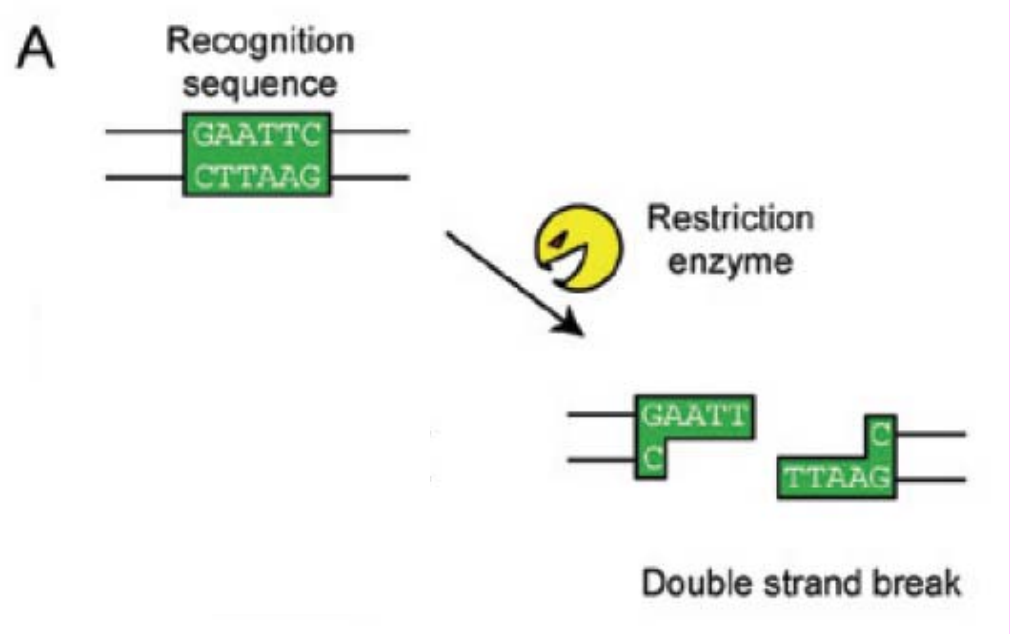
Ichizo Kobayashi*

…so I got what was left, something about restriction and modification.

Fortunately, I had a review article that defined the topic (I hoped).

Skimming through the article…

Biological interest of topic
Restriction-modification



SQ1. Would the enzyme shown in the figure be expected to cut the sequence

CTTAAG
GAATTC

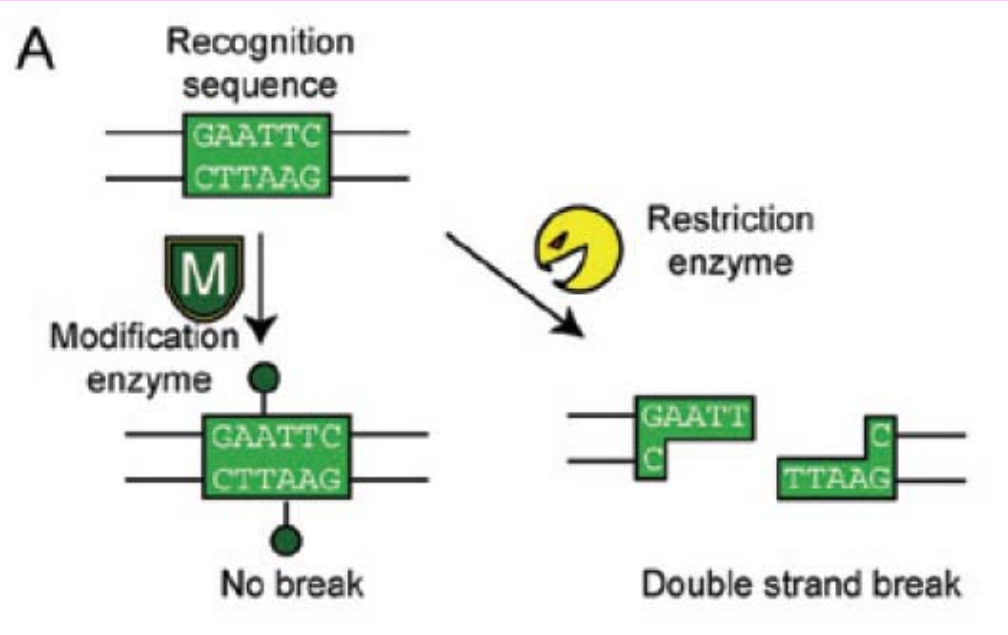
Kobayashi I (2001). Nucl Acids Res 29:3742-3756

I found this figure.

I learned that restriction enzymes recognize specific sequences on double-stranded DNA. (Hah! It's a palindrome!)

The enzyme cuts both strands of the DNA. Sounds suicidal…

Biological interest of topic
Restriction-modification

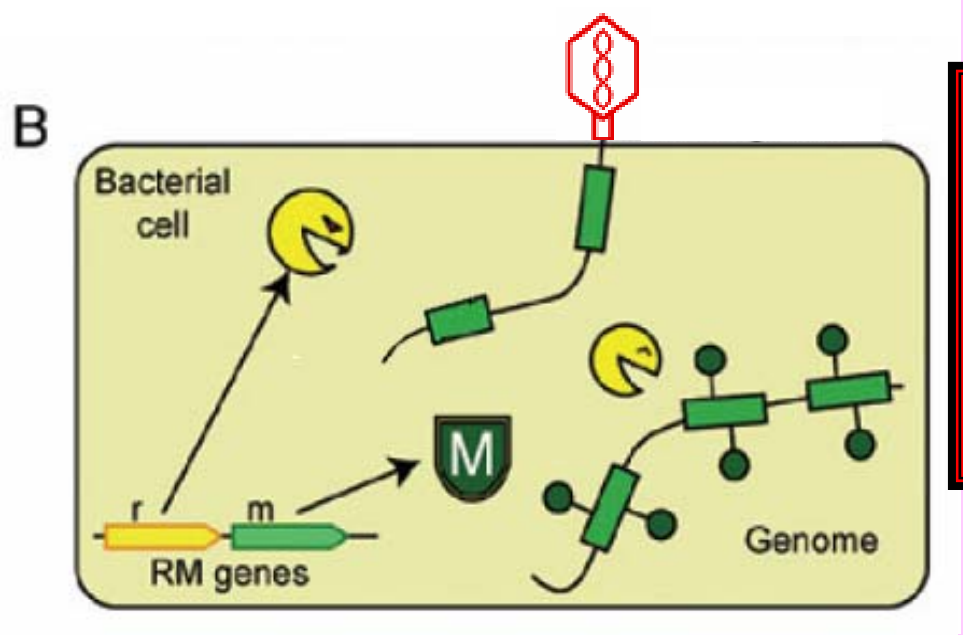


Kobayashi I (2001). Nucl Acids Res 29:3742-3756

The cell avoids suicide by the action of a second enzyme, one that modifies the cell's DNA so that it is impervious to the effects of the restriction enzyme.

If the net effect is no cutting, then why do restriction enzymes exist?

Biological interest of topic
Restriction-modification

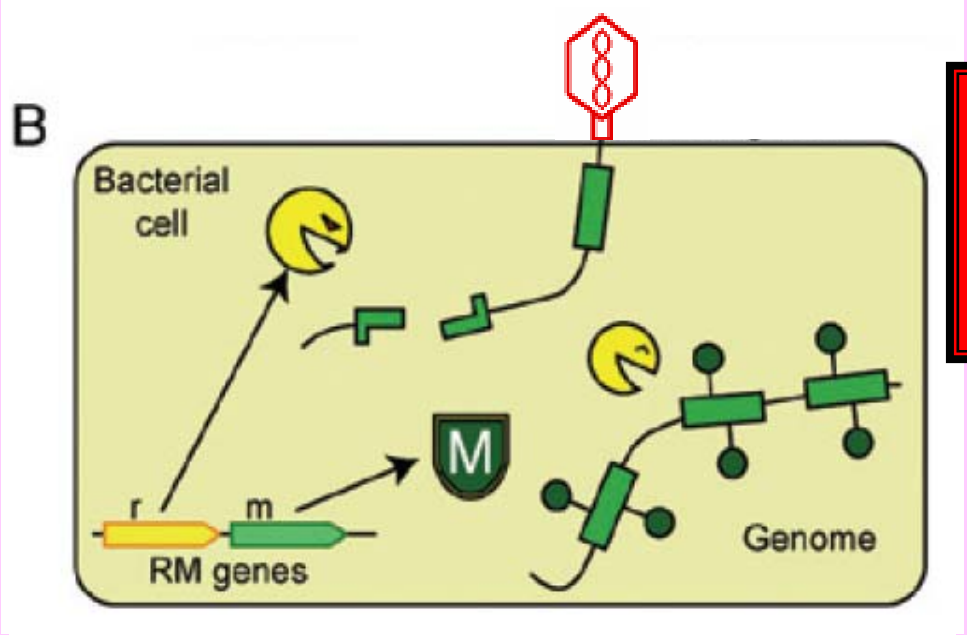


SQ2. Would you expect to find the restriction recognition sequence GAATTC in the genome of a typical phage?

Kobayashi I (2001). Nucl Acids Res 29:3742-3756

Not all DNA is modified. If a phage injects it's DNA into a cell, it's attacked by restriction enzymes, before the modification enzymes have time to modify it…

Biological interest of topic
Restriction-modification



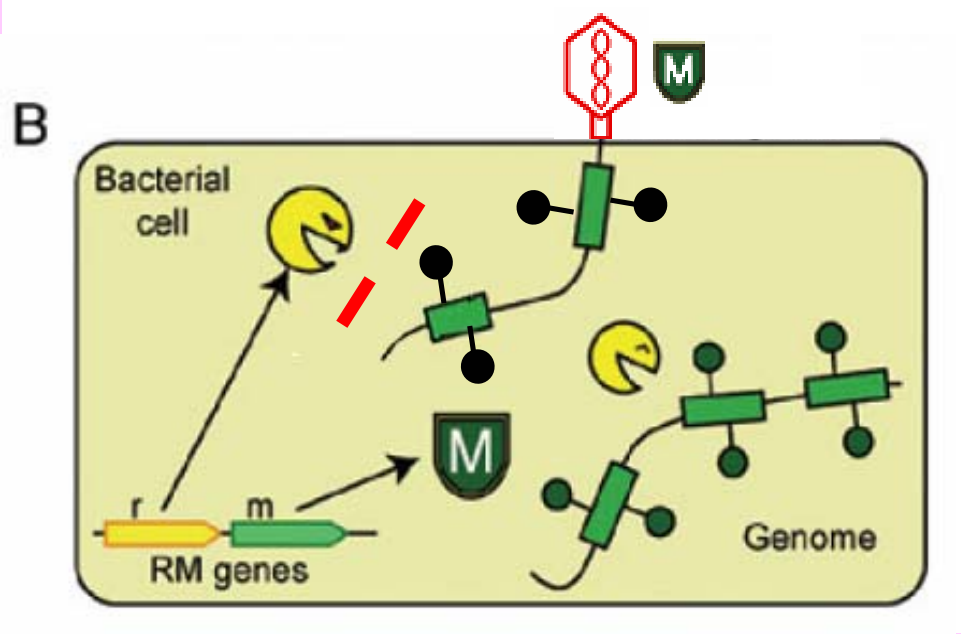
SQ3. What defense can you imagine that phages could employ against restriction?

Kobayashi I (2001). Nucl Acids Res 29:3742-3756

…and the phage is chopped to pieces. With DNA fragmented, lytic proteins can't be made, and the infection is stopped. But the cell DNA remains safe through modification.

How then do phage persist – nay, thrive! -- if there's this potent defense against them?

Biological interest of topic
Restriction-modification

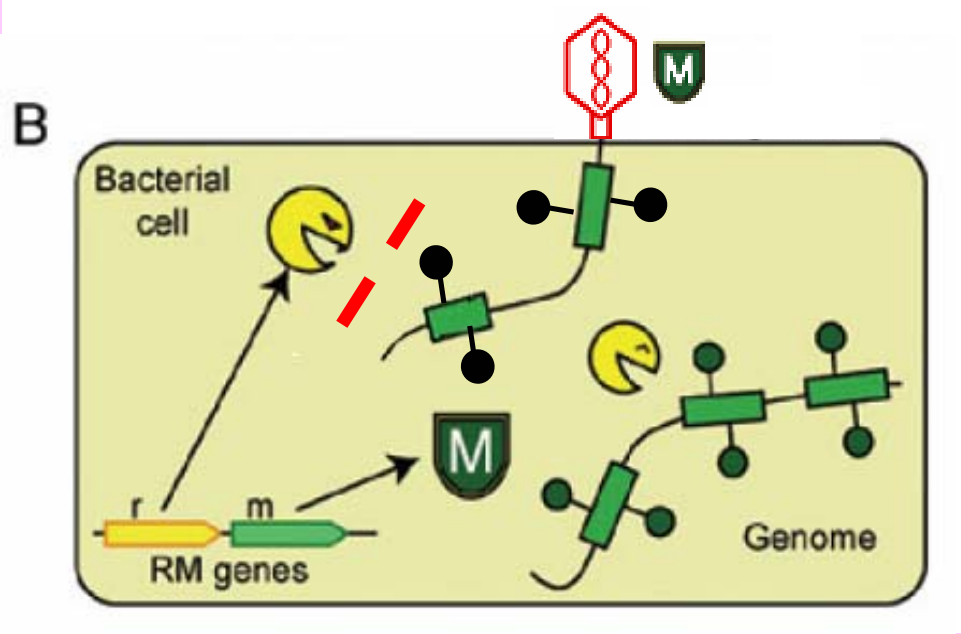


SQ4. What other defenses can you imagine?

Kobayashi I (2001). Nucl Acids Res 29:3742-3756

Phages also have their tricks. Lots of them. One defense is to encode a modification gene themselves. Then they're not affected by the corresponding restriction enzyme.

Biological interest of topic
Restriction-modification



Kobayashi I (2001). Nucl Acids Res 29:3742-3756

This is a valuable enzyme! You can see why it might spread amongst phage if there were a mechanism by which that could happen.

*Is there any reason to believe that there **is** such a mechanism? Maybe the article has an opinion on this…*

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

Further along in the article I saw this section, which certainly sounded like it has to do with mobility. Restriction-modification genes have been transferred horizontally? What's that? …and "lateral gene transfer"?

(A trip to the net…)

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

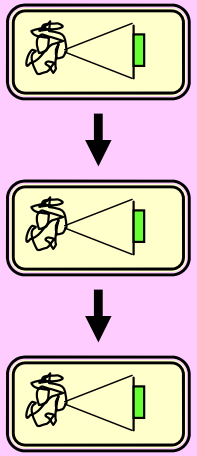
Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

…OK, now I saw that the two terms (horizontal/lateral gene transfer) are synonyms, and both stand in contrast to the more conventional vertical gene transfer.

In vertical gene transfer, genes are passed from parent cell to daughter cell through replication/division.



Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

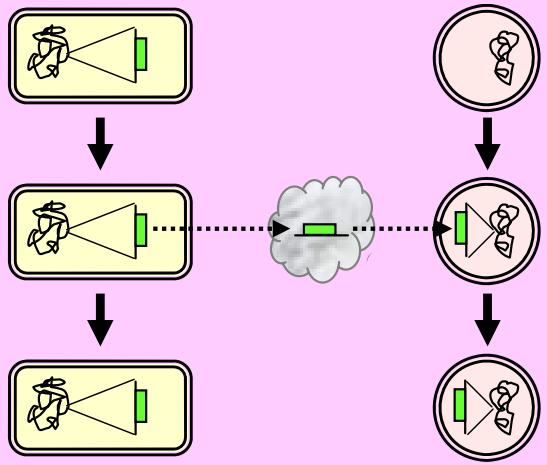
Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

In horizontal gene transfer, one or more genes (but not the whole genome) is transferred to a wholly different cell by one of several mechanisms, and there the gene may become part of the genome.

If RM genes are somehow more prone to this than others,...



SQ5. You know of one of the several mechanisms available to move DNA from one organism to another,… what is it?

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

SQ6. What kinds of information could we seek that might answer each of these questions?

…but are they more prone to move?

A germ of a project was beginning to form in my mind:

1. *Is self modification of DNA a common strategy employed by phages?*
2. *Have the genes for modification moved amongst phages by horizontal gene transfer?*
3. *If they have, what enables them to do so?*

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

One of the main benefits of a review article is to point you to research articles that tell you what the truth is regarding some point of interest, and the truth is always more complicated (and more interesting!) than the bland generalities that can fit into review articles.

So what were those research articles that talk about the mobility of RM genes? References 29…36…37,… now a trip to the end of the article.

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, **397**, 176–180.

29. Nobusato,A., Uchiyama,I. and Kobayashi,I. (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene*,

29, Nobusato et al… specific to Helicobacter… not sure about this.

of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl Acad. Sci. USA*, **97**, 9671–9676.

32. Kong,H., Lin,L.F., Porter,N., Stickel,S., Byrd,D., Posfai,J. and Roberts,R.J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.*, **28**, 3216–3223.

33. Lin,L.F., Posfai,J., Roberts,R.J. and Kong,H. (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*.

36, Jeltsch et al… Interesting if type-II restriction endonucleases are important.

69, 1816–1820.

36. Jeltsch,A., Kroger,M. and Pingoud,A. (1995) Evidence for an evolutionary relationship among type-II restriction endonucleases. *Gene*, **160**, 7–16.

37. Bujnicki,J.M. and Radlinska,M. (1999) Molecular phylogenetics of DNA 5mC-methyltransferases. *Acta. Microbiol. Pol.*, **48**, 19–30.

38. Jeltsch,A. and Pingoud,A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.*, **42**, 91–96.

Biological interest of topic
Restriction-modification

3742–3756 Nucleic Acids Research, 2001, Vol. 29, No. 18

SURVEY AND SUMMARY

Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution

Ichizo Kobayashi*

Evolutionary analyses suggest RM genes have undergone horizontal gene transfer

Comparisons of RM sequence alignments (often in the form of a phylogenetic tree) with sequence alignments of other genes, such as ribosomal RNA genes, in the genomes suggest that RM genes have undergone extensive lateral gene transfer (29,36,37). The GC content and/or codon usage of RM genes are often different from those of other genes in the genome (28,29,38,39). This is consistent with the notion that these RM genes joined the genome relatively recently, although it is difficult to estimate the time of their arrival.

How do I decide what's important?

Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, **397**, 176–180.

29. Nobusato,A., Uchiyama,I. and Kobayashi,I. (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene*, **259**, 89–98.
30. Akopyants,N.S., Fradkov,A., Diatchenko,L., Hill,J.E., Siebert,P.D., Lukyanov,S.A., Sverdlov,E.D. and Berg,D.E. (1998) PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA*, **95**, 13108–13113.
31. Xu,Q., Morgan,R.D., Roberts,R.J. and Blaser,M.J. (2000) Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proc. Natl Acad. Sci. USA*, **97**, 9671–9676.
32. Kong,H., Lin,L.F., Porter,N., Stickel,S., Byrd,D., Posfai,J. and Roberts,R.J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.*, **28**, 3216–3223.
33. Lin,L.F., Posfai,J., Roberts,R.J. and Kong,H. (2001) Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc. Natl Acad. Sci. USA*, **98**, 2740–2745.
34. Chinen,A., Uchiyama,I. and Kobayashi,I. (2000) Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome

37, Bujnicki & Radlinska…ditto, if 5mC-methyl-transferases are important…

160, 7–16.

37. Bujnicki,J.M. and Radlinska,M. (1999) Molecular phylogenetics of DNA 5mC-methyltransferases. *Acta. Microbiol. Pol.*, **48**, 19–30.
38. Jeltsch,A. and Pingoud,A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.*, **42**, 91–96.

BNFO 301

Introduction to Bioinformatics

Spring 2012

January February **March** April May

Monday	Wednesday
5: Genome Analysis **Article** - Gomathi et al (2007) + tour Questionnaire **Group meetings:** Organize approach to topic **Lab**: Intro to BioBIKE and Statistics **Problem Set 5** - Blast and Dotplot ***OR*** **Tour** - Was Mendel Right? (Part I and Part II) **Problem Set 6** - Statistics	7: Genome Analysis **Article** - Gomathi et al (2007) + tour **Research Project** - What to do? Questionnaire
14: *SPRING BREAK*	16: *SPRING BREAK*
19: Genome Analysis **Article** - Karlin (2001) + tour Questionnaire **Lab**: Problem Set 7: Genome analysis ***Please e-mail rough draft of article summary***	21: Genome Analysis (focus on phylogeny) **Skim**: Howe et al (2001). Trends in Genetics 17:147-152 *Manuscript Evolution* **Read**: Baldauf SL (2003). Trends in Genetics 19:345-351 *Phylogeny for the faint of heart* Questionnaire Problem Set 8: Research Project ***Please e-mail responses to Problem Set 7 by end of Friday***
26: Genome Analysis (catch up) **Article** - Karlin (2001) + tour (same as Mar 21) **Skim**: Howe et al (2001). Trends in Genetics 17:147-152 *Manuscript Evolution* **Read**: Baldauf SL (2003). Trends in Genetics 19:345-351 *Phylogeny for the faint of heart* Questionnaire **Lab**: Problem Set 7 and others	28: **General problem session** Questionnaire - What questions to discuss? **Review Session**: 10:00-10:55 Life Sciences Room 250 ***Exam 2 distributed*** What to expect Ground rules

A call for help… a hand extended?

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I. Typical research project focus

There are hardly any limits on a question that can serve as the focus of your research project except one – ***find something interesting*** – and that doesn't exclude much. However, I can offer a strategy of finding something interesting that has proven to be broadly applicable to a variety of topics within research groups. Understand that this is just one strategy. You might well find something better for yourself.

I.A. Identify a specific protein that has an important role within the broad focus of your group

Many people, given the task of carving out something they can call their own within the broad focus of a group, tend to cast their eyes towards general functions. For example, a person in the DNA replication group might think of nucleotide synthesis. This is certainly required for DNA

Let me take a look at this.

As it says, what's most important is that my project is interesting, and maybe the most interesting direction will have nothing to do with what this document suggests.

But let's play a long for a while to see where it goes…

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.A. Identify a specific protein that has an important role within the broad focus of your group

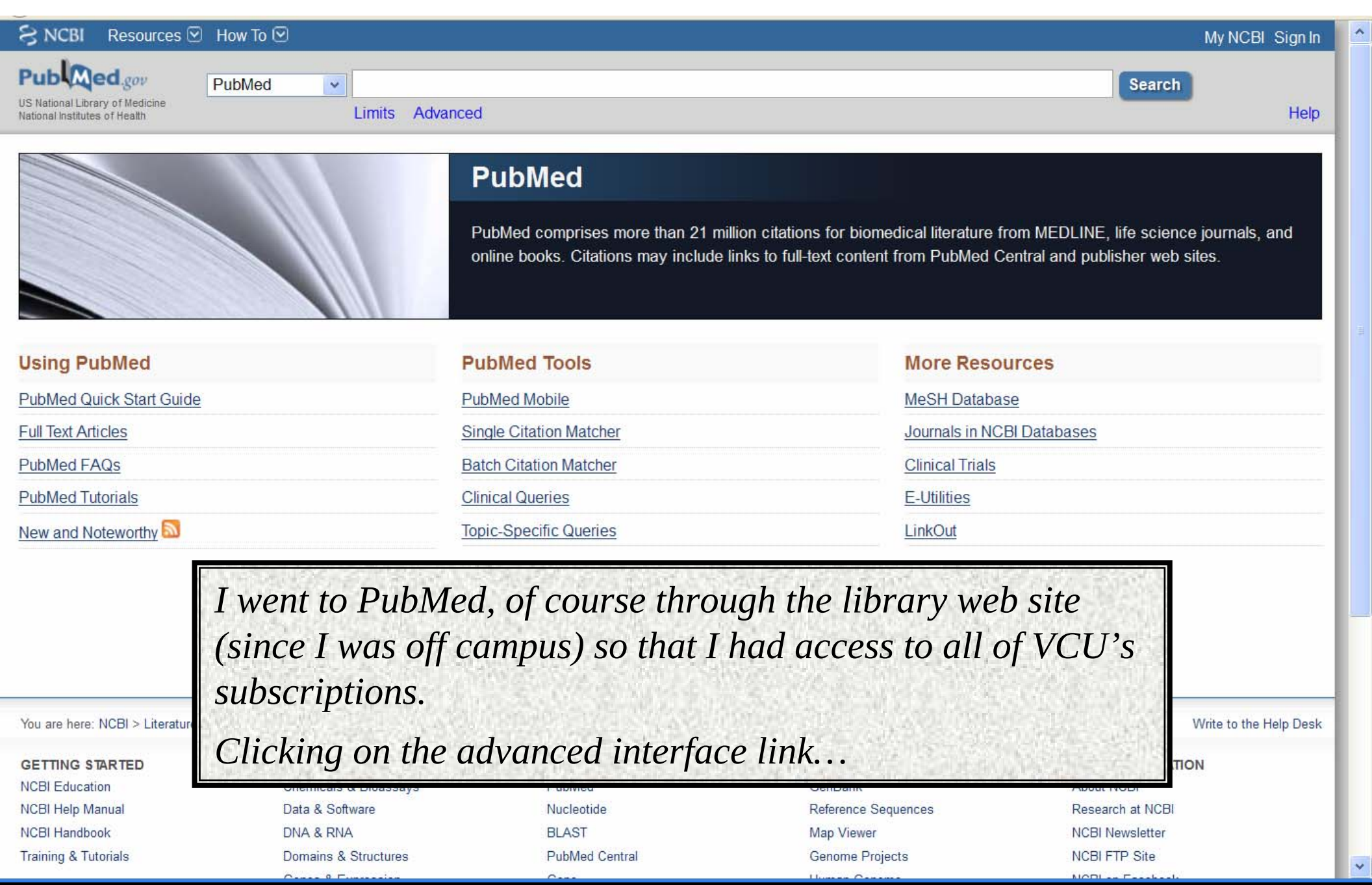
Many people, given the task of carving out something they can call their own within the broad focus of a group, tend to cast their eyes towards general functions. For example, a person in the DNA replication group might think of nucleotide synthesis. This is certainly required for DNA replication, but by itself, it doesn't give you a handle on what to look for in a phage genome. Genomes contain genes, and genes encode proteins. Much better to identify a specific sort of protein.

How to find such a protein? Review articles can provide an overview of proteins that are involved in a specific process. You're looking particularly for a protein that has conserved sequence motifs, something that you can search for in the phage proteins and thereby identify new instances of this type of protein. If you're lucky, the protein you adopt as your focus will have well conserve nces through their possession of simila nds like a specific protein to you exists similarity with one another. In order to d use their features for predictive purpos

SQ7. We're talking about amino acid sequence motifs, but what about DNA sequence motifs. You've come across a couple. Do you know of any?

I should identify a specific protein? From a review article? But my review article doesn't mention any specific proteins, just "restriction-modification"

I figured I needed to look for a different review article…


NCBI Resources How To
My NCBI Sign In
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
Search
Limits Advanced
Help
PubMed
PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.
Using PubMed
PubMed Quick Start Guide
Full Text Articles
PubMed FAQs
PubMed Tutorials
New and Noteworthy
PubMed Tools
PubMed Mobile
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Topic-Specific Queries
More Resources
MeSH Database
Journals in NCBI Databases
Clinical Trials
E-Utilities
LinkOut
I went to PubMed, of course through the library web site (since I was off campus) so that I had access to all of VCU's subscriptions.
Clicking on the advanced interface link…
You are here: NCBI > Literatur
Write to the Help Desk
GETTING STARTED
NCBI Education
NCBI Help Manual
NCBI Handbook
Training & Tutorials
Data & Software
DNA & RNA
Domains & Structures
Nucleotide
BLAST
PubMed Central
Reference Sequences
Map Viewer
Genome Projects
Research at NCBI
NCBI Newsletter
NCBI FTP Site

NCBI Resources How To My NCBI Sign In

PubMed Home More Resources Help

PubMed Advanced Search Builder

YouTube Tutorial

History deleted.

(restriction-modification) AND review[Publication Type]

Edit Clear

Builder

	All Fields	restriction-modification	Show index list
AND	Publication Type	review	Show index list
AND	All Fields		Show index list

Search or Add to history

History Clear history

There is no recent history

I decided to use 'restriction-modification' as the search term, and I confined the search to review articles using the Publication Type field.

Rather than go directly to the references, I prefer to see how many hits I get first, so I clicked Add to history.

PubMed Advanced Search Builder

YouTube Tutorial

Use the builder below to create your search

Edit Clear

Builder

All Fields — Show index list

AND All Fields — Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:22:38

78,… seems like a lot, but I wasn't plannig on reading them, just skimming the titles. So I gave it a try, by clicking the number.

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed (restriction-modification) AND review[Publication Type] Search

RSS Save search Limits Advanced Help

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to:

Results: 1 to 20 of 78

<< First < Prev Page 1 of 4 Next > Last >>

1. Translocation, switching and gating: potential roles for ATP in long-range communication on DNA by Type III restriction endonucleases.
   Szczelkun MD.
   Biochem Soc Trans. 2011 Apr;39(2):589-94. **Review.**
   PMID: 21428945 [PubMed - indexed for MEDLINE]
   Related citations

2. Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems.
   Ishikawa K, Fukuda E, Kobayashi I.
   DNA Res. 2010 Dec;17(6):325-42. Epub 2010 Nov 8.
   PMID: 21059708 [PubMed - indexed for MEDLINE]
   Related citations

3. **Restriction-modification** systems may be associated with Helicobacter pylori virulence.
   Ando T, Ishiguro K, Watanabe O, Miyake N, Kato T, Hibi S, Mimura S, Nakamura M, Miyahara R, Ohmiya N, Niwa Y, Goto H.
   J Gastroenterol Hepatol. 2010 May;25 Suppl 1:S95-8. **Review.**
   PMID: 20586875 [PubMed - indexed for MEDLINE]
   Related citations

4. Bacteriophage host range and bacterial resistance.
   Hyman P, Abedon ST.
   Adv Appl Microbiol. 2010;70:217-48. Epub 2010 Mar 6. **Review.**
   PMID: 20359459 [PubMed - indexed for MEDLINE]
   Related citations

5. Bacteriophage resistance mechanisms.
   Labrie SJ, Samson JE, Moineau S.
   Nat Rev Microbiol. 2010 May;8(5):317-27. Epub 2010
   PMID: 20348932 [PubMed - indexed for MEDLINE]
   Related citations

6. The type IIB restriction endonucleases.
   Marshall JJ, Halford SE.
   Biochem Soc Trans. 2010 Apr;38(2):410-6. **Review.**
   PMID: 20298193 [PubMed - indexed for MEDLINE]

Didn't sound like it was going to tell me what proteins are involved in RM.

Same, in fact they all sounded too specific.

Although this one sounded like it could tell me more about the biology.

Filter your results:

All (78)
Free Full Text (16)
Review (78)
Manage Filters

elated data
se: Select
Find items

h details
riction-modification[All
s] AND review[Publication
Search
See more...

Recent activity

Turn Off Clear

Restriction and modification systems. PubMed

(restriction-modification) AND review[Publication Type] (78) PubMed

tosine) AND ((phage) AND ethyltransferase) (205) PubMed

obile) AND ((((methyltransferase) AND otifs) AND alignment) AN... (2) PubMed

methyltransferase) AND motifs) AND alignment) AND bacteria (50) PubMed

See more...

54. Ito T.
Tanpakushitsu Kakusan Koso. 1993 Feb;38(3):541-50. **Review**. Japanese. No abstract available.
PMID: 8488288 [PubMed - indexed for MEDLINE]
Related citations

[Chemical cleavage of long chain DNA--molecular design for artificial restriction enzyme].
55. Sugiura Y.
Tanpakushitsu Kakusan Koso. 1993 Feb;38(3):516-23. **Review**. Japanese. No abstract available.
PMID: 8488285 [PubMed - indexed for MEDLINE]
Related citations

Genetics of Paracoccus denitrificans.
56. Steinrücke P, Ludwig B.
FEMS Microbiol Rev. 1993 Jan;10(1-2):83-117. **Review**.
PMID: 8431311 [PubMed - indexed for MEDLINE]
Related citations

On the origins, structures and functions of **restriction-modification** enzymes.
57. Heitman J.
Genet Eng (N Y). 1993;15:57-108. **Review**. No abstract available.
PMID: 7764063 [PubMed - indexed for MEDLINE]
Related citations

DNA triple-helix formation: an approach to artificial gene repressors?
58. Maher LJ 3rd.
Bioessays. 1992 Dec;14(12):807-15. **Review**.
PMID: 1365896 [PubMed - indexed for MEDLINE]
Related citations

Organization of **restriction-modification** systems.
59. Wilson GG.
Nucleic Acids Res. 1991 May 25;19(10):2539-66. **Review**.
PMID: 2041731 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

Restriction and modification systems.
60. Wilson GG, Murray NE.
Annu Rev Genet. 1991;25:585-627. **Review**. No abstract available.
PMID: 1812816 [PubMed - indexed for MEDLINE]
Related citations

Near the end of the list, I finally found some review articles (from 20 years ago) that sounded like they might describe RM proteins.

I suppose that's reasonable: there's more need for a review when a field of inquiry is new.

I decided to try the first, a free PubMed Central article, since that one will surely have full text.

<< First | < Prev | Page 3 of 4 | Next > | Last >>

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to:

Nucleic Acids Research, Vol. 19, No. 10 2539

Organization of restriction-modification systems

Geoffrey G.Wilson

Skimming through the article, I came to a section that described restriction enzymes and modification enzymes.

Hey, enzymes are proteins! This looked like it was just what I needed.

Start with restriction enzymes…

…uh oh. They're described as dissimilar from each other. Maybe I should move on.

You're looking particularly for a protein that has conserved sequence motifs…

Amino acid sequence comparisons

Restriction enzymes vs. modification enzymes. Over fifty type II R-M systems have been sequenced. No similarities have been seen between endonucleases and methyltransferases (49). Some similarities might be expected between companion enzymes since they recognize identical DNA sequences. The lack of similarity suggests that restriction and modification enzymes are unrelated, and that they recognize their targets by different strategies.

Restriction enzymes. Fifty four type II endonucleases have been sequenced. Apart from certain pairs of isoschizomers, the enzymes are dissimilar (49). This suggests that they arose independently during evolution, and not from a common ancestor by divergence of its target recognition domain (TRD). Isoschizomers that cleave the same sequence at the same position ('homoschizomers'?) are sometimes exceptions: *Eco*RI and *Rsr*I (G′AATTC) are closely similar, and probably diverged from a common ancestor (50). Not all homoschizomers are homologous, however: *Hae*III and *Ngo*PII (GG′CC) are entirely dissimilar (51,128). Isoschizomers that cleave the same sequence at different positions ('heteroschizomers'? (7)), for example *Sma*I (CCC′GGG) and *Xma*I (C′CCGGG), are also dissimilar (53,54). Since no common sequence motifs have been discerned among endonucleases, they cannot be recognized as such by inspection of their amino acid sequences.

Nucleic Acids Research, Vol. 19, No. 10 2539

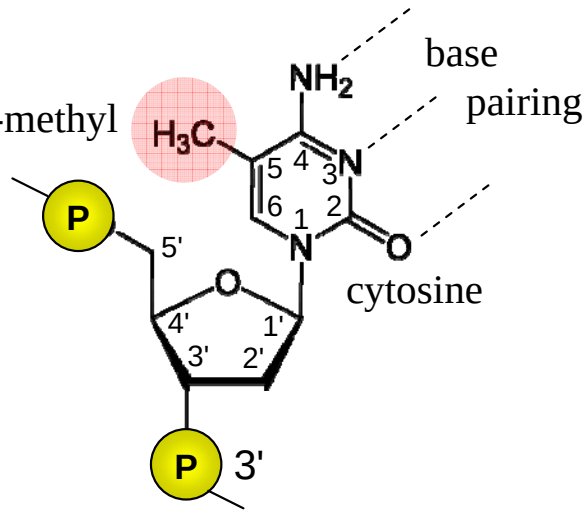
Organization of restriction-modification systems

Geoffrey G.Wilson

Much better!

Evidently modification enzymes are better conserved. And I learned that they're called methyltransferases.

And that there are different types. The first one described puts methyl groups on cytosines at position 5.



Modification enzymes. In contrast to the endonucleases, extensive similarities occur among the methyltransferases. Approximately ninety have been sequenced, and seven, or so, architectural classes have been distinguished (55). One class comprises enzymes that form 5-methylcytosine in DNA (m^5C-MTases). Members of this group possess ten, or so, common aa sequence motifs (56). Towards the CO_2H-terminus of these enzymes is a 'variable region' that is believed to form the TRD (57–59). The remaining classes comprise enzymes that form *N*4-methylcytosine (m^4C-MTases), and *N*6-methyladenine (m^6A-MTases). The m^4C-MTases and m^6A-MTases are quite similar, suggesting a common mechanism for methylating the exocyclic amino group of adenine and cytosine (55,60). The enzymes share two principal common sequence motifs. Surprisingly, the order of the motifs differs between certain of the classes (55,61).

Nucleic Acids Research, Vol. 19, No. 10 2539

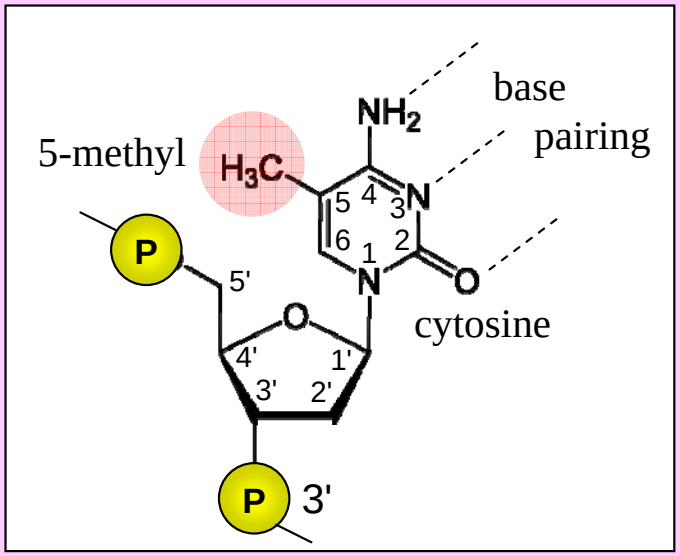
Organization of restriction-modification systems

Geoffrey G.Wilson

These cytosine methyltransferases turn out to have the most similarity amongst themselves, so I chose tentatively to start with them.

I modified the questions I had asked earlier…

Modification enzymes. In contrast to the endonucleases, extensive similarities occur among the methyltransferases. Approximately ninety have been sequenced, and seven, or so, architectural classes have been distinguished (55). One class comprises enzymes that form 5-methylcytosine in DNA (m^5C-MTases). Members of this group possess ten, or so, common aa sequence motifs (56). Towards the CO_2H-terminus of these enzymes is a 'variable region' that is believed to form the TRD (57–59). The remaining classes comprise enzymes that form *N*4-methylcytosine (m^4C-MTases), and *N*6-methyladenine (m^6A-MTases). The m^4C-MTases and m^6A-MTases are quite similar, suggesting a common mechanism for methylating the exocyclic amino group of adenine and cytosine (55,60). The enzymes share two principal common sequence motifs. Surprisingly, the order of the motifs differs between certain of the classes (55,61).



Nucleic Acids Research, Vol. 19, No. 10 2539

Organization of restriction-modification systems

Geoffrey G.Wilson

1. *Is modification of DNA by cytosine methyltransferases a common strategy employed by phages?*
2. *Have the genes for cytosine methyltransferases moved amongst phages by horizontal gene transfer?*
3. *If they have, what enables them to do so?*

Maybe I can find an article about them that will help…

Modification enzymes. In contrast to the endonucleases, extensive similarities occur among the methyltransferases. Approximately ninety have been sequenced, and seven, or so, architectural classes have been distinguished (55). One class comprises enzymes that form 5-methylcytosine in DNA (m^5C-MTases). Members of this group possess ten, or so, common aa sequence motifs (56). Towards the CO_2H-terminus of these enzymes is a 'variable region' that is believed to form the TRD (57–59). The remaining classes comprise enzymes that form *N*4-methylcytosine (m^4C-MTases), and *N*6-methyladenine (m^6A-MTases). The m^4C-MTases and m^6A-MTases are quite similar, suggesting a common mechanism for methylating the exocyclic amino group of adenine and cytosine (55,60). The enzymes share two principal common sequence motifs. Surprisingly, the order of the motifs differs between certain of the classes (55,61).

SQ8. Do you have a bit clearer idea of how these questions might be answerable?

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.B. Become an expert on some small slice of the project

It takes years to understand the complexities of any interesting biological problem. But to know what there is to understand -- surprisingly, that takes hardly any time at all! You can collect almost everything that has ever been published on a suitably constrained topic just through a trip to PubMed plus a bit of noodling, and your collection will be no less exhaustive (and probably more so) than that of the world's leading expert on the topic.

The references you find through this exercise can be quite valuable, even if you have not read a single one of the articles you've found. The list tells you what questions have been asked and where to find the answers that have been obtained. It may give you a sense that you have at least drawn a frame around what might otherwise seem a formless topic.

I was going to find an article about cytosine methyltransferases,… now I'm advised to find every article that's ever been written about them???

That sounds like it could be a lot. I hoped to cut down the number by confining the list to those articles focusing on bacteriophages.

Back to PubMed…

PubMed Advanced Search Builder

YouTube Tutorial

((phage) AND methyltransferase) AND cytosine

Edit Clear

Builder

	All Fields	phage	Show index list
AND	All Fields	methyltransferase	Show index list
AND	All Fields	cytosine	Show index list
AND	All Fields		Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:23:44

I hoped that limiting the search to these three terms wouldn't give me too much, maybe one or two dozen pertinent articles.

NCBI Resources How To My NCBI Sign In

PubMed Home More Resources Help

PubMed Advanced Search Builder

YouTube Tutorial

Use the builder below to create your search

Edit Clear

Builder

All Fields — Show index list

AND All Fields — Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#3	Add	Search **((methyltransferase) AND phage) AND cytosine**	205	03:53:34
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	7	03:23:44

Oof! 205 articles! Are there really that many articles related to cytosine methyltransferases in phages?

If so, then I would need to constrain my search further.

If not, then maybe my search strategy wasn't good enough.

First step was to see what I got…

You are here: NCBI > Literature > P

Write to the Help Desk

GETTING STARTED
NCBI Education
NCBI Help Manual
NCBI Handbook
Training & Tutorials

FORMATION
CBI
h at NCBI
wsletter
P Site
Facebook
Twitter
YouTube

Literature Genome Sequence Read Archive
Proteins SNP
Sequence Analysis Structure

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ((methyltransferase) AND phage) AND cytosine Search

RSS Save search Limits Advanced Help

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to:

Results: 1 to 20 of 205 Selected: 1 << First < Prev Page 1 of 11 Next > Last >>

Wolbachia prophage DNA adenine **methyltransferase** genes in different Drosophila-Wolbachia associations.
1. Saridaki A
PLoS One.
PMID: 2157
Related cita

Adenine! Wrong methyltransferase.

Recombinant DNA-**methyltransferase** M1.BspACI from Bacillus psychrodurans AC: purification and properties.
2. Tarasova
Biochemistr
PMID: 2131
Related cita

Didn't sound like a phage.

Identification of prophage gene z2389 in Escherichia coli EDL933 encoding a DNA **cytosine methyltransferase** for full protection of NotI sites.
3. Chiou CS, Li HY, Tung SK, Chen CY, Teng CH, Shu JC, Tseng IT, Hsu CY, Chen CC
Int J Med Mi
PMID: 2002
Related cita

This one sounded good! Check it.

Diversity and evolution of chromatin proteins encoded by DNA viruses.
4. de Souza RF, Iyer LM, Aravind L.
Biochim Biophys Acta. 2010 Mar-Apr;1799(3-4):302-18. Epub 2009 Oct 28. Review.
PMID: 19878744 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

Real-time kinetics of restriction-modification gene expression after entry into a new host cell.
5. Mruk I, Blumenthal RM.
Nucleic Acids Res. 2008 May;36(8):2581-93. Epub 2008 Mar 11.
PMID: 18334533 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

A view of an elemental naturalist at the DNA world (base composition, sequences, methylation).
6. Vanyushin BF.
Biochemistry (Mosc). 2007 Dec;72(12):1289-98.
PMID: 18205613 [PubMed - indexed for MEDLINE] Free Article
Related citations

Filter your results:
All (205)
Free Full Text (95)
Review (13)
Manage Filters

Find related data
Database: Select
Find items

Search details
(("methyltransferases"[MeSH Terms] OR "methyltransferases"[All Fields] OR "methyltransferase"[All Fields]) AND
Search See more...

Recent activity
Turn Off Clear
((methyltransferase) AND phage) AND cytosine (205) PubMed
Organization of restriction-modification systems. PMC
Organization of restriction-modification systems. PubMed
Restriction and modification systems. PubMed
(restriction-modification) AND review[Publication Type] (78) PubMed
See more...

14. Zinoviev VV, Yakishchik SI, Evdokimov AA, Malygin EG, Hattman S.
Nucleic Acids Res. 2004 Jul 27;32(13):3930-4. Print 2004.
PMID: 15280508 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

Bacteriophage T2Dam and T4Dam DNA-[N6-adenine]-**methyltransferases**.

15. Hattman S, Malygin EG.
Prog Nucleic Acid Res Mol Biol. 2004;77:67-126. Review. No abstract available.
PMID: 15196891 [PubMed - indexed for MEDLINE]
Related citations

A column method for determination of DNA **cytosine**-C5-**methyltransferase** activity.

16. Kim BY, Kwon OS, Joo SA, Park JA, Heo KY, Kim MS, Ahn JS.
Anal Biochem
PMID: 1476
Related cita

A biochemical method… I excluded this one.

VO1, a temperate bacteriophage of the type 19A multiresistant epidemic 8249 strain of Streptococcus pneumoniae: analysis of variability of lytic and putative C5 **methyltransferase** genes.

1. Obregón V, García P, López R, García JL.
Microb Drug
PMID: 1270
Related cita

C5? …Ah yes, check it.

Evidence for horizontal transfer of the EcoT38I restriction-modification gene to chromosomal DNA by the P2 **phage** and diversity of defective P2 prophages in Escherichia coli TH38 strains.

18. Kita K, Kawakami H, Tanaka H.
J Bacteriol. 2003 Apr;185(7):2296-305.
PMID: 12644501 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

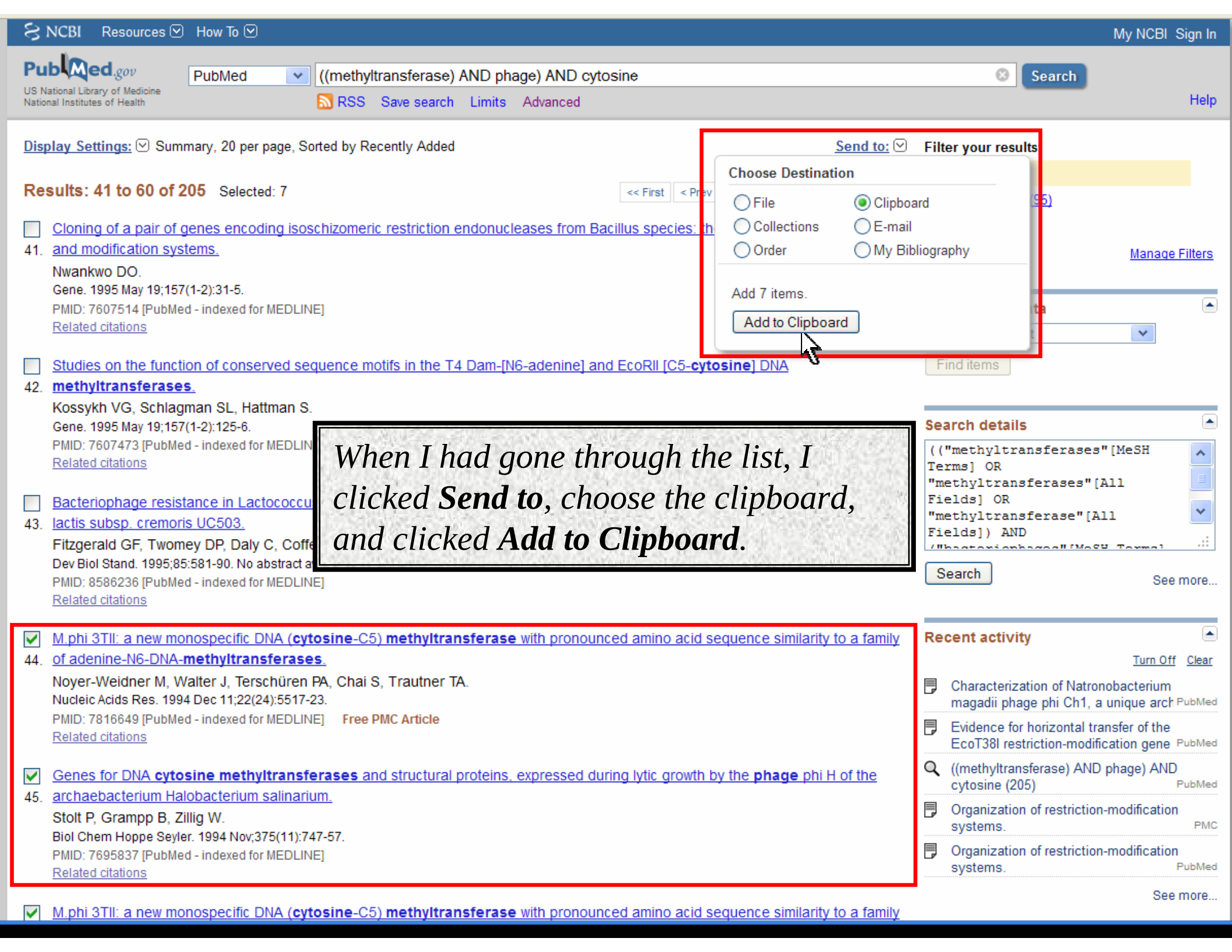
SQ9. C5,… What does this mean? Why did I like it?

Burkholderia thailandensis E125 harbors a temperate bacteriophage specific for Burkholderia mallei.

19. Woods DE, Jeddeloh JA, Fritz DL, DeShazer D.
J Bacteriol. 2002 Jul;184(14):4003-17.
PMID: 12081973 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

Mismatch repair in xenopus egg extracts is not strand-directed by DNA methylation.

20. Petranović M, Vlahović K, Zahradka D, Dzidić S, Radman M.
Neoplasma. 2000;47(6):375-81.
PMID: 11263862 [PubMed - indexed for MEDLINE]
Related citations


NCBI Resources How To
My NCBI Sign In
PubMed.gov
US National Library of Medicine
National Institutes of Health
PubMed
((methyltransferase) AND phage) AND cytosine
Search
RSS Save search Limits Advanced
Help
Display Settings: Summary, 20 per page, Sorted by Recently Added
Send to:
Filter your results
Choose Destination
File
Clipboard
Collections
E-mail
Order
My Bibliography
Add 7 items.
Add to Clipboard
Results: 41 to 60 of 205 Selected: 7
<< First
< Prev
Manage Filters
Cloning of a pair of genes encoding isoschizomeric restriction endonucleases from Bacillus species: th
41. and modification systems.
Nwankwo DO.
Gene. 1995 May 19;157(1-2):31-5.
PMID: 7607514 [PubMed - indexed for MEDLINE]
Related citations
Studies on the function of conserved sequence motifs in the T4 Dam-[N6-adenine] and EcoRII [C5-cytosine] DNA
42. methyltransferases.
Kossykh VG, Schlagman SL, Hattman S.
Gene. 1995 May 19;157(1-2):125-6.
PMID: 7607473 [PubMed - indexed for MEDLIN
Related citations
Find items
Search details
(("methyltransferases"[MeSH Terms] OR "methyltransferases"[All Fields] OR "methyltransferase"[All Fields]) AND
Search
See more...
When I had gone through the list, I clicked Send to, choose the clipboard, and clicked Add to Clipboard.
Bacteriophage resistance in Lactococcu
43. lactis subsp. cremoris UC503.
Fitzgerald GF, Twomey DP, Daly C, Coffe
Dev Biol Stand. 1995;85:581-90. No abstract a
PMID: 8586236 [PubMed - indexed for MEDLINE]
Related citations
M.phi 3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family
44. of adenine-N6-DNA-methyltransferases.
Noyer-Weidner M, Walter J, Terschüren PA, Chai S, Trautner TA.
Nucleic Acids Res. 1994 Dec 11;22(24):5517-23.
PMID: 7816649 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations
Genes for DNA cytosine methyltransferases and structural proteins, expressed during lytic growth by the phage phi H of the
45. archaebacterium Halobacterium salinarium.
Stolt P, Grampp B, Zillig W.
Biol Chem Hoppe Seyler. 1994 Nov;375(11):747-57.
PMID: 7695837 [PubMed - indexed for MEDLINE]
Related citations
Recent activity
Turn Off Clear
Characterization of Natronobacterium magadii phage phi Ch1, a unique arch PubMed
Evidence for horizontal transfer of the EcoT38I restriction-modification gene PubMed
((methyltransferase) AND phage) AND cytosine (205) PubMed
Organization of restriction-modification systems. PMC
Organization of restriction-modification systems. PubMed
See more...
M.phi 3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to: Clipboard: 7 items

7 items were added to the Clipboard.

Clipboard items will be lost after eight hours of inactivity. The maximum number of Clipboard items is 500.

Results: 41 to 60 of 205

<< First < Prev Page 3 of 11 Next > Last >>

41. Cloning of a pair of genes encoding isoschizomeric restriction endonucleases from Bacillus species: the BspEI and BspMII restriction and modification systems.
Nwankwo DO.
Gene. 1995 May 19;157(1-2):31-5.
PMID: 7607514 [PubMed - indexed for MEDLINE]
Related citations

42. Studies on the function of conserved sequence motifs in the T **methyltransferases**.
Kossykh VG, Schlagman SL, Hattman S.
Gene. 1995 May 19;157(1-2):125-6.
PMID: 7607473 [PubMed - indexed for MEDLINE]
Related citations

43. Bacteriophage resistance in Lactococcus: molecular characterization of the ScrFI restriction/modification system from Lactococcus lactis subsp. cremoris UC503.
Fitzgerald GF, Twomey DP, Daly C, Coffey AG.
Dev Biol Stand. 1995;85:581-90. No abstract available.
PMID: 8586236 [PubMed - indexed for MEDLINE]
Related citations

44. M.phi 3TII: a new monospecific DNA (**cytosine**-C5) **methyltransferase** with pronounced amino acid sequence similarity to a family of adenine-N6-DNA-**methyltransferases**.
Noyer-Weidner M, Walter J, Terschüren PA, Chai S, Trautner TA.
Nucleic Acids Res. 1994 Dec 11;22(24):5517-23.
PMID: 7816649 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations Item in clipboard

45. Genes for DNA **cytosine methyltransferases** and structural proteins, expressed during lytic growth by the **phage** phi H of the archaebacterium Halobacterium salinarium.
Stolt P, Grampp B, Zillig W.
Biol Chem Hoppe Seyler. 1994 Nov;375(11):747-57.
EDLINE]

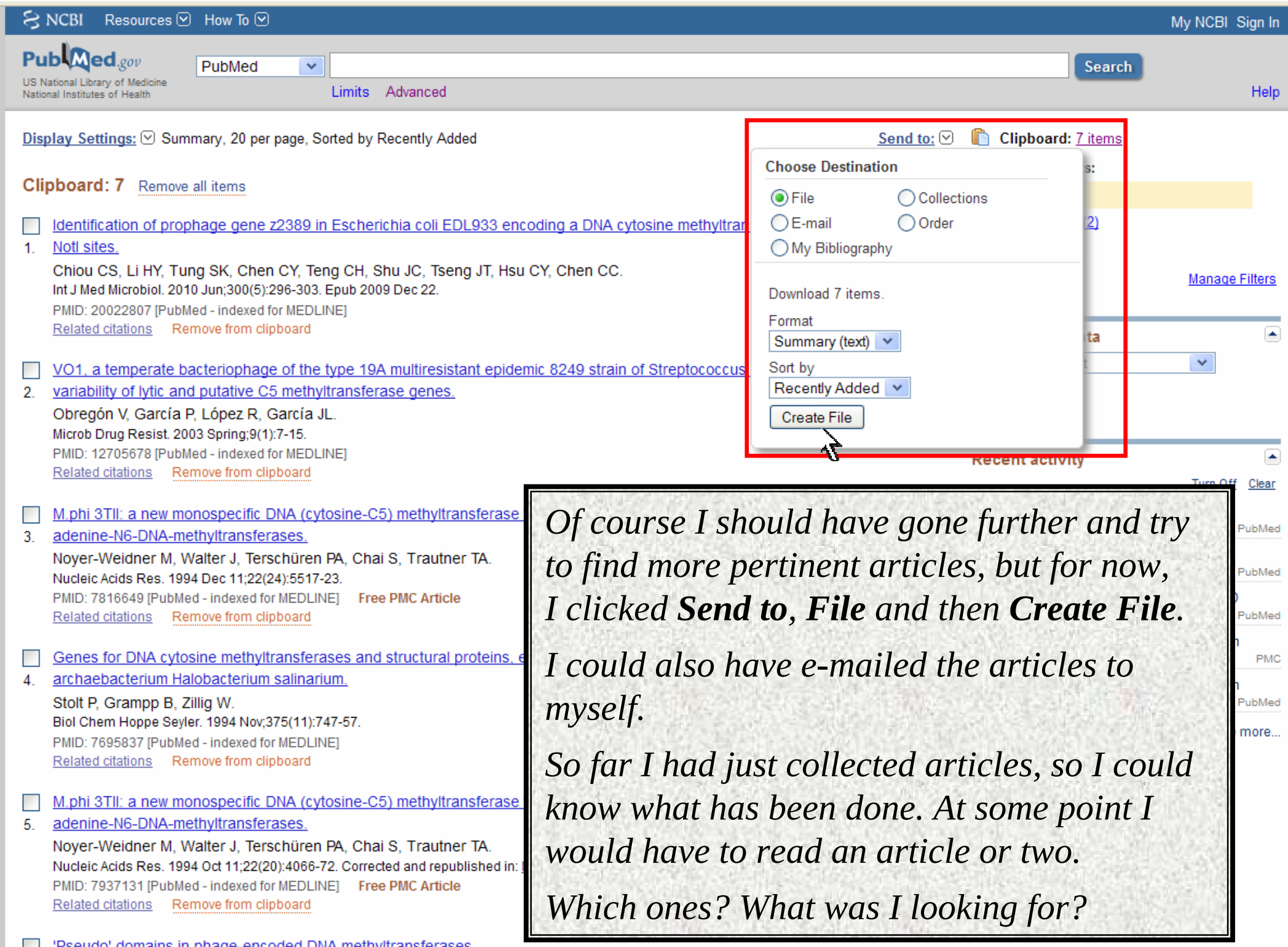
I evidently had checked 7 articles so far. To see what I had,I clicked those items…

Filter your results:
All (205)
Free Full Text (95)
Review (13)
Manage Filters

Find related data
Database: Select
Find items

Search details
(("methyltransferases"[MeSH Terms] OR "methyltransferases"[All Fields] OR "methyltransferase"[All Fields]) AND
Search
See more...

Recent activity
Turn Off Clear
Characterization of Natronobacterium magadii phage phi Ch1, a unique arch PubMed
Evidence for horizontal transfer of the EcoT38I restriction-modification gene PubMed
((methyltransferase) AND phage) AND cytosine (205) PubMed
Organization of restriction-modification systems. PMC
Organization of restriction-modification systems. PubMed



*Of course I should have gone further and try to find more pertinent articles, but for now, I clicked **Send to**, **File** and then **Create File**.*

I could also have e-mailed the articles to myself.

So far I had just collected articles, so I could know what has been done. At some point I would have to read an article or two.

Which ones? What was I looking for?

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.C. Identify the features that will enable you to find instances of your protein

Most of a protein's sequence is subject to random mutation without fatal damage to the protein's function. However, there are often certain residues that cannot be changed without a decrease in functionality to the extent that selection will weed out mutants with these changes from a phage's population. If you can identify these residues, then you have a powerful tool to determine whether a protein that bears similarity to the class you're interested in truly exhibit the desired function.

Sometimes specific amino acids are invariant amongst all members of a class of proteins, but more often what you find are common sequence ***motifs***, a collection of nearby amino acids that are conserved, more or less, as a group. You would do well to find an article that identifies conserved motifs in the protein class on which you have chosen to focus.

I needed to be able to identify cytosine methyltransferases in phage genomes, without relying on the English description of the gene (which can be wrong).

Sequence, not description, determines a protein's function.

Are there common sequence features – motifs – in the sequences of cytosine methyltransferases?

PubMed Advanced Search Builder

YouTube Tutorial

((methyltransferase) AND cytosine) AND motifs

Edit Clear

Builder

	All Fields	methyltransferase	Show index list
AND	All Fields	cytosine	Show index list
AND	All Fields	motifs	Show index list
AND	All Fields		Show index list

Search or Add to history

History

Clear history

Search	Add to builder	Query	Items found	Time
#3	Add	Search **((methyltransferase) AND phage) AND cytosine**	205	04:03:12
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:23:44
#0	Add	**pubmed** clipboard	7	04:04:00

I didn't know. Time to find out.

PubMed Advanced Search Builder

YouTube Tutorial

Use the builder below to create your search

Edit Clear

Builder

All Fields Show index list

AND All Fields Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#6	Add	Search **((methyltransferase) AND cytosine) AND motifs**	131	04:13:12
#3	Add	Search **((methyltransferase) AND phage) AND cytosine**	24	04:03:12
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:23:44
#0	Add	**pubmed** clipboard	7	04:04:00

131 articles is a lot, too much in fact, since I probably would have to read part of the article to determine if they said anything about sequence motifs.

Maybe looking at the results of this search would tell me how I could refine the search strategy.

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ((methyltransferase) AND cytosine) AND motifs Search

RSS Save search Limits Advanced Help

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Clipboard: 7 items

Results: 1 to 20 of 131 << First < Prev Page 1 of 7 Next > Last >>

1. Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 methylation in Arabidopsis thaliana.
   Enke RA, Dong Z, Bender J.
   PLoS Genet. 2011 Oct;7(10):e1002350. Epub 2011 Oct 27.
   PMID: 22046144 [PubMed - indexed for MEDLINE] Free PMC Article
   Related citations

2. The lytic phase of epstein-barr virus requires a viral genome with 5-methylcytosine residues in CpG sites.
   Kalla M, Göbel C, Hammerschmidt W.
   J Virol. 2012 Jan;86(1):447-58. Epub 2011 Oct 26.
   PMID: 22031942 [PubMed - indexed for MEDLINE]
   Related citations

3. Molecular dynamics simulations of human DNA **methyltransferase** 3B with selective inhibitor nanaomycin A.
   Caulfield T, Medina-Franco JL.
   J Struct Biol. 2011 Nov;176(2):185-91. Epub 2
   PMID: 21839172 [PubMed - indexed for MEDL
   Related citations

4. Twists and turns of DNA methylation.
   Frauer C, Leonhardt H.
   Proc Natl Acad Sci U S A. 2011 May 31;108(2
   PMID: 21593412 [PubMed - indexed for MEDL
   Related citations

5. Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines.
   Hagemann S, Heil O, Lyko F, Brueckner B.
   PLoS One. 2011 Mar 7;6(3):e17388.
   PMID: 21408221 [PubMed - indexed for MEDLINE] Free PMC Article
   Related citations

6. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation.
   Rigbolt KT, Prokhorova TA, Akimov V, Henningsen J, Johansen PT, Kratchmarova I, Kassem M, Mann M, Olsen JV, Blagoev B.
   Sci Signal. 2011 Mar 15;4(164):rs3.
   PMID: 21406692 [PubMed - indexed for MEDLINE]
   Related citations

Filter your results:
All (131)
Free Full Text (71)
Review (7)
Manage Filters

Find related data
Database: Select
Find items

Search details
(("methyltransferases"[MeSH

Histones,… human viruses,… human DNA and diseases,…

Looking into this, I find that I have the misfortune that my chosen protein has a relative that's involved in human health. However interesting that may be, it's getting in the way!

SQ10. What IS the relevance of cytosine methyltransferases to human DNA?

systems. PMC

NCBI Resources How To My NCBI Sign In

PubMed Home More Resources Help

PubMed Advanced Search Builder

YouTube Tutorial

((#6) NOT human) NOT histone

Edit Clear

Builder

	Recent Query	#6	
NOT	All Fields	human	Show index list
NOT	All Fields	histone	Show index list
AND	All Fields		Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#6	Add	Search **((methyltransferase) AND cytosine) AND motifs**	131	04:15:05
#3	Add	Search **((methyltransferase) AND phage) AND cytosine**	205	04:03:12
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:23:44
#0	Add	**pubmed** clipboard	7	04:04:00

I tried to get rid of these human-related articles by ridding the results of my last search (#6) of any hits that mentioned "human" or "histone".

That should increase the fraction that are pertinent to the proteins involved in bacteria and phages.

You are here: NCBI > Litera

Write to the Help Desk

GETTING STARTED
NCBI Education
NCBI Help Manual
NCBI Handbook
Training & Tutorials

NCBI Resources How To My NCBI Sign In

PubMed Home More Resources Help

PubMed Advanced Search Builder

YouTube Tutorial

Use the builder below to create your search

Edit Clear

Builder

All Fields — Show index list

AND All Fields — Show index list

Search or Add to history

History Clear history

Search	Add to builder	Query	Items found	Time
#7	Add	Search **((#6) NOT human) NOT histone**	77	04:16:58
#6	Add	Search **((methyltransferase) AND cytosine) AND motifs**	1[illegible]	04:15:05
#3	Add	Search **((methyltransferase) AND phage) AND cytosine**	205	04:03:12
#1	Add	Search **(restriction-modification) AND review[Publication Type]**	78	03:23:44
#0	Add	**pubmed** clipboard	7	04:04:00

77. Hmmm….didn't do as much good as I had hoped.

I could certainly go further in this direction, excluding more articles, but for now, I chose to make do with what I had.

You are here: NCBI > Litera

Write to the Help Desk

GETTING STARTED

NCBI Education
NCBI Help Manual
NCBI Handbook
Training & Tutorials

Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps

Nucleotide
BLAST
PubMed Central
Gene
Bookshelf
Protein

Reference Sequences
Map Viewer
Genome Projects
Human Genome
Mouse Genome
Influenza Virus

Research at NCBI
NCBI Newsletter
NCBI FTP Site
NCBI on Facebook
NCBI on Twitter
NCBI on YouTube

NCBI Resources How To My NCBI Sign In

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ((#6) NOT human) NOT histone Search

RSS Save search Limits Advanced Help

Display Settings: Summary, 20 per page, Sorted by Recently Added Send to: Clipboard: 7 items

Results: 1 to 20 of 77

<< First < Prev Page 1 of 4 Next > Last >>

1. Twists and turns of DNA methylation.
   Frauer C, Leonhardt H.
   Proc Natl Acad Sci U S A. 2011 May 31;108(22):8919-20. Epub 2011 May 18. No abstract available.
   PMID: 21593412 [PubMed - indexed for MEDLINE] Free PMC Article
   Related citations

2. [Secondary structure of SsoII-like (cytosine-5)-DNA methyltransferases N-terminal region determined by circular dichroism spectroscopy].
   Riazanova AIu, Molochkov NV, Abrosimova LA, Alekseevskiĭ AV, Kariagina AS, Protsenko AS, Friedhoff P, Oretskaia TS, Kubareva EA.
   Mol Biol (Mosk). 2010 Sep-Oct;44(5):911-21. Russian.
   PMID: 21090246 [PubMed - indexed for MEDLINE]
   Related citations

3. The BsaHI restriction-modification system: cloning, sequencing and analysis of conserved motifs.
   Neely RK, Roberts RJ.
   BMC Mol Biol. 2008 May 14;9:48.
   PMID: 18479503 [PubMed - indexed for MEDLINE] Free PMC Article
   Related citations

4. The … pression patterns during ovarian and …
   Fen…
   Zool…
   PMID…
   Rela…

5. The … cess DNA during mitosis.
   Schaefer M, Steringer JP, Lyko F.
   PLoS One. 2008 Jan 9;3(1):e1414.
   PMID: 18183295 [PubMed - indexed for MEDLINE] Free PMC Article
   Related citations

6. Cysteine of sequence motif VI is essential for nucleophilic catalysis by yeast tRNA m5C methyltransferase.
   Walbott H, Husson C, Auxilien S, Golinelli-Pimpaneau B.
   RNA. 2007 Jul;13(7):967-73. Epub 2007 May 2.

This article looked promising!

"... analysis of conserved motifs"

That sounded like what I was looking for.

Filter your results:

All (77)
Free Full Text (42)
Review (2)

Manage Filters

Find related data

Database: Select

Find items

Search details

```
((#6) NOT ("humans"[MeSH Terms]
OR "humans"[All Fields] OR
"human"[All Fields])) NOT
("histones"[MeSH Terms] OR
"histones"[All Fields] OR
"histone"[All Fields])
```

Search See more...

Recent activity

Turn Off Clear

- Molecular phylogenetics of DNA 5mC-methyltransferases. PubMed
- (((((methyltransferase) AND cytosine) AND motifs)) NOT human) NOT... (77) PubMed
- ((methyltransferase) AND cytosine) AND motifs (131) PubMed
- Characterization of Natronobacterium magadii phage phi Ch1, a unique arch PubMed
- Evidence for horizontal transfer of the EcoT38I restriction-modification gene PubMed

BMC Mol Biol. 2008 May 14;9:48.

The BsaHI restriction-modification system: cloning, sequencing and analysis of conserved motifs.

Neely RK, Roberts RJ.

School of Chemistry, The University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK. r.neely@ed.ac.uk

Abstract

BACKGROUND: Restriction and modification enzymes typically recognise short DNA sequences of between two and eight bases in length. Understanding the mechanism of this recognition represents a significant challenge that we begin to address for the BsaHI restriction-modification system, which recognises the six base sequence GRCGYC.

RESULTS: The DNA sequences of the genes for the BsaHI methyltransferase, bsaHIM, and restriction endonuclease, bsaHIR, have been determined (GenBank accession #EU386360), cloned and expressed in E. coli. Both the restriction endonuclease and methyltransferase enzymes share significant similarity with a group of 6 other enzymes comprising the restriction-modification systems HgiDI and HgiGI and the putative HindVP, NlaCORFDP, NpuORFC228P and SplZORFNP restriction-modification systems. A sequence alignment of these homologues shows that their amino acid sequences are largely conserved and highlights several motifs of interest. We target one such conserved motif, reading SPERRFD, at the C-terminal end of the bsaHIR gene. A mutational analysis of these amino acids indicates that the motif is crucial for enzymatic activity. Sequence alignment of the methyltransferase gene reveals a short motif within the target recognition domain that is conserved among enzymes recognising the same sequences. Thus, this motif may be used as a diagnostic tool to define the recognition sequences of the cytosine C5 methyltransferases.

CONCLUSION: We have cloned and sequenced the BsaHI restriction and modification enzymes. We have identified a region of the R. BsaHI enzyme that is crucial for its activity. Analysis of the amino acid sequence of the BsaHI methyltransferase enzyme led us to propose two new motifs that can be used in the diagnosis of the recognition sequence of the cytosine C5-methyltransferases.

PMID: 18479503 [PubMed - indexed

Lots of good words from Neely and Roberts.

I signed on for the full article.

Images from this publicati

MeSH Terms, Substances

LinkOut - more resources

Related citations

Cloning of the BssHII restriction-modification system in Escherichia [Nucleic Acids Res. 1997]

Cloning and analysis of a bifunctional methyltransferase/restrictio [BMC Mol Biol. 2009]

[Gene cloning, comparative analysis of the protein structures from Fs [Mol Biol (Mosk). 2007]

Review Structure, function, and mechanism of HhaI DNA metl [Crit Rev Biochem Mol Biol. 2002]

Review Chemistry and biology of DNA methyltransfera [Crit Rev Biochem Mol Biol. 1996]

See reviews...

See all...

Related information

Related Citations

Nucleotide

References for this PMC Article

Substance (MeSH Keyword)

Taxonomy via GenBank

Protein

Free in PMC

Recent activity

Turn Off Clear

The BsaHI restriction-modification system: cloning, sequencing and analysis of c PubMed

Molecular phylogenetics of DNA 5mC-methyltransferases. PubMed

((((methyltransferase) AND cytosine) AND

BMC Molecular Biology

BioMed Central

Research article

Open Access

The BsaHI restriction-modification system: Cloning, sequencing and analysis of conserved motifs

Robert K Neely*[1] and Richard J Roberts[2]

Address: [1]School of Chemistry, The University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, UK and [2]New England Biolabs Inc., 240 County Road, Ipswich, Massachusetts, 01938, USA

Email: Robert K Neely* - r.neely@ed.ac.uk; Richard J Roberts - roberts@neb.com

* Corresponding author

Published: 14 May 2008

BMC Molecular Biology 2008, **9**:48 doi:10.1186/14

This article is available from: http://www.biomedc

Abstract

Background: Restriction between two and eight represents a significant ch system, which recognises t

I skimmed the article to see if there was anything interesting enough to warrant a more careful reading…

First of all, was the methyltransferase of this restriction-modification system a cytosine methyltransferase?

*Easiest way to find out was to search, so I opened up **Advanced Search** from the **Edit** menu.*

Search

Arrange Windows

Looking For:
cytosine in the current document

Results:
1 documents with 25 instances

New Search

Results:

- 1471-2199-9-48.fm
  - the **cytosine** C5 methyltransferases. Conc
  - the **cytosine** C5-methyltransferases. Back
  - and **cytosine** bases: GR/CGYC, where '/' is
  - other **cytosine** C5 methyltransferases reve
  - a **cytosine** C5 methyltransferase [13]. To
  - central **cytosine** of the recognition sequenc
  - **cytosine** bases of its GRCGYC recognition
  - **cytosine** C5 methyltransferase structural r
  - target **cytosine** is a guanine, a conserved
  - **cytosine** C5-methyltransferases with know
  - targeting **cytosine** C5-methyltransferases.
  - of **cytosine** C5 methyltransferases. The cc
  - recognising **cytosine** C5-methyltranferases
  - the **cytosine** C5 methylating enzymes targ
  - **cytosine** C5 methyltransferases. Showing
  - a **cytosine** C5 methyltransferase that has
  - central **cytosine** of its GRCGYC recognition
  - **cytosine** C5 methyltransferase. Furthermc
  - **cytosine** C5 methyltransferases. Should th
  - Extrahelical **Cytosine** and Rearranged Base
  - DNA-(**cytosine**-C5)-methyltransferases. El
  - DNA-(**cytosine**-C5)-methyltransferases re
  - M: **Cytosine**-specific type II DNA methyltra
  - from **Cytosine** Methyltransferases. Nucl Ac
  - DNA **cytosine**-5 methyltransferase HhaI b

Neely-2008-BsaHI-Resmod.pdf - Adobe Reader

File Edit View Window Help

5 / 10 93% Tools Sign Comment

BMC Molecular Biology 2008, 9:48 http://www.biomedcentral.com/1471-2199/9/48

This allowed me to rapidly examine all instances of the word 'cytosine' in the article.

I soon came to a winner.

No doubt about it. There was good reason to believe M.BsaHI is a cytosine methyltransferase.

what com
dues, wh
bonding f
the R351A
striking w
neighbour
ble to that
in the acti
dues sugg
ensuring t

M. BsaHI

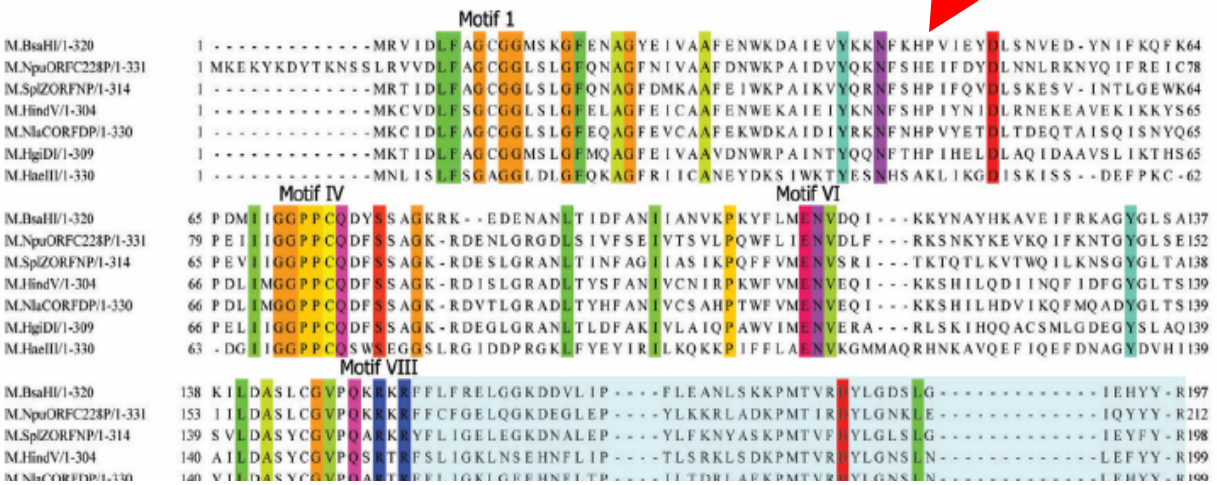
Figure 5 shows that the M. BsaHI methyltransferase contains all of the conserved motifs of a cytosine C5 methyltransferase [13]. To determine the target base for methylation, pUC19 plasmid DNA was methylated with the M. BsaHI enzyme. Figure 6 shows the result of subsequent digestion of the DNA with the R. HpaII and R. H restriction enzymes. The single overlapping HhaI/Bs site (CGCGCC (where boldface bases represent the H recognition sequence and the underlined bases are the

aligned along with the sequence for M. HaeIII, which also has a known structure [8] but shares more similarity with M. BsaHI, as shown in Figure 5.

The TL motif at the centre of the TRD is shared by M. BsaHI (TL...), its homologues and M. HaeIII (IV...). The

...and look what else I found:

eleven bases upstream of this motif in M. HaeIII, this con-



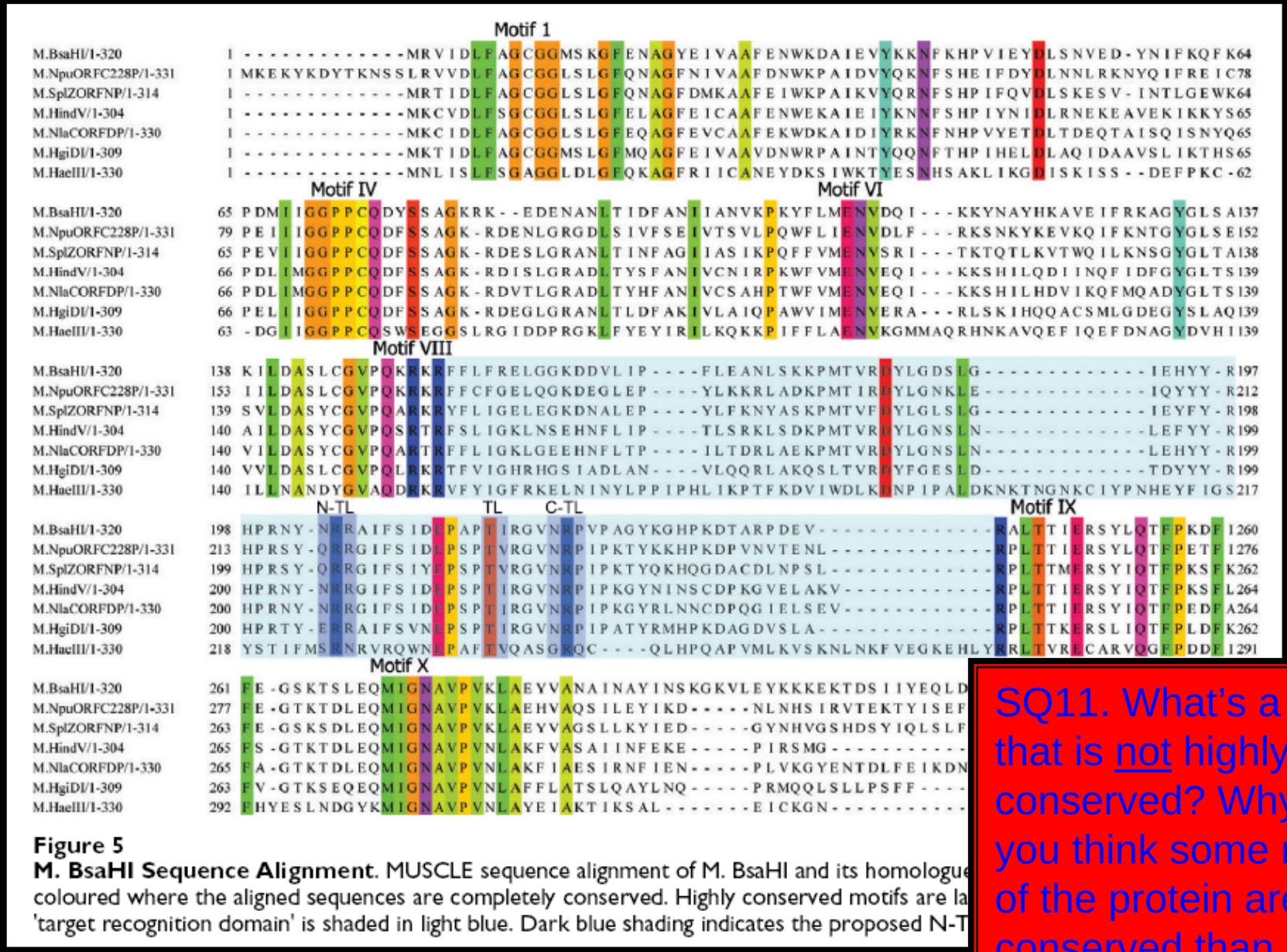


Figure 5
M. BsaHI Sequence Alignment. MUSCLE sequence alignment of M. BsaHI and its homologue
coloured where the aligned sequences are completely conserved. Highly conserved motifs are la
'target recognition domain' is shaded in light blue. Dark blue shading indicates the proposed N-T

SQ11. What's a region that is not highly conserved? Why do you think some regions of the protein are more conserved than others?

This was wonderful! Motifs! I could now see which amino acids are found in many cytosine methyltransferases. These might help me identify undiscovered methyltransferases.

34. A tobacco NtMET1 cDNA encoding a DNA methyltransferase: molecular characterization and abnormal phenotypes of transgenic tobacco plants.
Nakano Y, Steward N, Sekine M, Kusano T, Sano H.
Plant Cell Physiol. 2000 Apr;41(4):448-57.
PMID: 10845458 [PubMed - indexed for MEDLINE]
Related citations

35. Hybrid mouse-prokaryotic DNA (cytosine-5) methyltransferases retain the
Pradhan S, Roberts RJ.
EMBO J. 2000 May 2;19(9):2103-14.
PMID: 10790376 [PubMed - indexed for MEDLINE] Free PMC Article
Related citations

36. Multiple DNA methyltransferase genes in Arabidopsis thaliana.
Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ.
Plant Mol Biol. 1999 Sep;41(2):269-78.
PMID: 10579493 [PubMed - indexed for MEDLINE]
Related citations

160, 7–16.
37. Bujnicki,J.M. and Radlinska,M. (1999) Molecular phylogenetics of DNA 5mC-methyltransferases. *Acta. Microbiol. Pol.*, **48**, 19–30.

But before I left this PubMed search,… was there anything else that might be valuable?

Hey, there's my friends Bujnicki and Radlinska!

This had become an article to read for two reasons: mobility and motifs.

I resolved to give it a try.

scobolus immersus.

Related citations

40. Molecular phylogenetics of DNA 5mC-methyltransferases.
Bujnicki JM, Radlinski M.
Acta Microbiol Pol. 1999;48(1):19-30.
PMID: 10467693 [PubMed - indexed for MEDLINE]
Related citations

<< First | < Prev | Page 2 of 4 | Next > | Last >>

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to:

Display Settings: Abstract Send to:

Clipboard: 7 items

Acta Microbiol Pol. 1999;48(1):19-30.

Molecular phylogenetics of DNA 5mC-methyltransferases.

Bujnicki JM, Radlinska M.

Department of Virology, University of Warsaw, Poland. iamb@ibbrain.ibb.waw.pl

Abstract

We have identified a total of 88 members of the DNA-(cytosine-5) methyltransferase (5mC MTase) family whose sequences have been deposited in the databases. The results of a comparison of these sequences is presented in the form of an alignment-based phylogenetic tree and sequence logos for 10 conserved motifs. Phylogenetic analysis showed that members of the family aggregate into subfamilies which are usually consistent with their target specificity. However, it was also shown that similar target specificity does not necessarily imply close homology of the catalytic domain of MTases, which strongly supports the hypothesis that target recognition evolved independently of catalytic properties. This analysis also indicate that the 5mC MTase was present in the cenancestor (last common ancestor) of eubacteria, archaebacteria, and eukaryotes. The phylogeny of the 5mC MTases catalytic domain provides the basis for establishing the patterns of evolutionary change that characterize this family of proteins with conserved structural core and variable and mobile modules not directly involved in formation of the active site.

PMID: 10467693 [PubMed - indexed for MEDLINE]

MeSH Terms, Substances

LinkOut - more resources

Bad news. No link to the full article.

Was that the end of the road?
Did I need to give up finding this article?

BY NO MEANS!

Try the journal route through the library…

Related citations

Plant cytosine-5 DNA methyltransferases: structure, function, and molecu [Genomics. 2007]

Evolutionary connection between the catalytic subunits of DNA-dependc [BMC Struct Biol. 2003]

M.phi 3TII: a new monospecific DNA (cytosine-C5) methyltr [Nucleic Acids Res. 1994]

Review Protein phylogenies and signature sequences: A reap [Microbiol Mol Biol Rev. 1998]

Review The neomuran origin of archaebacteria, the negibacterial [Int J Syst Evol Microbiol. 2002]

See reviews...

See all...

Cited by 5 PubMed Central articles

Genome comparison and context analysis reveals putative mobile [Nucleic Acids Res. 2010]

Sequence-structure-function studies of tRNA:m5C methyltran [Nucleic Acids Res. 2004]

A DNA methyltransferase can protect the genome from postdisturbance [J Bacteriol. 2002]

See all...

Related information

Related Citations

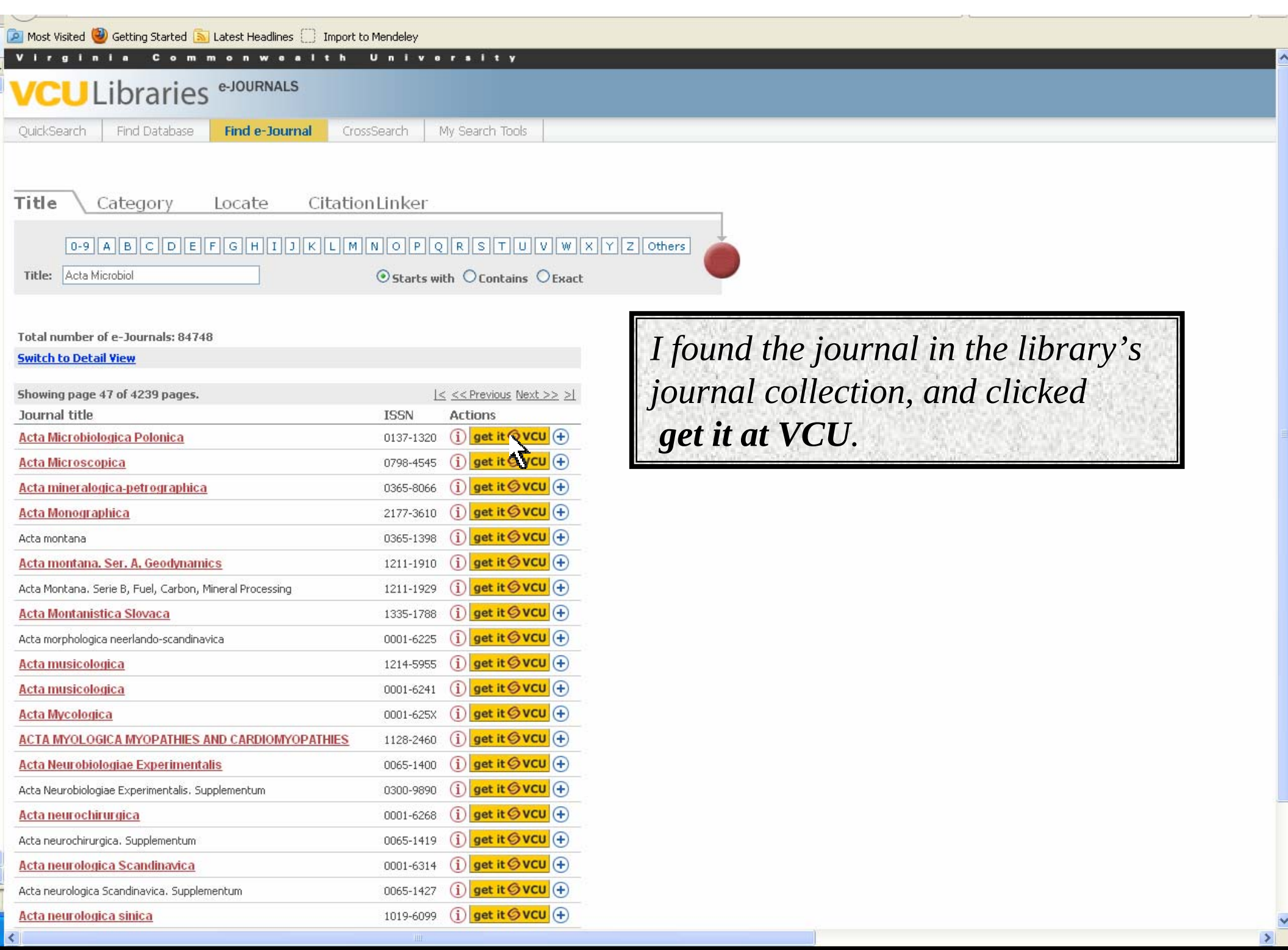
Cited in PMC

Cited in Books

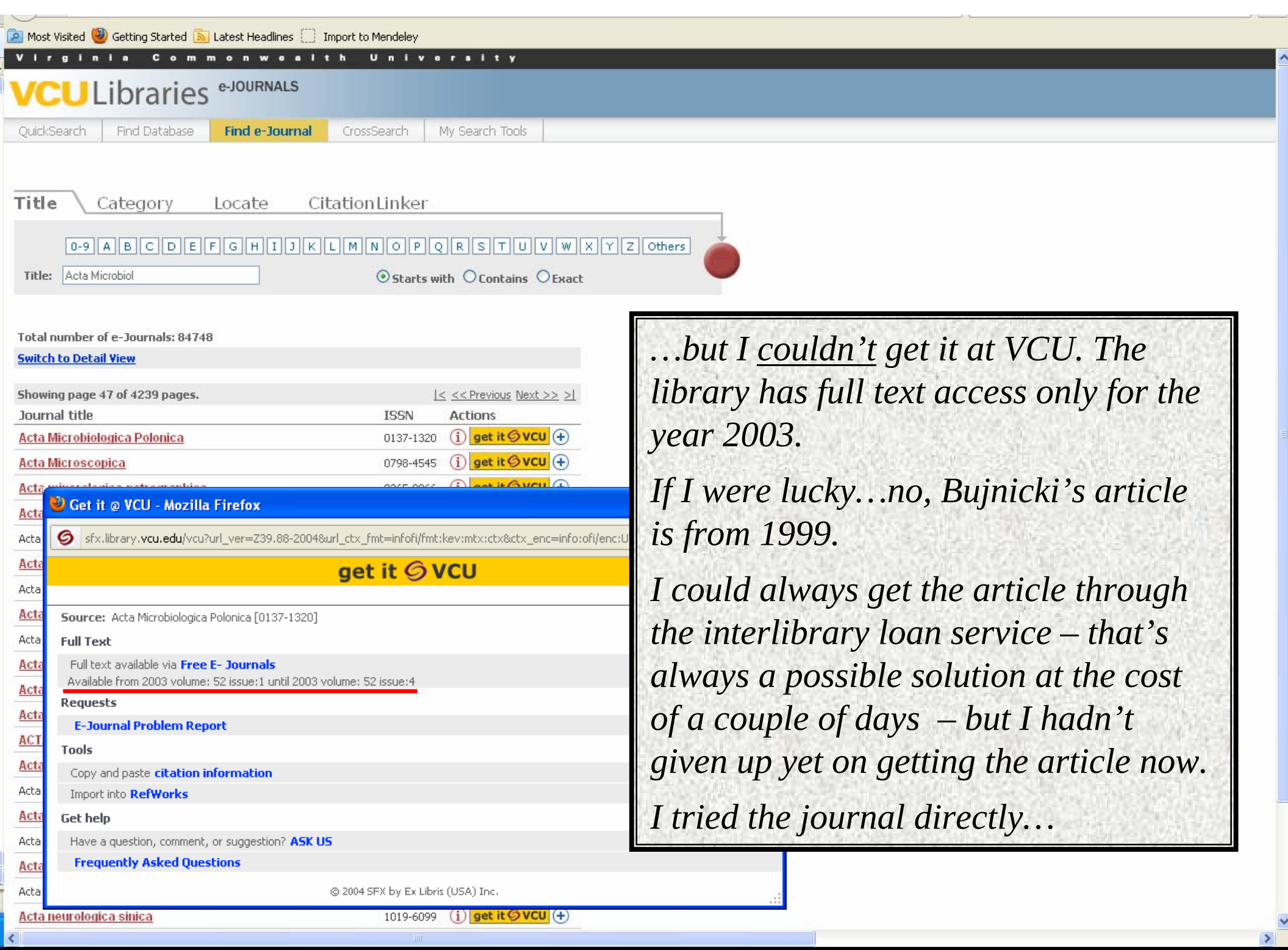
Recent activity

Turn Off Clear

Molecular phylogenetics of DNA 5mC-methyltransferases. PubMed



I found the journal in the library's journal collection, and clicked ***get it at VCU****.*



…but I couldn't get it at VCU. The library has full text access only for the year 2003.

If I were lucky…no, Bujnicki's article is from 1999.

I could always get the article through the interlibrary loan service – that's always a possible solution at the cost of a couple of days – but I hadn't given up yet on getting the article now.

I tried the journal directly…

Acta Microbiologica Polonica

Contents of Archive

Acta Microbiologica Polonica
1999, Vol. 48, No. 1

CONTENTS

Google got me to the journal, and I navigated through the archives to the proper issue.

Victory! The table of contents provided a link!

Acta Mirobiologica Polonica
1999, Vol. 48, No 1, 19-30

Molecular Phylogenetics of DNA 5mC-Methyltransferases

JANUSZ M. BUJNICKI*+ and MONIKA RADLINSKA+

Department of Virology, Institute of Microbiology, University of Warsaw,
Nowy Swiat 67, 00-046 Warsaw, Poland

Received in revised form 29 December, 1998

We have identified a total of 88 members of the DNA-(cytosine-5) methyltransferase (5mC MTase) family whose sequences have been deposited in the databases. The results of a comparison of these sequences is presented in the form of an alignment-based phylogenetic tree and sequence logos for 10 conserved motifs. Phylogenetic analysis showed that members of the family aggregate into subfamilies which are usually consistent with their target specificity. However, it was also shown that similar target specificity does not necessarily imply close homology of the catalytic domain of MTases, which strongly supports the hypothesis that target recognition evolved independently of catalytic properties. This analysis also indicate that the 5mC MTase was present in the cenancestor (last common ancestor) of eubacteria, archaebacteria, and eukaryotes. The phylogeny of the 5mC MTases catalytic domain provides the basis for establishing the patterns of evolutionary change that characterize this family of proteins with conserved structural core and variable and mobile modules not directly involved in formation of the active site.

* To whom correspondence should be addressed at both authors' present adress:
+ Henry Ford Health System, Molecular Biology Research Program, One Ford Place, Suite 5D, Detroit, MI 48202, USA. e-mail: iamb@ibbrain.ibb.waw.pl

Unfortunately, the link took me to a single page. They just had abstracts on-line.

Now give up? ...Never! (well, almost never)

Maybe Bujnicki or someone had posted a full length version of the article somewhere?

+You Search Images Maps YouTube News Gmail Documents Calendar More

Google Molecular phylogenetics of DNA 5mC-methyltransferases Sign in

Search About 777 results (0.36 seconds)

Everything
Images
Maps
Videos
News
Shopping
More

Richmond, VA
Change location

Show search tools

Scholarly articles for **Molecular phylogenetics of DNA 5mC-methyltransferases**
... **phylogenetics of DNA 5mC-methyltransferases** - Bujnicki - Cited by 19
Phylogenomic analysis of 16S rRNA:(guanine-N2) ... - BUJNICKI - Cited by 33
... Trm4p and its relationship to **DNA**: m5C and RNA: ... - Bujnicki - Cited by 45

[PDF] **Molecular Phylogenetics of DNA** SmC-**Methyltransferases**
ftp://212.87.21.35/iamb/papers/1999.AMP.D...-m5C.pdf
File Format: PDF/Adobe Acrobat - Quick View
by JM BUJNICKI - 1999 - Cited by 19 - Related articles
Molecular Phylogenetics of DNA SmC-**Methyltransferases** ... We have identified a total of 88 members of the **DNA**-(cytosine-5) **methyltransferase** (**5mC** MTase) ...

Molecular phylogenetics of DNA 5mC-methyltransferases.
www.ncbi.nlm.nih.gov/pubmed/10467693
by JM Bujnicki - 1999 - Cited by 19 - Related articles
Molecular phylogenetics of DNA 5mC-methyltransferases. Bujnicki JM, Radlinska M. Department of Virology, University of Warsaw, Poland.

[PDF] **Molecular Phylogenetics of DNA 5mC- Methyltransferases**
protein.uta.fi/~brshen/PHY/Tam007_JaniHaukka.pdf
File Format: PDF/Adobe Acrobat - Quick View
Molecular Phylogenetics of. **DNA 5mC-**. **Methyltransferases**. JANUSZ M. BUJNICKI and MONIKA RADLINSKA. Phylogenetics 23.04.2007. Jani Haukka ...

on PubMed - PubMed Result
www.biology-direct.com/pubmed/related/19105819
5: Bujnicki JM, Radlinska M. **Molecular phylogenetics of DNA 5mC-methyltransferases**. Acta Microbiol Pol. 1999;48(1):19-30. PubMed PMID: 10467693.

Acta Microbiol Pol Acta microbiologica Polonica 48 0137-1320 ...
opencitations.net/doc/expression:pmid/10467693.rdf
Acta Microbiol Pol Acta microbiologica Polonica 48 0137-1320 **Molecular phylogenetics of DNA 5mC-methyltransferases** 10467693 1999-01-01T00:00:00Z ...

Phylogenomic analysis of 16S rRNA:(guanine-N2) **methyltransferases**
www.fasebj.org/content/14/14/2365.full
by JM BUJNICKI - 2000 - Cited by 33 - Related articles
Bioinformatics Unit, International Institute of Molecular and Cell Biology
Molecular phylogenetics of DNA 5mC-methyltransferases

https://www.google.com/calendar?tab=wc

Indeed! Someone had!

Click the link…

Acta Microbiologica Polonica
1999, Vol. 48, No. 1, 19–30

Molecular Phylogenetics of DNA 5mC-Methyltransferases

JANUSZ M. BUJNICKI[+*] and MONIKA RADLINSKA[+]

Department of Virology, Institute of Microbiology, University of Warsaw,
Nowy Swiat 67, 00-046 Warsaw, Poland

Received in revised form 29 December, 1998

Abstract

We have identified a total of 88 members of the DNA-(cytosine-5) methyltransferase (5mC MTase) family whose sequences have been deposited in the databases. The results of a comparison of these sequences is presented in the form of an alignment-based phylogenetic tree and sequence logos for 10 conserved motifs. Phylogenetic analysis showed that members of the family aggregate into subfamilies which are usually consistent with their target specificity. However, it was also shown that similar target specificity does not necessarily imply close homology of the catalytic domain of MTases, which strongly supports the hypothesis that target recognition evolved independently of catalytic properties. This analysis also indicate that the 5mC MTase was present in the cenancestor (last common ancestor) of eubacteria, archaebacteria, and eukaryotes. The phylogeny of the 5mC MTases catalytic domain provides the basis for establishing the patterns of evolutionary change that characterize this family of proteins with conserved structural core and variable and mobile modules not directly involved in formation of the active site.

Introduction

Enzymatic transfer of the methyl group from S-adenosyl-L-methionine (AdoMet) to certain nucleotides in DNA is the most common form of biological DNA modification. DNA methylation in eukaryotes has been implicated in the control of gene regulation, genomic imprinting and embryonic development (L i *et al.*, 1992). In prokaryotes, DNA methylation affects such diverse phenomena as protection of DNA against digestion by endonucleases, control of initiation of DNA replication, targeting the correction of errors in DNA replication, and definition of origins of packing in maturation of phage DNA (reviewed by N o y e r - W e i d n e r and T r a u t n e r, 1993).

DNA methyltransferases (MTases) can be divided into those that methylate the exocyclic amino group of adenines and cytosines (amino-MTases) and those that

…and there I was!

As usual, I skimmed the article first, looking for something pertinent regarding motifs.

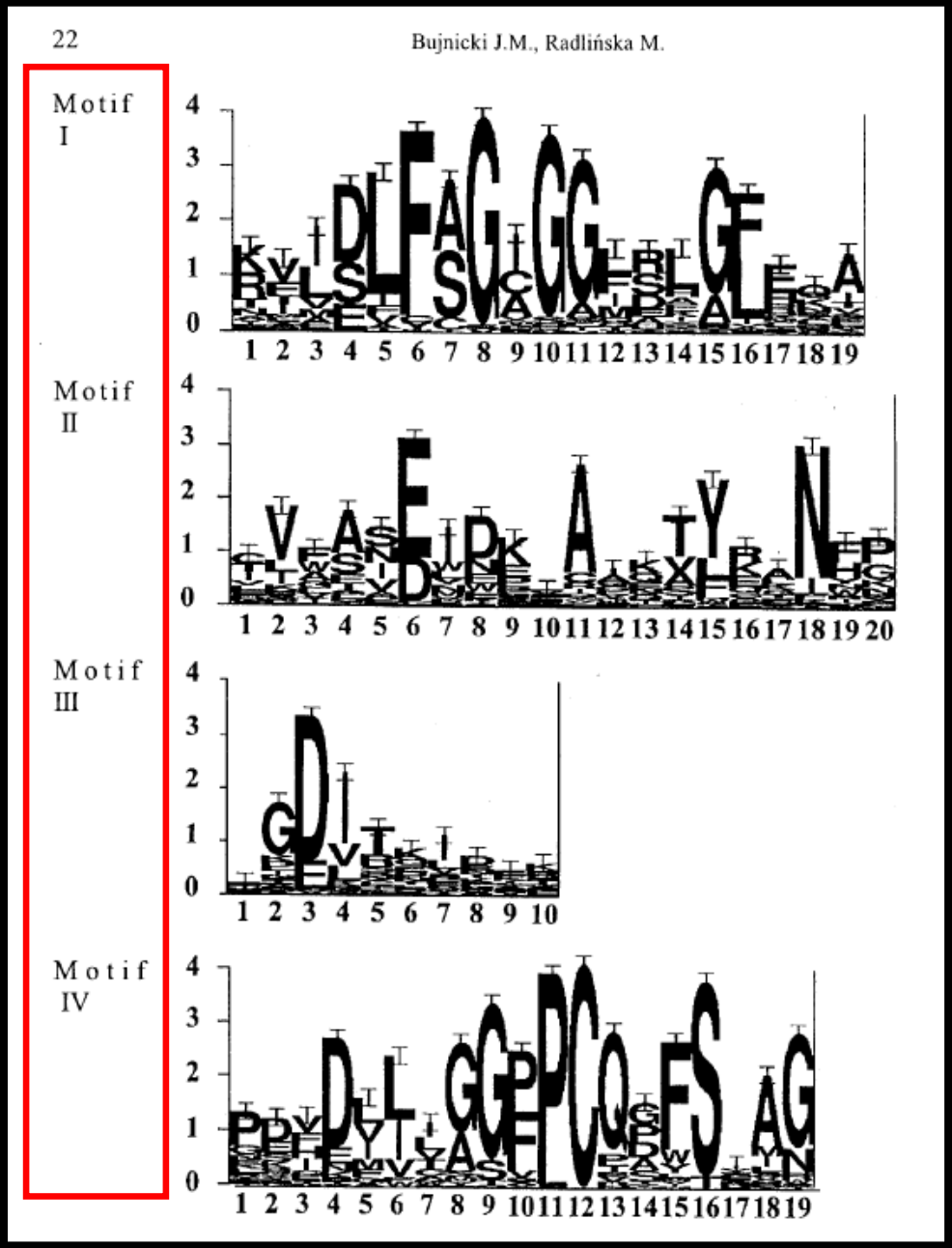


Fig. 1. Sequence logos displaying both significant residues and subtle sequence patterns of ten conservedmotifs derived from an alignment of 88 5mC-MTases.

The residues are stacked on top of each other in increasing order of their frequencies, so that the general consensus at every position can be found by reading the top amino acid residue. The height of the entire stack in a logo represents the total information content of the sequences at that position. The vertical bar shows the graduation of the information content measured in bits (Schneider and Stephens, 1990).

…and found motifs galore! From the title of the figure, I gathered that the motifs were derived from 88 methyltransferases, far more than the paltry 7 sequences used by Neely and Roberts.

But what did the display mean?

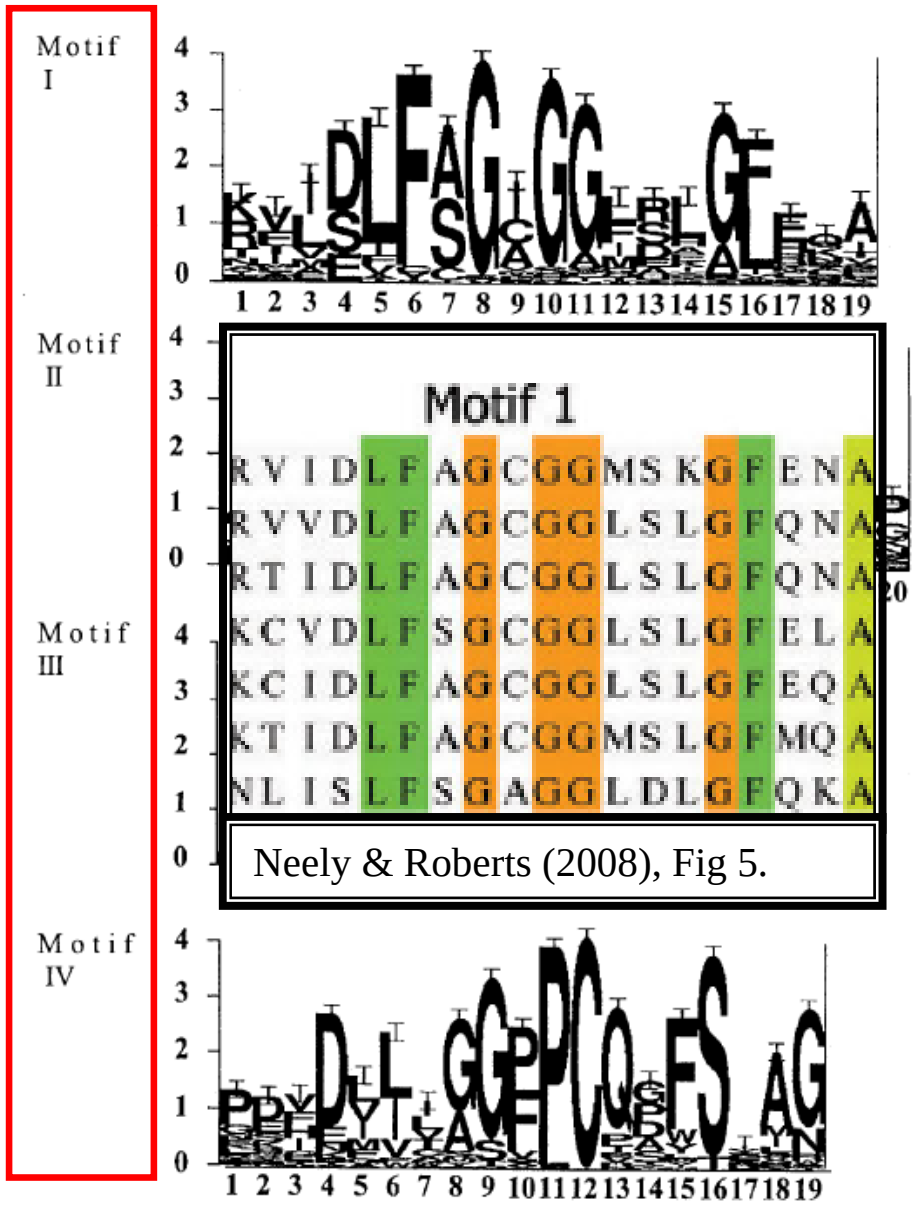


Fig. 1. Sequence logos displaying both significant residues and subtle sequence patterns of ten conservedmotifs derived from an alignment of 88 5mC-MTases.

The residues are stacked on top of each other in increasing order of their frequencies, so that the general consensus at every position can be found by reading the top amino acid residue. The height of the entire stack in a logo represents the total information content of the sequences at that position. The vertical bar shows the graduation of the information content measured in bits (Schneider and Stephens, 1990).

The figure legend called the display "sequence logos". From the legend and some exploration on the web (including the cited article by Schneider & Stephens), I understood that the tall letters are highly conserved, the relative height corresponding to their frequency at the given position in the protein.

By comparing the rendition of Motif I from Neely & Roberts' figure with that from Bujnicki & Radlinska, the nature of logos became clear.

I now had sequence motifs for cytosine methyltransferases.

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.D. Learn how to find the critical features yourself

It's one thing to find a pretty picture of a motif of your favorite class of proteins and quite another to find them yourself in protein sequences. Yet that is what you need to do if your goal is to identify proteins that have not previously been identified. Take the proven cases that you find in articles and collect their sequences within BioBIKE. Then with them in hand, use the function MOTIFS-IN to identify the motifs that are shown in the article.

How to identify these motifs in phage cytosine methyltransferases?

Before taking that step, I first needed to make sure that I could find them in proteins in which they are known to occur, i.e. the proteins used by Neely & Roberts.

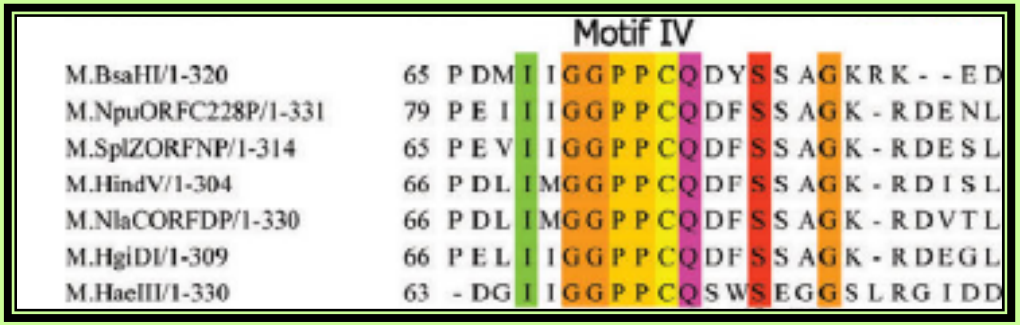
How to get the sequences of these proteins?

HELP! FILE EDIT EXIT PROBLEM ALL ANNOTATION ARITHMETIC DEFINITION FLOW-LOGIC GENES-PROTEINS GENOME

INPUT-OUTPUT LISTS-TABLES OTHER-COMMANDS STRINGS-SEQUENCES DATA FAVORITES ORGANISMS SUBSYSTEMS

Help:

SEQUENCE-SIMILAR-TO each "PDLIMGGPPCQDFSSAGKRD..." PROTEIN-VS-PROTEIN Options



These are bacterial sequences, so PhAnToMe/BioBIKE may have them.

I copied the sequence of the most conserved region of one of the proteins into the query box, and executed the function.

SQ12. Which protein sequence did I choose?

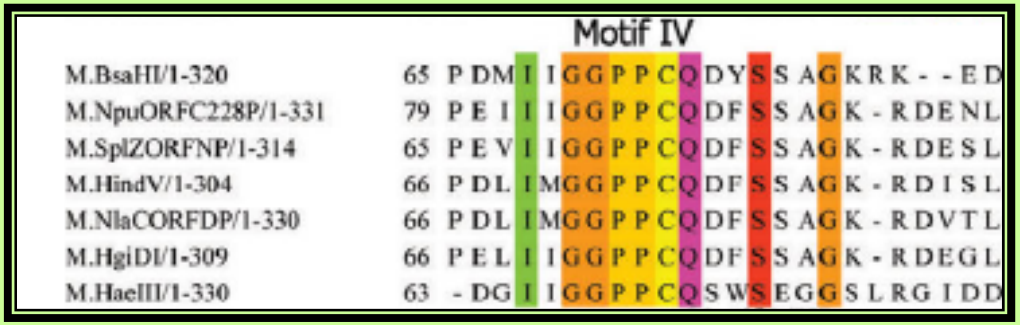
Results

HELP! FILE EDIT EXIT PROBLEM ALL ANNOTATION ARITHMETIC DEFINITION FLOW-LOGIC GENES-PROTEINS GENOME

INPUT-OUTPUT LISTS-TABLES OTHER-COMMANDS STRINGS-SEQUENCES DATA FAVORITES ORGANISMS SUBSYSTEMS

Help:

SEQUENCE-SIMILAR-TO each "PDLIMGGPPCQDFSSAGKRD..." PROTEIN-VS-PROTEIN Options



The function returned proteins. Three of them had names similar to those in Neely & Roberts' figure. I confirmed that they were really the same.

vpl execution printout - Mozilla Firefox

File Edit View History Bookmarks Tools Help

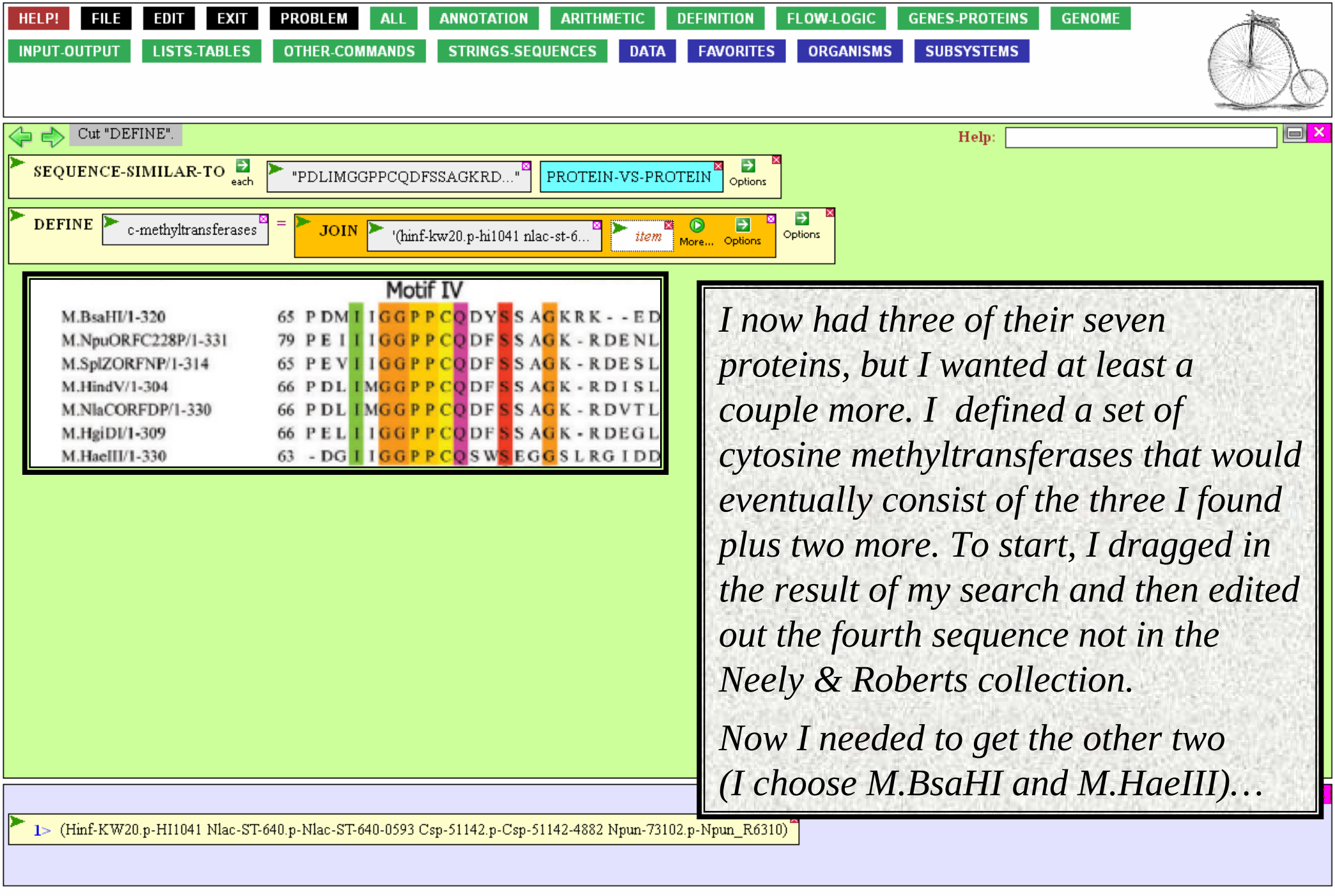
vpl execution printout

edwards.sdsu.edu:7003/Topdir1//home/biobike/user Google

	QUERY	Q-START	Q-END	TARGET	T-START	T-END	E-VALUE	%ID
1.	Seq1	1	23	Hinf-KW20.p-HI1041	66	88	3.0d-6	100.0
2.	Seq1	1	23	Nlac-ST-640.p-Nlac-ST-640-0593	66	88	9.0d-6	91.3
3.	Seq1	1	23	Csp-51142.p-Csp-51142-4882	68	90	5.0d-5	91.3
4.	Seq1	1	23	Npun-73102.p-Npun_R6310	79	101	7.0d-4	78.26

SQ13. How could I confirm I had the right three sequences?

1> (Hinf-KW20.p-HI1041 Nlac-ST-640.p-Nlac-ST-640-0593 Csp-51142.p-Csp-51142-4882 Npun-73102.p-Npun_R6310)


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
Cut "DEFINE".
Help:
SEQUENCE-SIMILAR-TO
each
"PDLIMGGPPCQDFSSAGKRD..."
PROTEIN-VS-PROTEIN
Options
DEFINE
c-methyltransferases
=
JOIN
'(hinf-kw20.p-hi1041 nlac-st-6...
item
More...
Options
Options
Motif IV
M.BsaHI/1-320 65 P D M I I G G P P C Q D Y S S A G K R K - - E D
M.NpuORFC228P/1-331 79 P E I I I G G P P C Q D F S S A G K - R D E N L
M.SplZORFNP/1-314 65 P E V I I G G P P C Q D F S S A G K - R D E S L
M.HindV/1-304 66 P D L I M G G P P C Q D F S S A G K - R D I S L
M.NlaCORFDP/1-330 66 P D L I M G G P P C Q D F S S A G K - R D V T L
M.HgiDI/1-309 66 P E L I I G G P P C Q D F S S A G K - R D E G L
M.HaeIII/1-330 63 - D G I I G G P P C Q S W S E G G S L R G I D D
I now had three of their seven proteins, but I wanted at least a couple more. I defined a set of cytosine methyltransferases that would eventually consist of the three I found plus two more. To start, I dragged in the result of my search and then edited out the fourth sequence not in the Neely & Roberts collection.
Now I needed to get the other two (I choose M.BsaHI and M.HaeIII)…
1> (Hinf-KW20.p-HI1041 Nlac-ST-640.p-Nlac-ST-640-0593 Csp-51142.p-Csp-51142-4882 Npun-73102.p-Npun_R6310)

NCBI Resources How To My NCBI Sign In

NCBI
National Center for Biotechnology Information

Protein M.BsaHI Search

NCBI Home
Resource List (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

Get Started

- Tools: Analyze data using NCBI soft
- Downloads: Get NCBI data or softw
- How-To's: Learn how to accomplish
- Submissions: Submit data to GenBa

Genomic Structural Variation

dbVar archives large scale genomic variation data and associates defined variants with phenotypic information.

" 1 2 3 4 5 6 7 8

Popular Resources

PubMed

Bookshelf

PubMed Central

Genetic Testing Registry (GTR)

There are many places one can find protein sequences. The NCBI site is amongst the most general.

I went to the course web site, ***Resources & Links****, and followed the link to NCBI.*

There I chose the Protein database, and specified M.BsaHI (the name given by Neely & Roberts).

NCBI Resources How To My NCBI Sign In

Protein

Protein Search

Limits Advanced Help

Display Settings: GenPept Send to: Change region shown

BsaHI DNA methyltransferase [Geobacillus stearothermophilus]

Customize view

GenBank: ABY86226.1

Analyze this sequence

FASTA Graphics

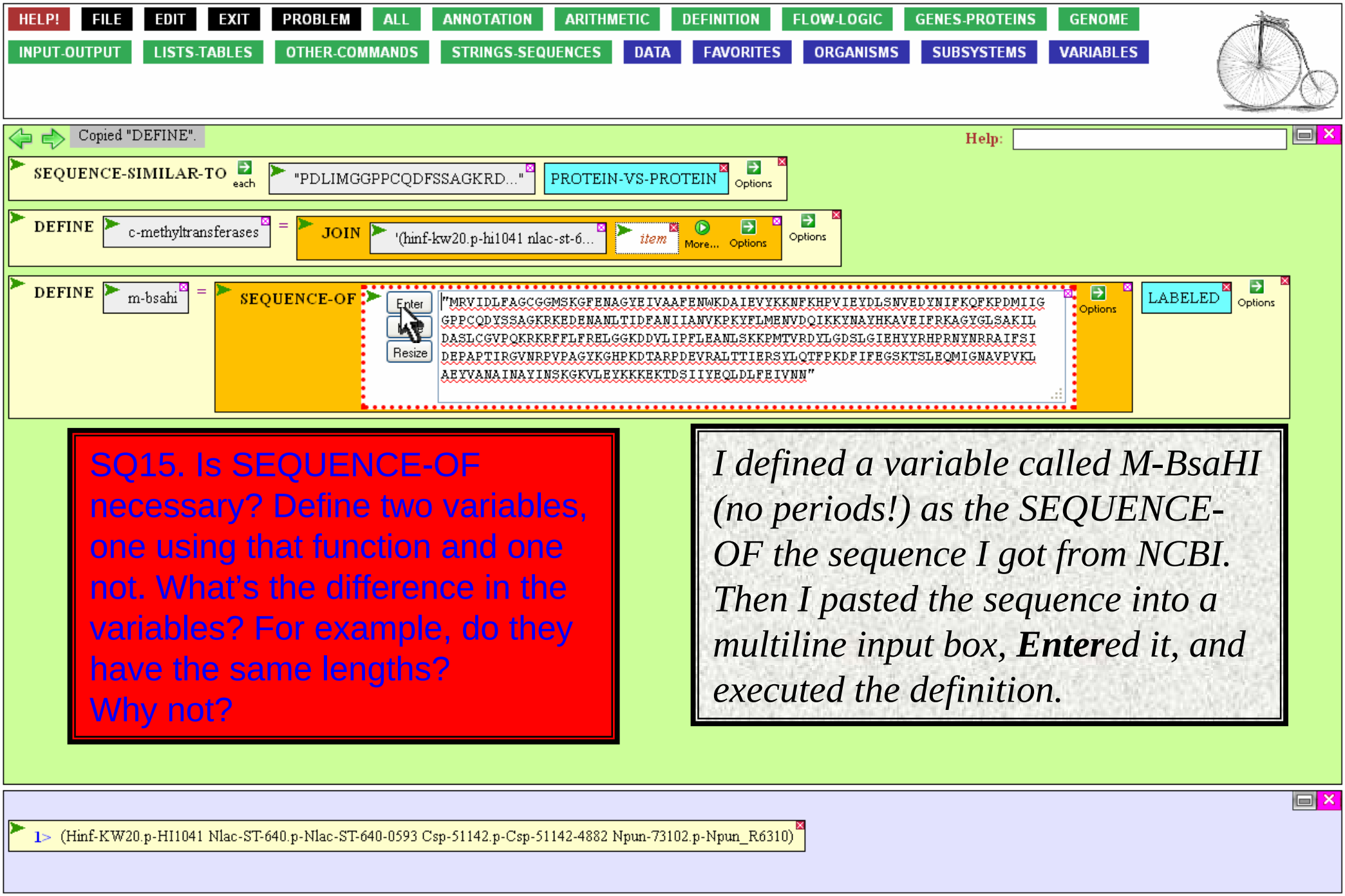
Go to:

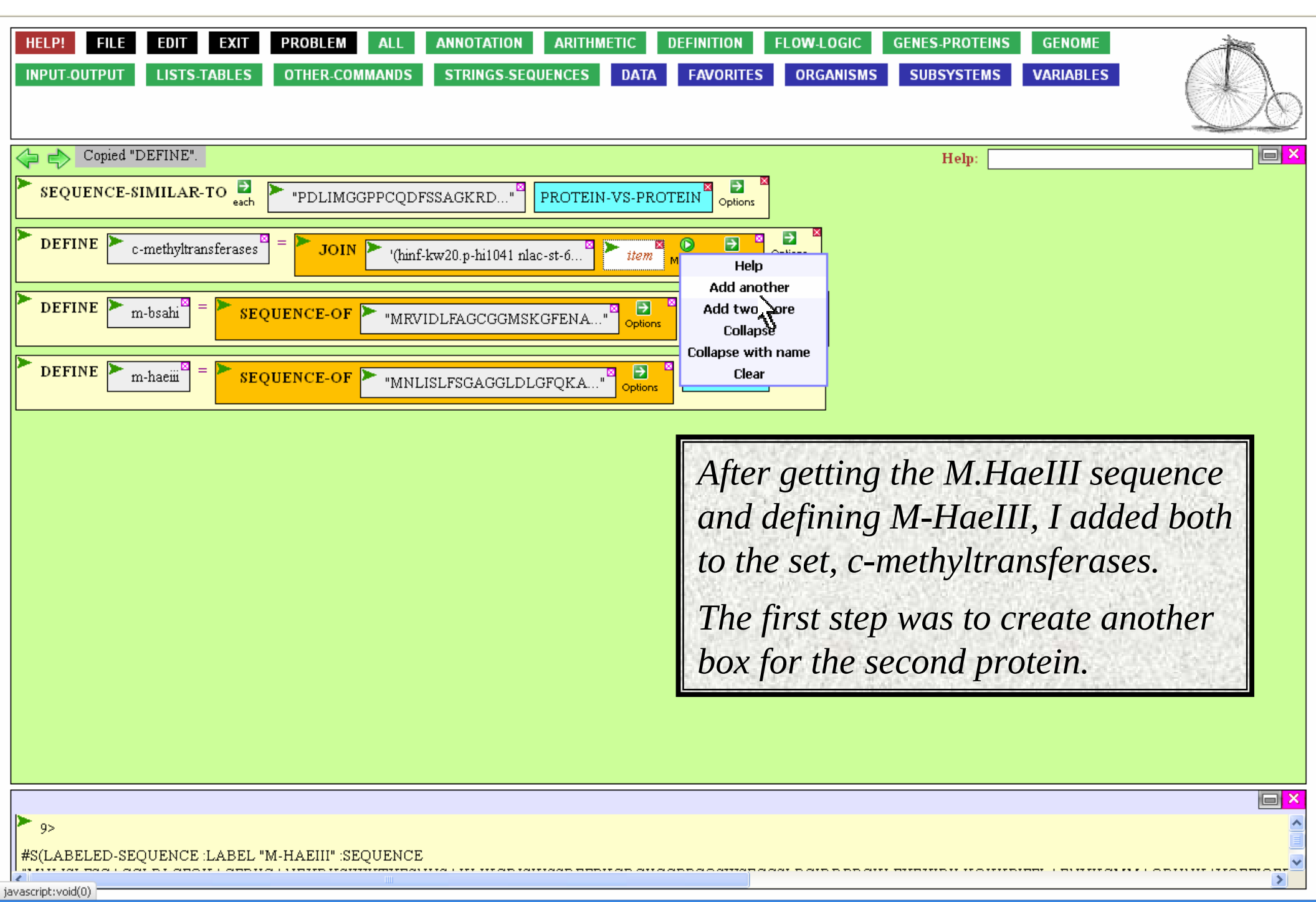
```
LOCUS       ABY86226                  323 aa
DEFINITION  BsaHI DNA methyltransferase [Geobacill
ACCESSION   ABY86226
VERSION     ABY86226.1  GI:166237008
DBSOURCE    accession EU386360.1
KEYWORDS    .
SOURCE      Geobacillus stearothermophilus
  ORGANISM  Geobacillus stearothermophilus
            Bacteria; Firmicutes; Bacillales; Baci
REFERENCE   1  (residues 1 to 323)
  AUTHORS   Neely,R.K. and Roberts,R.J.
  TITLE     The BsaHI restriction-modification system: cloning, s
            analysis of conserved motifs
  JOURNAL   BMC Mol. Biol. 9, 48 (2008)
   PUBMED   18479503
  REMARK    Publication Status: Online-Only
```

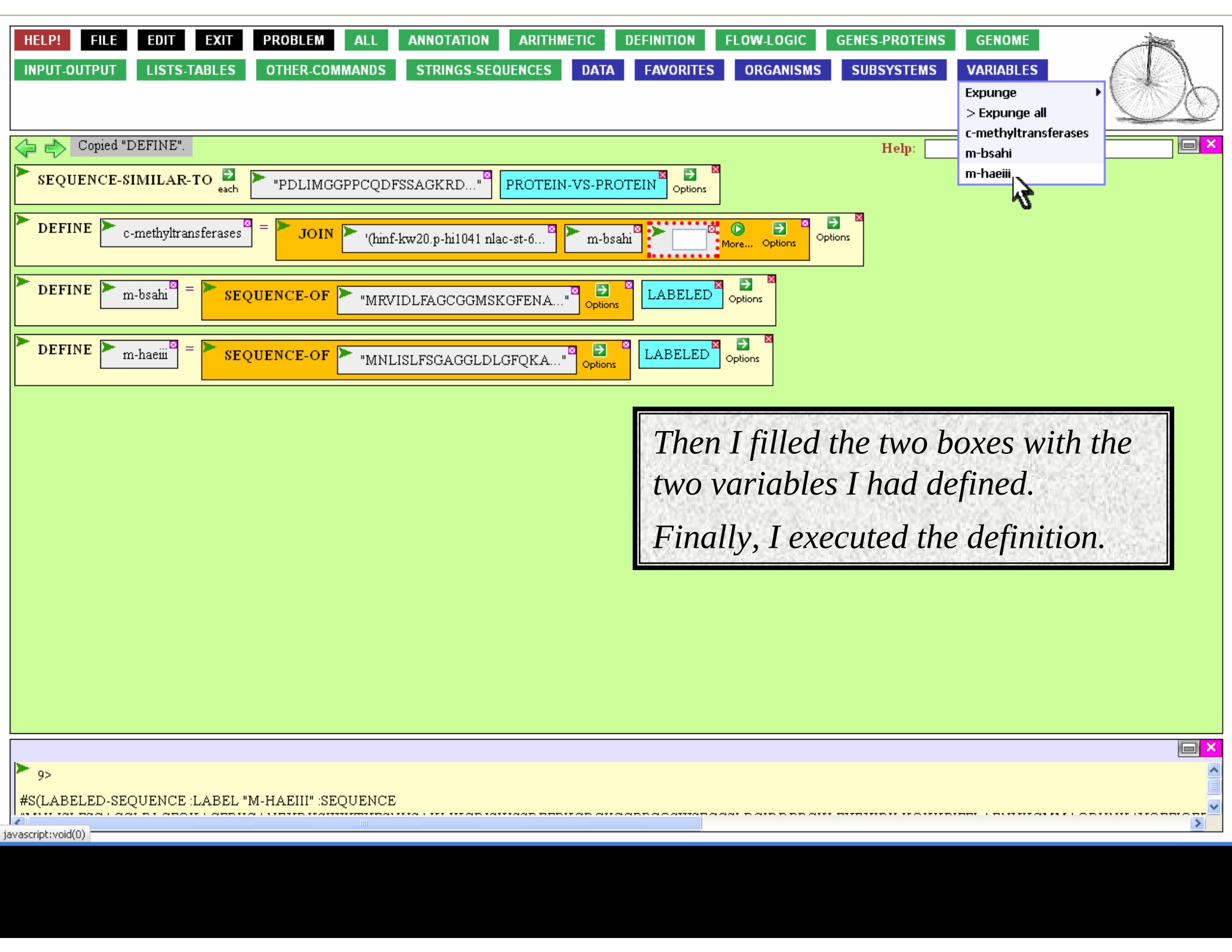
After verifying that this site had the right sequence, I wanted to copy the sequence and paste it into BioBIKE.

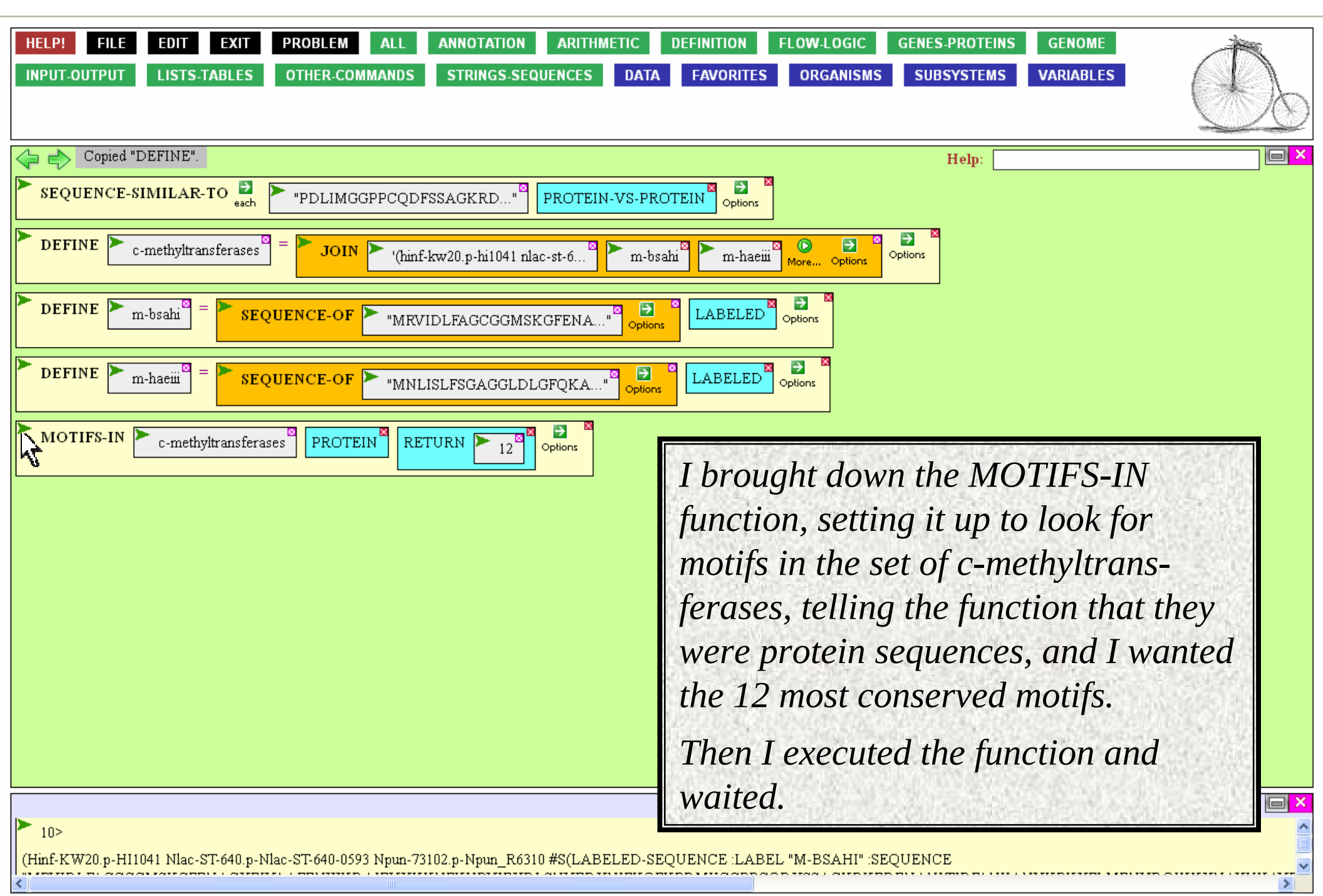
To get it into a format that has just sequence, I clicked FastA, copied the resulting sequence (not the header), and returned to BioBIKE.

SQ14. How did I verify that the site had the right sequence?


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO
each
"PDLIMGGPPCQDFSSAGKRD..."
PROTEIN-VS-PROTEIN
Options
DEFINE
c-methyltransferases
=
JOIN
'(hinf-kw20.p-hi1041 nlac-st-6...
item
More...
Options
Options
DEFINE
m-bsahi
=
SEQUENCE-OF
Enter
Resize
"MRVIDLFAGCGGMSKGFENAGYEIVAAFENWKDAIEVYKKNFKHPVIEYDLSNVEDYNIFKQFKPDMIIG
GPPCQDYSSAGKRKEDENANLTIDFANIIANVKPKYFLMENVDQIKKYNAYHKAVEIFRKAGYGLSAKIL
DASLCGVPQKRKRFFLFRELGGKDDVLIPFLEANLSKKPMTVRDYLGDSLGIEHVYRHPRNYNRRAIFSI
DEPAPTIRGVNRPVPAGYKGHPKDTARPDEVRALTTIERSYLQTFPKDFIFEGSKTSLEQMIGNAVPVKL
AEYVANAINAYINSKGKVLEYKKKEKTDSIIYEQLDLFEIVNN"
Options
LABELED
Options
SQ15. Is SEQUENCE-OF necessary? Define two variables, one using that function and one not. What's the difference in the variables? For example, do they have the same lengths?
Why not?
I defined a variable called M-BsaHI (no periods!) as the SEQUENCE-OF the sequence I got from NCBI. Then I pasted the sequence into a multiline input box, Entered it, and executed the definition.
1> (Hinf-KW20.p-HI1041 Nlac-ST-640.p-Nlac-ST-640-0593 Csp-51142.p-Csp-51142-4882 Npun-73102.p-Npun_R6310)


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO
each
"PDLIMGGPPCQDFSSAGKRD..."
PROTEIN-VS-PROTEIN
Options
DEFINE
c-methyltransferases
=
JOIN
'(hinf-kw20.p-hi1041 nlac-st-6...
item
Help
Add another
Add two more
Collapse
Collapse with name
Clear
DEFINE
m-bsahi
=
SEQUENCE-OF
"MRVIDLFAGCGGMSKGFENA..."
Options
DEFINE
m-haeiii
=
SEQUENCE-OF
"MNLISLFSGAGGLDLGFQKA..."
Options
After getting the M.HaeIII sequence and defining M-HaeIII, I added both to the set, c-methyltransferases.
The first step was to create another box for the second protein.
9>
#S(LABELED-SEQUENCE :LABEL "M-HAEIII" :SEQUENCE
javascript:void(0)


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Expunge
> Expunge all
c-methyltransferases
m-bsahi
m-haeiii
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO
each
"PDLIMGGPPCQDFSSAGKRD..."
PROTEIN-VS-PROTEIN
Options
DEFINE
c-methyltransferases
=
JOIN
'(hinf-kw20.p-hi1041 nlac-st-6...
m-bsahi
More...
Options
Options
DEFINE
m-bsahi
=
SEQUENCE-OF
"MRVIDLFAGCGGMSKGFENA..."
Options
LABELED
Options
DEFINE
m-haeiii
=
SEQUENCE-OF
"MNLISLFSGAGGLDLGFQKA..."
Options
LABELED
Options
Then I filled the two boxes with the two variables I had defined.
Finally, I executed the definition.
9>
#S(LABELED-SEQUENCE :LABEL "M-HAEIII" :SEQUENCE
javascript:void(0)


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO
each
"PDLIMGGPPCQDFSSAGKRD..."
PROTEIN-VS-PROTEIN
Options
DEFINE
c-methyltransferases
=
JOIN
'(hinf-kw20.p-hi1041 nlac-st-6...
m-bsahi
m-haeiii
More...
Options
Options
DEFINE
m-bsahi
=
SEQUENCE-OF
"MRVIDLFAGCGGMSKGFENA..."
Options
LABELED
Options
DEFINE
m-haeiii
=
SEQUENCE-OF
"MNLISLFSGAGGLDLGFQKA..."
Options
LABELED
Options
MOTIFS-IN
c-methyltransferases
PROTEIN
RETURN
12
Options
I brought down the MOTIFS-IN function, setting it up to look for motifs in the set of c-methyltrans-ferases, telling the function that they were protein sequences, and I wanted the 12 most conserved motifs.
Then I executed the function and waited.
10>
(Hinf-KW20.p-HI1041 Nlac-ST-640.p-Nlac-ST-640-0593 Npun-73102.p-Npun_R6310 #S(LABELED-SEQUENCE :LABEL "M-BSAHI" :SEQUENCE

MEME - Motif discovery tool

MEME version 3.0 (Release date: 2002/04/02 00:11:59)

For further information on how to interpret these results or to get a copy of the MEME software please access http://meme.sdsc.edu.

This file may be used as input to the MAST algorithm for searching sequence databases for matches to groups of motifs. MAST is available for interactive use and downloading at http://meme.sdsc.edu.

REFERENCE

If you use this program in your research, please cite:

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press, Menlo Park, California, 1994.

TRAINING SET

```
DATAFILE= /home/biobike/tmp/JELHAI-3540104660-2864.fa
ALPHABET= ACDEFGHIKLMNPQRSTVWY
```

Sequence name	Weight	Length	Sequence name	Weight	Length
Hinf-KW20.p-HI1041	1.0000	304	Nlac-ST-640.p-Nlac-ST-64	1.0000	330
Npun-73102.p-Npun_R6310	1.0000	331	M-HAEIII	1.0000	330
M-BSAHI	1.0000	323			

COMMAND LINE

```
This information can also be useful in the event you wish to report a
problem with the MEME software.

command: meme -nostatus -nmotifs 12 -mod zoops -protein /home/biobike/tmp/JELHAI-3540104

model:  mod=         zoops    nmotifs=        12    evt=           inf
object function=  E-value of product of p-values
width:  minw=            8    maxw=           50    minic=        0.00
width:  wg=             11    ws=              1    endgaps=       yes
nsites: minsites=        2    maxsites=        5    wnsites=       0.8
```

MOTIFS-IN calls Meme (just as SEQUENCE-SIMILAR-TO calls Blast).

First I confirmed that Meme considered all five of the sequences I gave it, then I scrolled down to the end of the output…

Combined block diagrams: non-overlapping sites with p-value < 0.0001

Name	Combined p-value	Motifs
Hinf-KW20.p-HI1041	6.56e-253	3 7 5 6 4 8 1 10 2
Nlac-ST-640.p-Nlac-ST-64	7.21e-272	3 7 5 6 4 8 1 10 2 11
Npun-73102.p-Npun_R6310	4.05e-237	3 7 5 6 4 8 1 12 2 9
M-HAEIII	1.50e-84	3 9 5 6 4 11 12 2
M-BSAHI	3.73e-224	3 7 5 6 4 8 1 2
SCALE		1 25 50 75 100 125 150 175 200 225 250 275 300 325 350

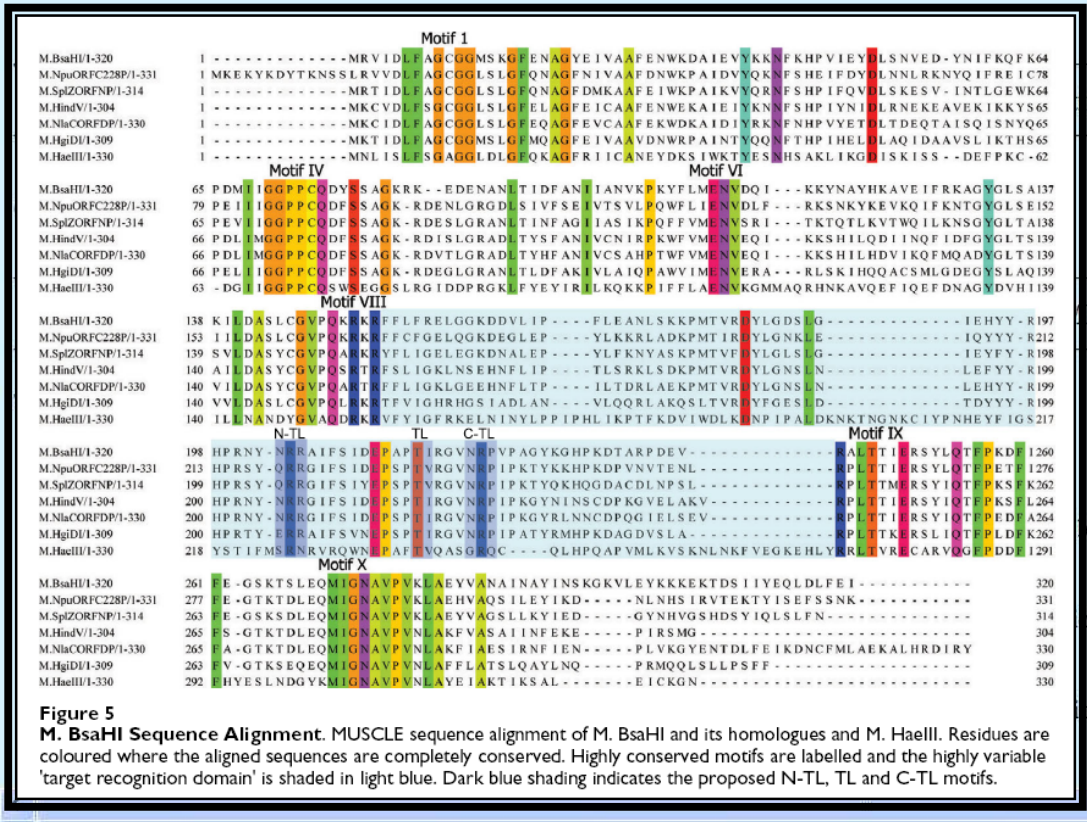


Figure 5
M. BsaHI Sequence Alignment. MUSCLE sequence alignment of M. BsaHI and its homologues and M. HaeIII. Residues are coloured where the aligned sequences are completely conserved. Highly conserved motifs are labelled and the highly variable 'target recognition domain' is shaded in light blue. Dark blue shading indicates the proposed N-TL, TL and C-TL motifs.

…where I found a summary of the motifs found. The motifs are number in the order they were found (most conserved to least conserved).

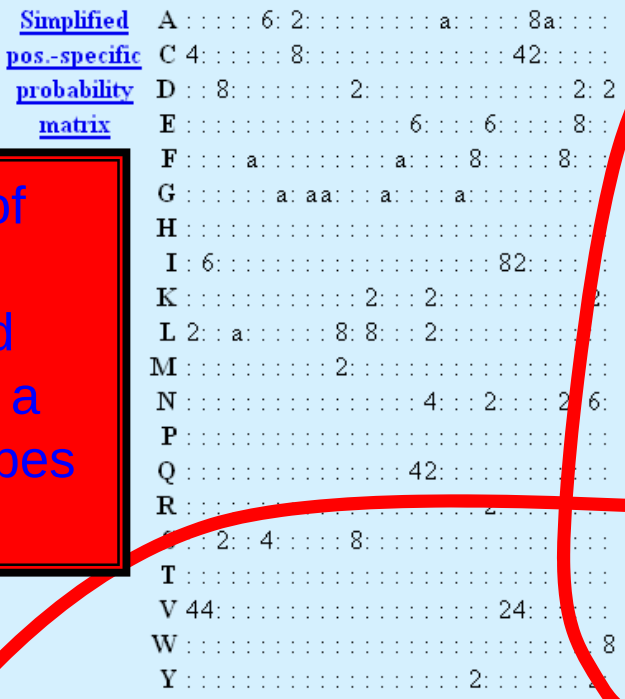
Meme's numbering has nothing to do with the motif numbering of Neely & Roberts. Meme called the first motif (geographically) Motif 3.

I scrolled up to the section of the output concerning Motif 3.

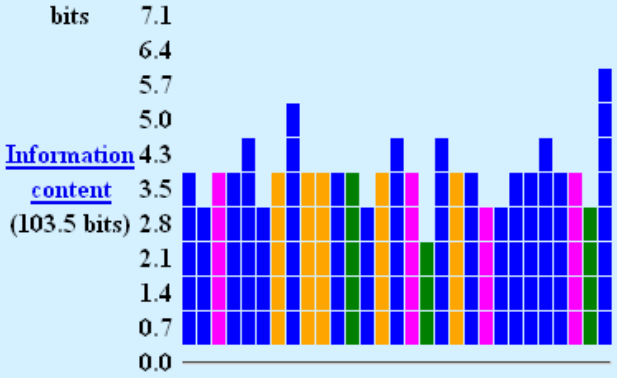
PN

MOTIF 3 width = 29 sites = 5 llr = 359 E-value = 2.7e-046

Simplified pos.-specific probability matrix



SQ16. The E-value of Motif 3 is given as $2.7x10^{-46}$. How would you use this value in a sentence that describes what it means?



Multilevel consensus sequence

```
CIDLFAGCGGLSLGFENAGFEICAAFENW
VVS  S A  MDK  QK  YNVVC NDKD
L              L   R I    Y
               Q
```

*This section tells me about the motif Meme found. The **consensus sequence** is the most common amino acids at each position in the motif.*

*The **E-value** is defined similarly to the E-value of Blast.*

0.0

Multilevel consensus sequence	CIDLFAGCGGLSLGFENAGFEICAAFENW VVS S A MDK QK YNVVC NDKD L L R I Y Q

NAME	START	P-VALUE	SITES
Hinf-KW20.p-HI1041	3	3.46e-35	MK CVDLFSGCGGLSLGFELAGFEICAAFENW EKAIEIYKNN
Nlac-ST-640.p-Nlac-ST-64	3	1.79e-34	MK CIDLFAGCGGLSLGFEQAGFEVCAAFEKW DKAIDIYRKN
Npun-73102.p-Npun_R6310	16	1.43e-32	KDYTKNSSLR VVDLFAGCGGLSLGFQNAGFNIVAAFDNW KPAIDVYQKN
M-BSAHI	3	1.85e-32	MR VIDLFAGCGGMSKGFENAGYEIVAAFENW KDAIEVYKKN
M-HAEIII	3	5.03e-22	MN LISLFSGAGGLDLGFQKAGFRIICANEYD KSIWKTYESN

SQ17. Are they? Check.

Motif 3 block diagrams

Name	Lowest p-value	Motifs
Hinf-KW20.p-HI1041	3.5e-35	3
Nlac-ST-640.p-Nlac-ST-64	1.8e-34	3
Npun-73102.p-Npun_R6310	1.4e-32	3
M-BSAHI	1.8e-32	3
M-HAEIII	5e-22	3
SCALE		1 25 50 75 100 125 150 175 200 225

Scrolling down…

The section also provides the actual sequences for this motif in each of the five sequences.

I confirmed that these are the same sequences reported by Neely & Roberts?

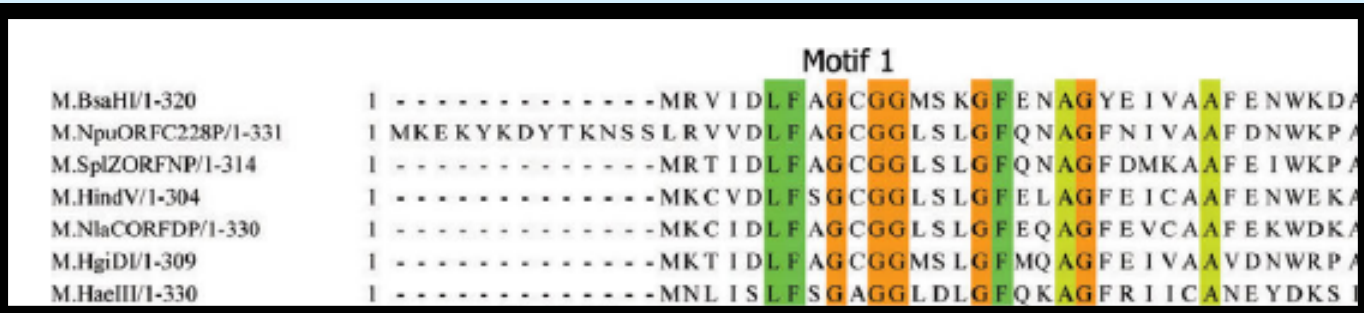
Motif 3 in BLOCKS format

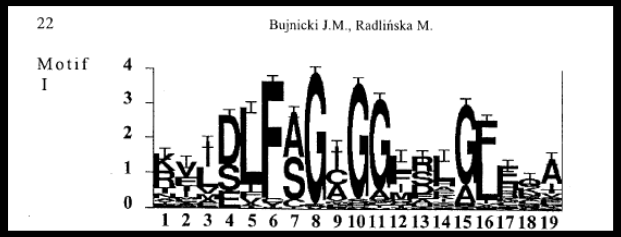
```
BL   MOTIF 3 width=29 seqs=5
Hinf-KW20.p-HI1041       (    3) CVDLFSGCGGLSLGFELAGFEICAAFENW  1
Nlac-ST-640.p-Nlac-ST-64 (    3) CIDLFAGCGGLSLGFEQAGFEVCAAFEKW  1
Npun-73102.p-Npun_R6310  (   16) VVDLFAGCGGLSLGFQNAGFNIVAAFDNW  1
M-BSAHI                  (    3) VIDLFAGCGGMSKGFENAGYEIVAAFENW  1
M-HAEIII                 (    3) LISLFSGAGGLDLGFQKAGFRIICANEYD  1
//
```

0.0

Multilevel consensus sequence	CIDLFAGCGGLSLGFENAGFEICAAFENW VVS S A MDK QK YNVVC NDKD L L R I Y Q

NAME	START	P-VALUE	SITES
Hinf-KW20.p-HI1041	3	3.46e-35	MK CVDLFSGCGGLSLGFELAGFEICAAFENW EKAIEIYKNN
Nlac-ST-640.p-Nlac-ST-64	3	1.79e-34	MK CIDLFAGCGGLSLGFEQAGFEVCAAFEKW DKAIDIYRKN
Npun-73102.p-Npun_R6310	16	1.43e-32	KDYTKNSSLR VVDLFAGCGGLSLGFQNAGFNIVAAFDNW KPAIDVYQKN
M-BSAHI	3	1.85e-32	MR VIDLFAGCGGMSKGFENAGYEIVAAFENW KDAIEVYKKN
M-HAEIII	3	5.03e-22	MN LISLFSGAGGLDLGFQKAGFRIICANEYD KSIWKTYESN





SQ18. Find a similar equivalence with another motif found by MOTIFS-IN.

No doubt about it. Meme's Motif 3 is the same as Neely & Roberts and Bujnicki & Radlinska's Motif I.

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.E. Seek identifiable features in phage genes, particularly those without annotation

Once you've convinced yourself that you can find protein sequence motifs that have been described as important in the functioning of your favorite protein class, add proteins you have reason to believe may belong to that class and rerun MOTIFS-IN. You can get candidate proteins by a variety of means, e.g., by their provisional annotation or by sequence similarity. However you may find them, add them to the list of proteins of proven function and determine if the candidates have all of the sequence motifs typical of the class.

1. *Is modification of DNA by cytosine methyltransferases a common strategy employed by phages?*
2. *Have the genes for cytosine methyl-transferases moved amongst phages by horizontal gene transfer?*
3. *If they have, what enables them to do so?*

So I've proved to myself that I can find motifs in proteins that have been shown by others to be cytosine methyltransferases.

Does that get me any farther in my project?

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.E. Seek identifiable features in phage genes, particularly those without annotation

Once you've convinced yourself that you can find protein sequence motifs that have been described as important in the functioning of your favorite protein class, add proteins you have reason to believe may belong to that class and rerun MOTIFS-IN. You can get candidate proteins by a variety of means, e.g., by their provisional annotation or by sequence similarity. However you may find them, add them to the list of proteins of proven function and determine if the candidates have all of the sequence motifs typical of the class.

1. Is modification of DNA by cytosine methyltransferases a common strategy employed by phages?

I re-examine my first question…

"Common strategy" seemed difficult to address, but I saw now how I could translate this into something within my grasp…

BNFO301: Introduction to Bioinformatics

Advice on Pursuing a Phage Genome Research Project

I.E. Seek identifiable features in phage genes, particularly those without annotation

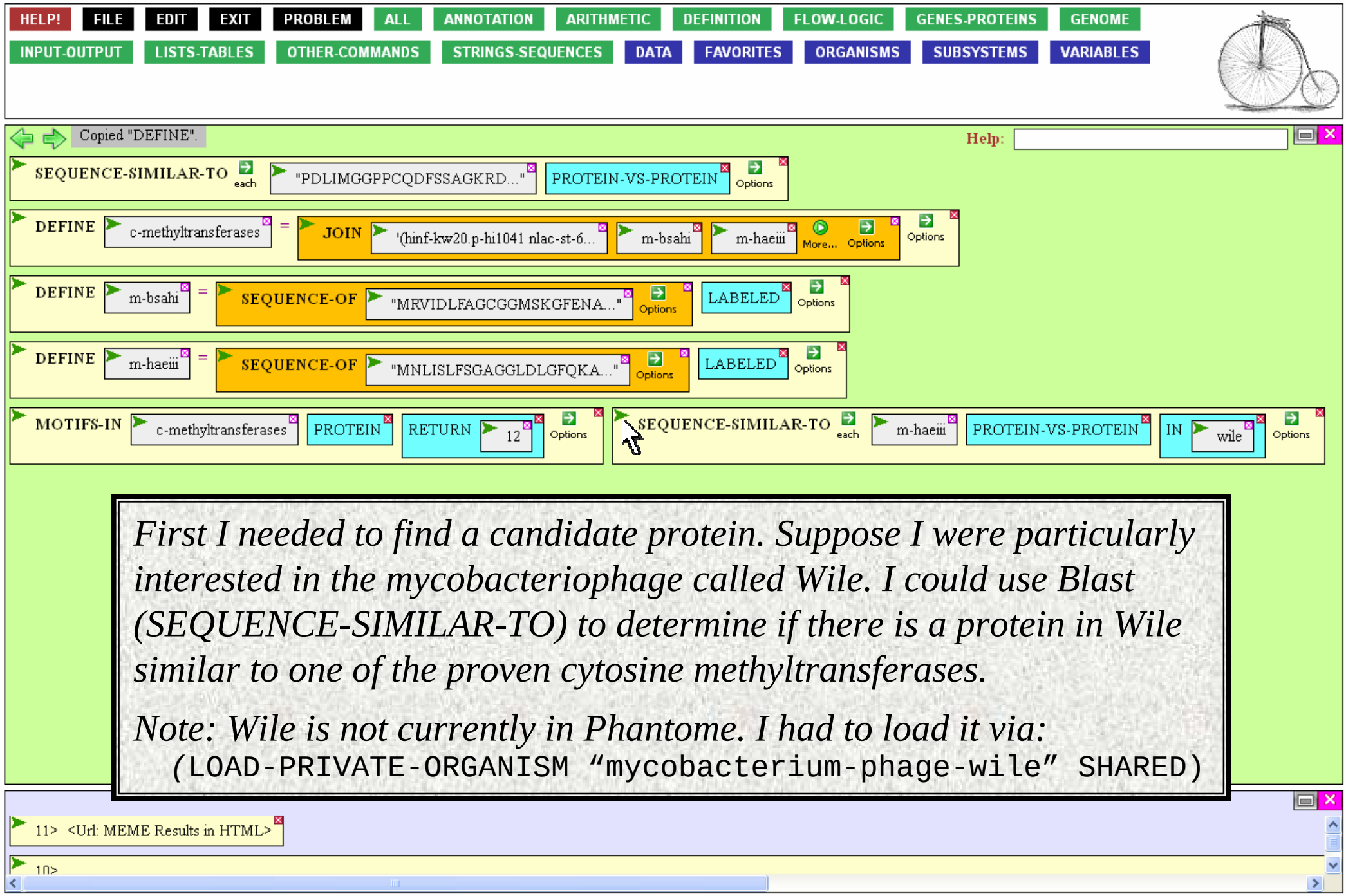
Once you've convinced yourself that you can find protein sequence motifs that have been described as important in the functioning of your favorite protein class, add proteins you have reason to believe may belong to that class and rerun MOTIFS-IN. You can get candidate proteins by a variety of means, e.g., by their provisional annotation or by sequence similarity. However you may find them, add them to the list of proteins of proven function and determine if the candidates have all of the sequence motifs typical of the class.

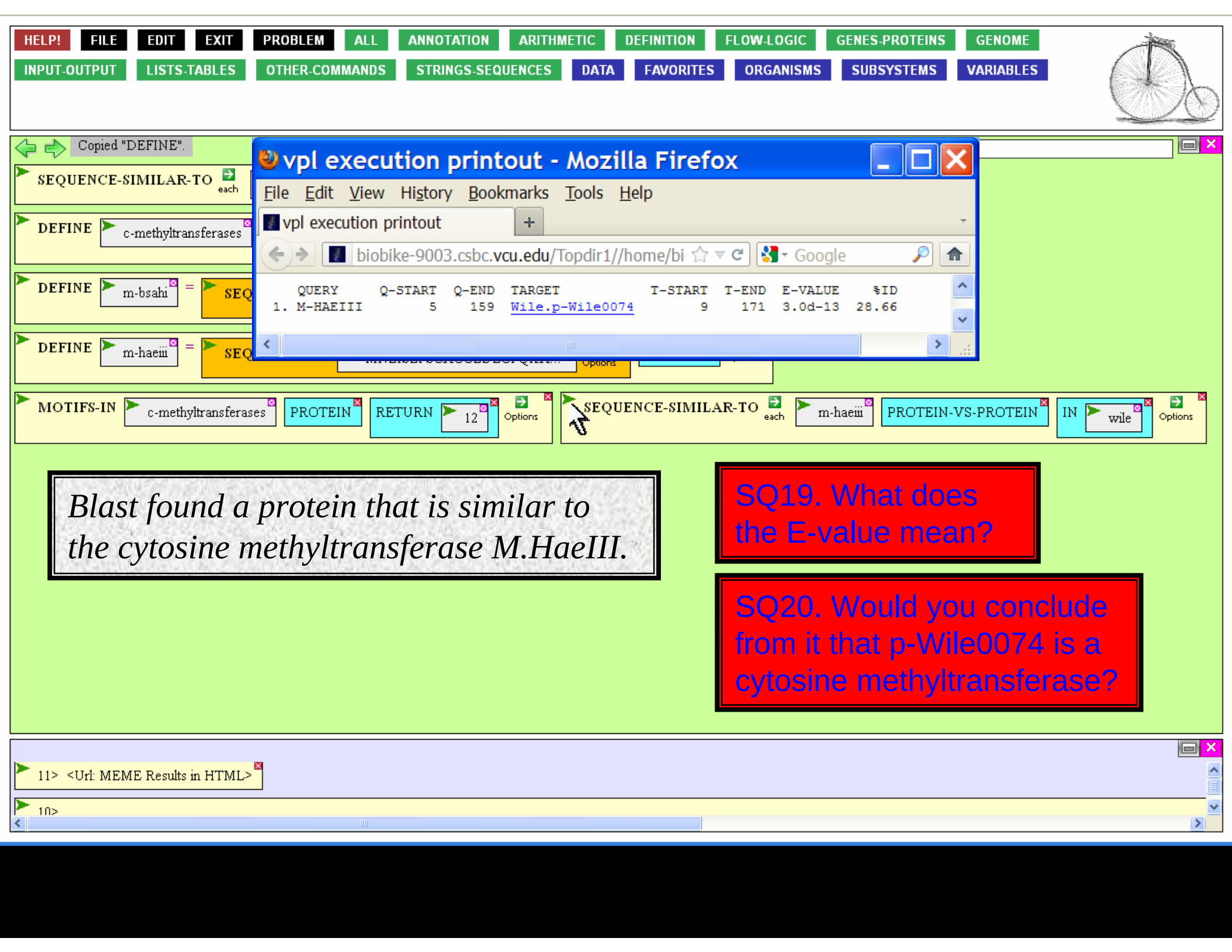
1. Is modification of DNA by cytosine methyltransferases a common strategy employed by phages?

1. Are genes encoding proteins with all the universal motifs of cytosine methyltransferases commonly found in phages?

This new version posed a question that I could answer, because now I had an operational definition of a modification gene.

But for this to work, I needed to find phage proteins whose motifs I could examine. It didn't seem likely I could do this visual analysis with every protein of every phage!


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO each "PDLIMGGPPCQDFSSAGKRD..." PROTEIN-VS-PROTEIN Options
DEFINE c-methyltransferases = JOIN '(hinf-kw20.p-hi1041 nlac-st-6... m-bsahi m-haeiii More... Options Options
DEFINE m-bsahi = SEQUENCE-OF "MRVIDLFAGCGGMSKGFENA..." Options LABELED Options
DEFINE m-haeiii = SEQUENCE-OF "MNLISLFSGAGGLDLGFQKA..." Options LABELED Options
MOTIFS-IN c-methyltransferases PROTEIN RETURN 12 Options
SEQUENCE-SIMILAR-TO each m-haeiii PROTEIN-VS-PROTEIN IN wile Options
First I needed to find a candidate protein. Suppose I were particularly interested in the mycobacteriophage called Wile. I could use Blast (SEQUENCE-SIMILAR-TO) to determine if there is a protein in Wile similar to one of the proven cytosine methyltransferases.
Note: Wile is not currently in Phantome. I had to load it via:
(LOAD-PRIVATE-ORGANISM "mycobacterium-phage-wile" SHARED)
11> <Url: MEME Results in HTML>
10>



vpl execution printout - Mozilla Firefox

File Edit View History Bookmarks Tools Help

vpl execution printout

biobike-9003.csbc.vcu.edu/Topdir1//home/bi

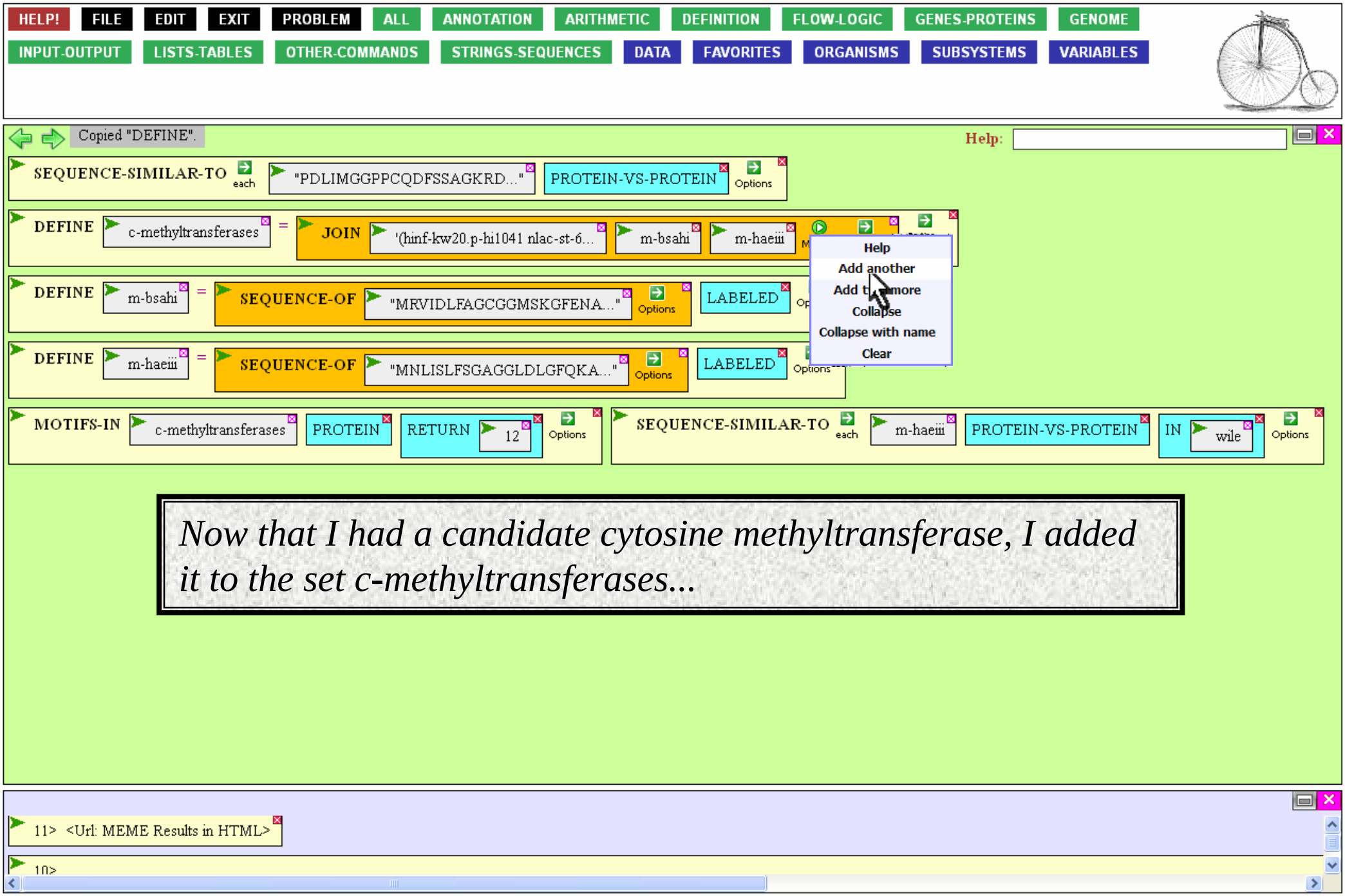
Google

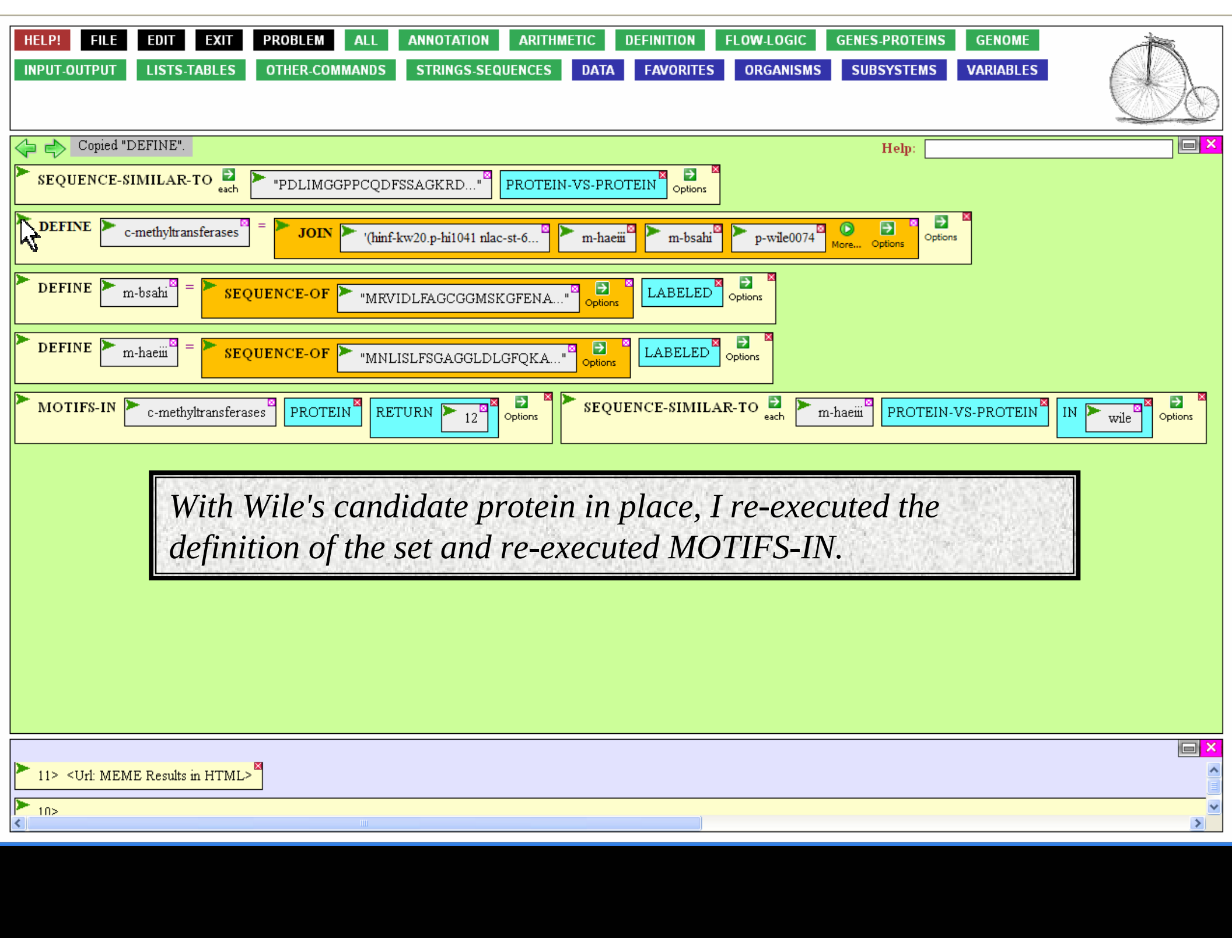
	QUERY	Q-START	Q-END	TARGET	T-START	T-END	E-VALUE	%ID
1.	M-HAEIII	5	159	Wile.p-Wile0074	9	171	3.0d-13	28.66

Blast found a protein that is similar to the cytosine methyltransferase M.HaeIII.

SQ19. What does the E-value mean?

SQ20. Would you conclude from it that p-Wile0074 is a cytosine methyltransferase?


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO each "PDLIMGGPPCQDFSSAGKRD..." PROTEIN-VS-PROTEIN Options
DEFINE c-methyltransferases = JOIN '(hinf-kw20.p-hi1041 nlac-st-6... m-bsahi m-haeiii
Help
Add another
Add t more
Collapse
Collapse with name
Clear
DEFINE m-bsahi = SEQUENCE-OF "MRVIDLFAGCGGMSKGFENA..." Options LABELED
DEFINE m-haeiii = SEQUENCE-OF "MNLISLFSGAGGLDLGFQKA..." Options LABELED
MOTIFS-IN c-methyltransferases PROTEIN RETURN 12 Options
SEQUENCE-SIMILAR-TO each m-haeiii PROTEIN-VS-PROTEIN IN wile Options
Now that I had a candidate cytosine methyltransferase, I added it to the set c-methyltransferases...
11> <Url: MEME Results in HTML>
10>


HELP!
FILE
EDIT
EXIT
PROBLEM
ALL
ANNOTATION
ARITHMETIC
DEFINITION
FLOW-LOGIC
GENES-PROTEINS
GENOME
INPUT-OUTPUT
LISTS-TABLES
OTHER-COMMANDS
STRINGS-SEQUENCES
DATA
FAVORITES
ORGANISMS
SUBSYSTEMS
VARIABLES
Copied "DEFINE".
Help:
SEQUENCE-SIMILAR-TO each "PDLIMGGPPCQDFSSAGKRD..." PROTEIN-VS-PROTEIN Options
DEFINE c-methyltransferases = JOIN '(hinf-kw20.p-hi1041 nlac-st-6... m-haeiii m-bsahi p-wile0074 More... Options Options
DEFINE m-bsahi = SEQUENCE-OF "MRVIDLFAGCGGMSKGFENA..." Options LABELED Options
DEFINE m-haeiii = SEQUENCE-OF "MNLISLFSGAGGLDLGFQKA..." Options LABELED Options
MOTIFS-IN c-methyltransferases PROTEIN RETURN 12 Options
SEQUENCE-SIMILAR-TO each m-haeiii PROTEIN-VS-PROTEIN IN wile Options
With Wile's candidate protein in place, I re-executed the definition of the set and re-executed MOTIFS-IN.
11> <Url: MEME Results in HTML>
10>

Command line Training Set First Motif Summary of Motifs Termination Explanation

MEME - Motif discovery tool

MEME version 3.0 (Release date: 2002/04/02 00:11:59)

For further information on how to interpret these results or to get a copy of the M

This file may be used as input to the MAST algorithm for searching sequence data ve use and downloading at http://meme.sdsc.edu.

REFEREN

If you use this program in your research, please cite:

Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation ma Second International Conference on Intelligent Systems for Molecular Biology, p

Sure enough, Meme reported that it now was considering six proteins.

Well, this was the magic moment: Did p-Wile0074 have the same motifs as the proven cytosine methyltransferases?

I scrolled down to the bottom…

TRAINING SET

```
DATAFILE= /home/biobike/tmp/JELHAI-3540105687-9527.fa
ALPHABET= ACDEFGHIKLMNPQRSTVWY
```

Sequence name	Weight	Length	Sequence name	Weight	Length
Hinf-KW20.p-HI1041	1.0000	304	Nlac-ST-640.p-Nlac-ST-64	1.0000	330
Npun-73102.p-Npun_R6310	1.0000	331	M-HAEIII	1.0000	330
M-BSAHI	1.0000	323	P-WILE0074	1.0000	177

Time 7.05 secs.

P N

SUMMARY OF MOTIFS

Combined block diagrams: non-overlapping sites with p-value < 0.0001

SQ21. Considering the evidence you've seen, do you think wile0074 encodes a cytosine methyltransferase?

SQ22. How can you explain the structure of p-Wile0074?

Name	Combined p-value	Motifs
Hinf-KW20.p-HI1041	4.25e-249	3 7 5 6 4 8 1 9 2
Nlac-ST-640.p-Nlac-ST-64	3.43e-258	3 7 5 6 4 8 1 9 2
Npun-73102.p-Npun_R6310	2.93e-235	3 7 5 6 4 8 1 11 2 10
M-HAEIII	3.72e-85	3 10 5 6 4 12 11 2
M-BSAHI	8.98e-223	3 7 5 6 4 8 1 2
P-WILE0074	2.38e-51	3 12 5 6 4
SCALE		1 25 50 75 100 125 150 175 200 225 250 275 300 325 350

Well, that was surprising. I wasn't so concerned that p-Wile0074 lacks the motifs labeled here 7, 8, and 1. After all, M.HaeIII also lacks those motifs and it works fine. But everyone (so far) has motif #2,… except for p-Wile0074. Furthermore, it can't' have any motifs beyond #4, because the protein stops!

Welcome to

Introduction to Bioinformatics

Wednesday, 7 March

Genome Analysis

So the curtain comes down just as things are starting to become interesting. We'll return to our story in Part II, but before that let's step back and notice:

- The project was focused (eventually) on a question of scientific interest

- That question took shape gradually, becoming ever closer to something that could be answered with available tools

- At each stage, what could be checked was checked