

CRISPI : a CRISPr Interactive database - user manual

CRISPI : a CRISPr Interactive database - user manual

by Christine Rousseau

Published February, 3th, 2009

Table of Contents

1. Introduction	1
2. Main page	3
2.1. User Parameters	3
2.2. Consult the CRISPR Database	4
2.2.1. Database summary	4
2.2.2. Genome selection.....	4
2.3. BLAST a personal sequence against the CRISPR database	4
2.3.1. BLAST parameter selection.....	5
2.3.2. BLAST results.....	5
2.4. Find CRISPR in personal sequence	6
3. CRISPR database consultation	9
3.1. CRISPR list	9
3.2. CRISPR details	11

List of Figures

2-1. The CRISPI main page.....	3
2-2. The database summary.....	4
2-3. Select a genome.....	4
2-4. BLAST parameters	5
2-5. BLAST results.....	5
2-6. Find out CRISPRs	6
3-1. CRISPR consultation list	9
3-2. Link to NCBI	9
3-3. Graphical circular view of a genome.....	10
3-4. CRISPR details.....	11
3-5. Pygram image.....	12
3-6. Genome information view	12
3-7. Consensus weblogo image.....	13

Chapter 1

Introduction

We propose a web user interface to explore the CRISPR's database at various levels, a Blast interface to extract information from this database and the possibility to find CRISPR structures in a personal sequence.



Chapter 2

Main page

The main page of *CRISPI* contains three search forms divided into two sections: search parameters on the left and help messages on the right. Help messages are hidden until some values are typed in form's fields or the mouse is over the field or the question mark. Three search forms are available : the first one to consult the CRISPR database, the second to run BLAST on a personal sequence against the CRISPR database and the third to find CRISPR in a personal sequence.

Figure 2-1. The CRISPI main page.

The screenshot shows the CRISPI main page with the following sections:

- The World of CRISPR**: Includes a [User Manuel](#) link.
- User Parameter**: A form with an "Email (optional)" field.
- Consult CRISPR Database**: Contains a **Database Summary** table and a **Genome selection** form.
- Blast a personal sequence against CRISPR database**: A form for sequence analysis.
- Help**: A link to "Upload your sequence file."

Database Summary

	Genomes analysed	CRISPRs found	Genomes without CRISPR	Units found	Spacers found	Putatifis genes CAS
Archaea	53	291	4	5955	5664	948
Bacteria	731	1990	134	23123	21133	0

Last Update - 2008-11-04 11:41:16

Genome selection

- Accession number
- Name
- Select genome
- Taxonomy browser

Blast a personal sequence against CRISPR database

- Sequence type: DNA Protein
- Paste your sequence : (fasta format)
- Or give your sequence file name :
- against databank : spacers database
- Modify Blast parameters

2.1. User Parameters

In this area, you can enter an email address. It will be used by CRISPI to send you a message at the



end of the job. The email message contains a reference to a confidential result file that can be accessed for 10 days before deletion. Note that this option is only available for the BLAST and Find CRISPR functions : consulting the CRISPR database requires no computation and results are immediately available through the interactive viewer. Entering an email address is optional, but is recommended. Providing your email address ensures that CRISPI will always inform you on job termination.

2.2. Consult the CRISPR Database

Microbial genomes have been pre-processed for an efficient search of particular CRISPRs structures.

2.2.1. Database summary

This table summarizes the statistics of the database and provides a global overview of CRISPRs present in the database.

Figure 2-2. The database summary

Database Summary

	Genomes analysed	CRISPRs found	Genomes without CRISPR	Units found	Spacers found	Putatifs genes CAS
Archaea	53	291	4	5955	5664	948
Bacteria	731	1990	134	23123	21133	0

Last Update = 2008-11-04 11:41:16

2.2.2. Genome selection

Microbial genomes can be selected in four ways :

- By entering the accession number.
- By entering the genome name (or a part of it).
- By selecting a genome in the genome list (lexicographically sorted).
- By selecting a genome in the taxonomy browser.

Figure 2-3. Select a genome

Genome selection

Accession number
 Name
 Select genome
 Taxonomy browser

Once all parameters have been edited, click on the [Display Crisprs] button to go to the next page.



2.3. BLAST a personal sequence against the CRISPR database

2.3.1. BLAST parameter selection

The query sequence must be in Fasta format (DNA or Protein). One can either paste the query sequence into the input field or upload it from a file on the local machine (files with multiple sequences are allowed). BLAST can be launched against units or spacers. For protein sequence, a preferred substitution matrix must be specified. Default BLAST parameters have been optimized, be careful if you modify them.

Figure 2-4. BLAST parameters

Blast a personal sequence against CRISPR database

Sequence type : DNA Protein

Paste your sequence : (fasta format)

Or give your sequence file name :

against databank :

Matrix :

Modify Blast parameters

eValue :

Word size :

Gap open cost :

Gap extension cost :

Low complexity :

2.3.2. BLAST results

The BLAST result page is cross-linked with the CRISPR database.

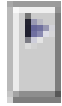


Figure 2-5. BLAST results

BLAST results

Your parameters :

- Databank : Spacers Database
- Sequence type : DNA
- eValue : 1
- Word size : 6
- Gap open cost : 1
- Gap extension cost : 2
- Low complexity : F

Your results :

- [Results BLAST \(html format\)](#)
- [Results BLAST \(text format\)](#)

BLASTN 2.2.15 [Oct-15-2006]

Reference:
 Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query= gi|51980126|ref|NC_006268.1| Sulfolobus virus STSV1, complete genome (75,294 letters)

Database:
 /home/genouest/bioinfo/Serveur-GPO/outils/repeatsAnalysis/CRISPR/spacersDBA
 5817 sequences; 233,100 total letters

Searching.....done

Sequences producing significant alignments:	Score	E
	(bits)	Value
NC_009440 Metallosphaera sedula DSM 5348 CRISPR no 2 spacer 25 from 1098879 to 1098922	38	0.045
NC_002754 Sulfolobus solfataricus P2 CRISPR no 4 spacer 13 from 1306414 to 1306455	38	0.045
NC_000033 <i>Saccharothrix reissii</i> F1 CRISPR no 1 spacer 6 from 221022 to 221070	26	0.17

List of Crispr for genome Metallosphaera sedula DSM 5348

Genome info
Metallosphaera sedula DSM 5348 Kingdom : Archaea
 RefSeq : NC_009440 [Graphical view of genome](#)
 Sequence size : 2191517
 5 CRISPR 20 putatives genes CAS

Genomic region around CRISPR 1

Select all Clear all Download selected genes

	BEGIN	END	NAME	STRAND	SUBTYPE	CONSENSUS		NB UNITS	eVALUE
						sequence	size		
<input type="checkbox"/>	1073299	1074216	Msed_1137 hypothetical protein	←	CRISPR-associated protein Cas6 (Cas6)				1.1e-07
<input type="checkbox"/>	1077185	1077943	Msed_1140 CRISPR-associated HD domain-containing protein	←	CRISPR-associated HD domain protein (Cas3)				2.7e-14
<input type="checkbox"/>	1077940	1079418	Msed_1141 CRISPR-associated helicase Cas3 family protein protein	←	CRISPR-associated helicase Cas3 (Cas3)				1.6e-41
<input type="checkbox"/>	1079402	1080127	Msed_1142 CRISPR-associated Cas5 family protein	←	CRISPR-associated protein Cas5, Apem subtype (Apem)				1.5e-36
<input type="checkbox"/>	1080127	1081098	Msed_1143 CRISPR-associated autoregulator DevR family protein protein	←	CRISPR-associated regulatory protein, Csa2 family (Apem)				6.2e-130
<input type="checkbox"/>	1081125	1081490	Msed_1144 CRISPR-associated RAMP Csa5 family protein	←	CRISPR-associated protein, Csa5 family (Apem)				8.4e-17
<input type="checkbox"/>			Msed_1145 CRISPR-associated DNA-binding		CRISPR Inrc-related DNA-binding				


2.4. Find CRISPR in personal sequence


The query sequence must be in Fasta format (only DNA sequences are allowed). One can either paste the query sequence into the input field or upload it from a file on the local machine (multiple sequences files are not allowed). Results are summarized in a table. Results can be notified by email (optional).



Figure 2-6. Find out CRISPRs

Find CRISPR in personal sequence

 Paste your sequence :
(fasta format)

 Or give your sequence file name :



Chapter 3

CRISPR database consultation

3.1. CRISPR list

Once a genome has been selected, results are summarized inside tables. Each CRISPR is highlighted and all *CAS* genes found in its vicinity are displayed. Annotations contain various elements such as positions, consensus unit, link to external NCBI information, link to genomic graphical circular view.

Figure 3-1. CRISPR consultation list

The World of CRISPR page 2 - List of CRISPR

List of Crispr for genome *Pyrococcus furiosus* DSM 3638

Genome info

Pyrococcus furiosus DSM 3638 Kingdom : Archaea
 RefSeq : NC_003413 [Graphical view of genome](#)
 Sequence size : 1908256
 6 CRISPR 30 putatives genes CAS

Genomic region around CRISPR 3

Select all Clear all Download selected genes

	BEGIN	END	NAME	STRAND	SUBTYPE	CONSENSUS		NB UNITS	eVALUE
						sequence	size		
<input type="checkbox"/>	611563	614166	PF0592 ATP-dependent RNA helicase, putative	←	CRISPR-associated helicase, Cyano-type (Unaffiliated)				6.1e-05
<input type="checkbox"/>	623119	625176	CRISPR repeat region			CTTTC AATTCTATTTGGTCTTATTGT AAC	30	31	
<input type="checkbox"/>	650200	651246	PF0637 hypothetical protein	←	CRISPR-associated protein, Csa4 family (Apern)				2.8e-218
<input type="checkbox"/>	651243	652391	PF0638 hypothetical protein	←	CRISPR-associated protein, Csa4 family (Apern)				1.9e-236
<input type="checkbox"/>	652404	653117	PF0639 hypothetical protein	←	CRISPR-associated HD domain protein (Cas3)				1.6e-51
<input type="checkbox"/>	653114	655090	PF0640 ATP-dependent RNA helicase, putative	←	CRISPR-associated HD domain protein (Cas3)				0.029
<input type="checkbox"/>	655060	655851	PF0641 hypothetical protein	←	CRISPR-associated protein Cas5, Apern subtype (Apern)				1.7e-34
<input type="checkbox"/>	655852	656877	PF0642 hypothetical protein	←	CRISPR-associated regulatory protein, Csa2 family (Apern)				3.5e-185
<input type="checkbox"/>	657364	658026	PF0644 hypothetical protein	→	CRISPR locus-related DNA-binding protein (Apern)				7.6e-13

select a line for more details

Genomic region around CRISPR 4

Select all Clear all Download selected genes

	BEGIN	END	NAME	STRAND	SUBTYPE	CONSENSUS		NB UNITS	eVALUE
						sequence	size		
<input type="checkbox"/>	676584	677936	PF0667 adenylosuccinate lyase	←	CRISPR-associated protein Cas6 (Cas6)				0.029
<input type="checkbox"/>	686783	688945	PF0677 ski2-like helicase	←	CRISPR-associated helicase Cas3 (Cas3)				0.00034



Figure 3-2. Link to NCBI

The World of CRISPR page 2 - List of CRISPR

List of Crispr for genome *Pyrococcus furiosus* DSM 3638

Genome info

Pyrococcus furiosus DSM 3638 Kingdom : Archaea

RefSeq : [NC_003413](#) **Graphical view of genome**

Sequence size : 1908256 30 putatives genes CAS

6 CRISPR

Genomic

Select all

NCBI **Nucleotide**

Search for

Display Show Send to Hide: sequence all but gene, CDS and mRNA feat

Range: from to Reverse complemented strand Features:

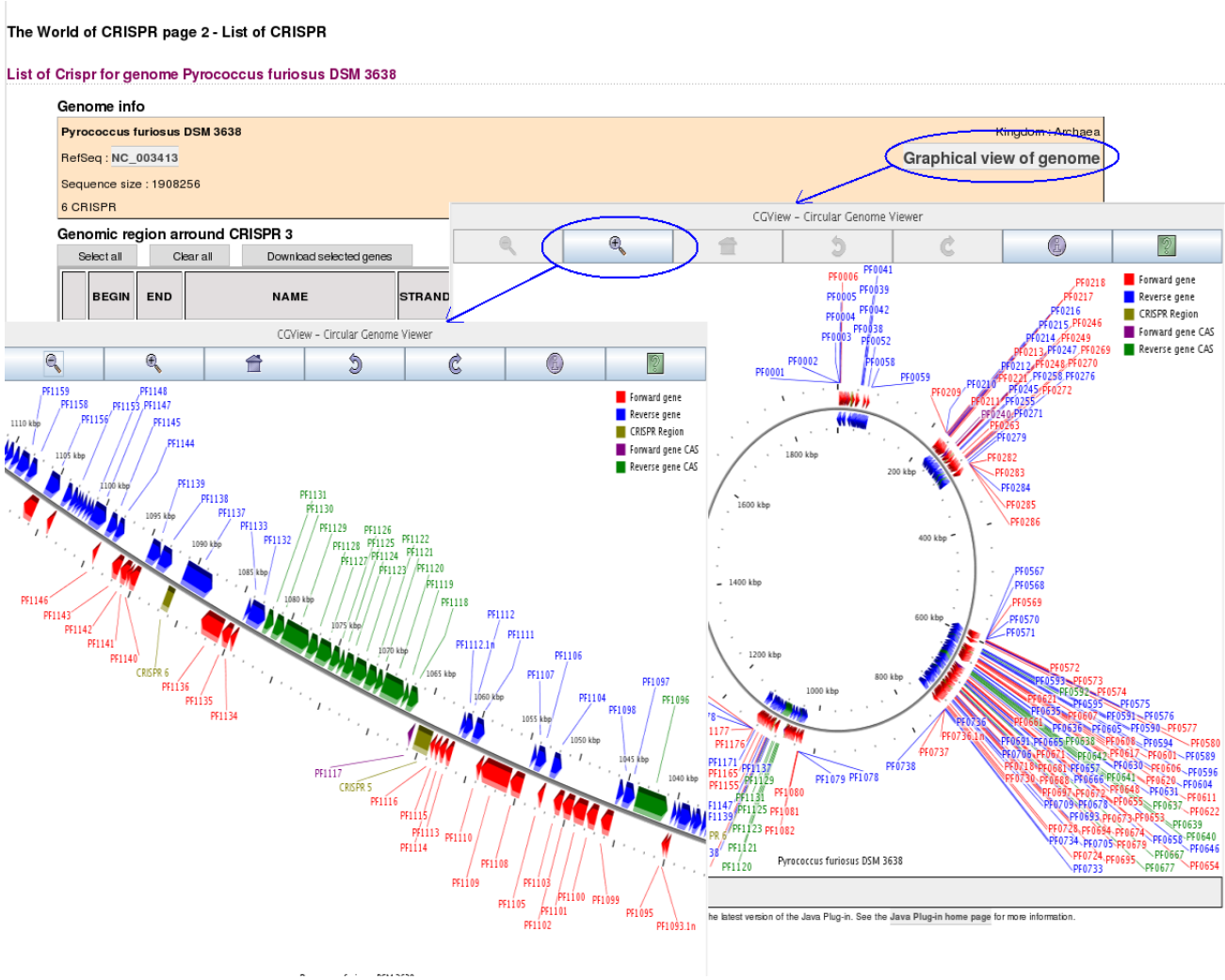
Accession	size	NB UNITS	eVALUE
TGGTCTTATTGTAAC	30	31	6.1e-05
This record may be shown in a reduced form. Use 'Hide' checkboxes for control.			
<input checked="" type="checkbox"/> 1: NC_003413 . Reports <i>Pyrococcus furiosus</i> ...[gi:18976372]			2.8e-218
Comment Features			1.9e-236
65240 LOCUS NC_003413 1908256 bp DNA circular BCT 23-JUL-2008			1.6e-51
65311 DEFINITION <i>Pyrococcus furiosus</i> DSM 3638, complete genome.			0.029
65506 ACCESSION NC_003413			
65506 VERSION NC_003413.1 GI:18976372			
65506 PROJECT GenomeProject:287			1.7e-34
65506 KEYWORDS .			
65506 SOURCE <i>Pyrococcus furiosus</i> DSM 3638			
65585 ORGANISM <i>Pyrococcus furiosus</i> DSM 3638			3.5e-185
65736 REFERENCE 1 (bases 1 to 1908256)			7.6e-13
AUTHORS Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D., Yeh,R.K., Weiss,R.B. and Dunn,D.M.			
TITLE Genomic sequence of hyperthermophile, <i>Pyrococcus furiosus</i> : implications for physiology and enzymology			
JOURNAL Meth. Enzymol. 330, 134-157 (2001)			
PUBMED 11210495			
REFERENCE 2 (bases 1 to 1908256)			
AUTHORS Diruggiero,J., Dunn,D., Maeder,D.L., Holley-Shanks,R., Chatard,J., Horlacher,R., Robb,F.T., Boos,W. and Weiss,R.B.			0.029
TITLE Evidence of recent lateral gene transfer among hyperthermophilic archaea			
JOURNAL Mol. Microbiol. 38 (4), 684-693 (2000)			
PUBMED 11115105			0.00034
REFERENCE 3 (bases 1 to 1908256)			

Genomic

Select all

Accession	size	NB UNITS	eVALUE
TGGTCTTATTGTAAC	30	31	6.1e-05
<input checked="" type="checkbox"/> 1: NC_003413 . Reports <i>Pyrococcus furiosus</i> ...[gi:18976372]			2.8e-218
Comment Features			1.9e-236
65240 LOCUS NC_003413 1908256 bp DNA circular BCT 23-JUL-2008			1.6e-51
65311 DEFINITION <i>Pyrococcus furiosus</i> DSM 3638, complete genome.			0.029
65506 ACCESSION NC_003413			
65506 VERSION NC_003413.1 GI:18976372			
65506 PROJECT GenomeProject:287			1.7e-34
65506 KEYWORDS .			
65506 SOURCE <i>Pyrococcus furiosus</i> DSM 3638			
65585 ORGANISM <i>Pyrococcus furiosus</i> DSM 3638			3.5e-185
65736 REFERENCE 1 (bases 1 to 1908256)			7.6e-13
AUTHORS Robb,F.T., Maeder,D.L., Brown,J.R., DiRuggiero,J., Stump,M.D., Yeh,R.K., Weiss,R.B. and Dunn,D.M.			
TITLE Genomic sequence of hyperthermophile, <i>Pyrococcus furiosus</i> : implications for physiology and enzymology			
JOURNAL Meth. Enzymol. 330, 134-157 (2001)			
PUBMED 11210495			
REFERENCE 2 (bases 1 to 1908256)			
AUTHORS Diruggiero,J., Dunn,D., Maeder,D.L., Holley-Shanks,R., Chatard,J., Horlacher,R., Robb,F.T., Boos,W. and Weiss,R.B.			0.029
TITLE Evidence of recent lateral gene transfer among hyperthermophilic archaea			
JOURNAL Mol. Microbiol. 38 (4), 684-693 (2000)			
PUBMED 11115105			0.00034
REFERENCE 3 (bases 1 to 1908256)			

Figure 3-3. Graphical circular view of a genome.



3.2. CRISPR details

Clicking on the consensus link takes the user to the corresponding CRISPR's details. Various informations are displayed such as unit and spacer coordinates, unit and spacer sequences, Pygram image and consensus WebLogo image.



Figure 3-4. CRISPR details

Crispr detail

		Select all units	Select all spacers	Clear all	Download selected data	Extract flanking sequences	Extract CRISPR sequence
		BEGIN	END	SEQUENCE			SIZE
<input type="checkbox"/>	unit	260714	260743	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	260744	260782	TGTTCAATCGCACTTCTTCTTGACTCTGCTCCACTTAG			39
<input type="checkbox"/>	unit	260783	260812	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	260813	260849	GCGTTAATGAACAATAAGCCTGACACGAACATAAAACA			37
<input type="checkbox"/>	unit	260850	260879	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	260880	260916	TTTCAATAAGGAGTTTCAAATGGCTCGATGGAATTAT			37
<input type="checkbox"/>	unit	260917	260946	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	260947	260988	AGAACCTCGCAATCTCCTTGGAGAAAGGCCCTTCTGAGA			42
<input type="checkbox"/>	unit	260989	261018	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261019	261056	CTTCTCGAAGTCGTAGTTTAGTGTGCAAGATATTCT			38
<input type="checkbox"/>	unit	261057	261086	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261087	261123	TTCTAGAAGTTCTTTCGAGAGCCAAGAGCGCGGCT			37
<input type="checkbox"/>	unit	261124	261153	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261154	261190	GTGTAATAAGAATTTCTTCATAACTTTTCTGAAGA			37
<input type="checkbox"/>	unit	261191	261220	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261221	261258	ACATAAGAGACGGAGAGCTTGTATCTGGAAAGCAAAT			38
<input type="checkbox"/>	unit	261259	261288	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261289	261331	AGTTATAACAAACCAGCATGACCCGCTCTCGGCTGAGTGCCAG			43
<input type="checkbox"/>	unit	261332	261361	GTTCCAATAAGACTACAAAAGAATTGAAAG			30
<input type="checkbox"/>	spacer	261362	261398	CTACCATGGCCATCACCATAATTAATTGTAAGTTAG			37
<input type="checkbox"/>	unit	261399	261428	GTTCCAATAAGACTACAAAAGAATTGAAAG			30

Figure 3-5. Pygram image

The World of CRISPR page 3 - Crispr's details

Pyrococcus furiosus DSM 3638 (from 260714 to 262113)

Pygram view

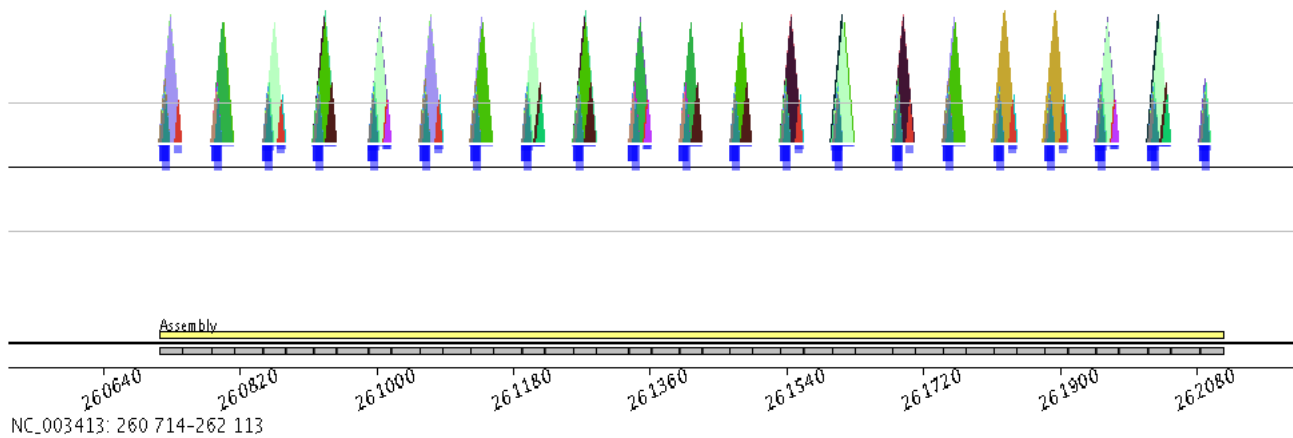




Figure 3-6. Genome information view

Crispr info

Pyrococcus furiosus DSM 3638	Kingdom : Archaea
RefSeq : NC_003413	
Consensus : GTTCCAATAAGACTACAAAAGAATTGAAAG (30 bp)	21 units
Begin position 260714	End position 262113

Figure 3-7. Consensus weblogo image

Consensus view

