



Plateforme de Bioinformatique
de l'Université Paris-Sud

The CRISPR webserver: a database and tools
for the identification and the characterisation of
CRISPRs and their associated genes



Principales ressources eBio

Tool	Description	Unique visitors/yr	Citations (last 2 yr)	URL
CRISPRdb	Resources for CRISPRs found in prokaryote genomes	30019	242 (140)	http://crispr.u-psud.fr/crispr/
ORENZA	list of Enzyme Activities for which no sequences are available in the main sequence protein databases	7800	20 (4)	http://www.orenza.u-psud.fr/
Varna	Interactive drawing and editing of the RNA secondary structure	5810	205 (150)	http://varna.lri.fr/
Podospora	<i>Podospora anserina</i> Genome Project	4836	156 (58)	http://podospora.igmors.u-psud.fr/
Arnold	Easy identification of Rho-independent transcription terminators	3458	26 (22)	http://rna.igmors.u-psud.fr/toolbox/arnold/
FUNGIPath	Database to compare the primary and secondary metabolism of different fungal species	3000	17 (11)	http://embg.igmors.u-psud.fr/fungipath/
Bacterial Genotyping	resources for bacterial micro-evolution studies using MLVA or CRISPR typing	2919	60 (26)	http://mlva.u-psud.fr/mlvav4/genotyping/
CRISPRFinder	Easy detection of CRISPRs in user-submitted sequence data	2273	261 (161)	http://crispr.u-psud.fr/Server/CRISPRfinder.php/
Erpin	Easy RNA Profile Identification tool, an RNA motif search program	1068	176 (40)	http://rna.igmors.u-psud.fr/Software/erpin.php
NAPPdb	Prediction of functional non-coding RNA in Bacteria and Archaea	288	53 (21)	http://napp.u-psud.fr/
GenRGens	Software tool dedicated to randomly generating genomic sequences and structures	279	59 (19)	http://www.lri.fr/~genrgens/

Plus de 20 serveurs et logiciels en téléchargement.

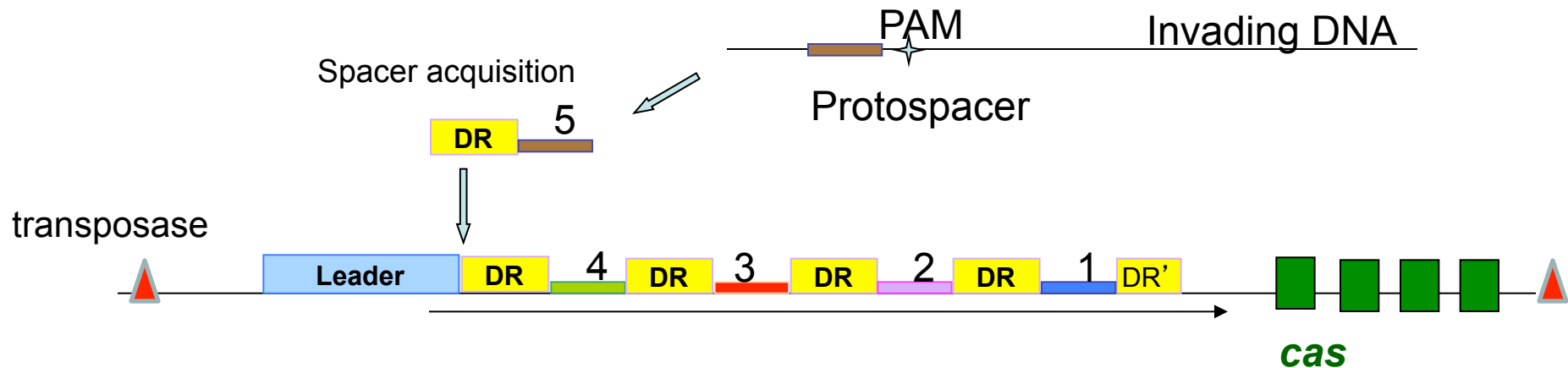
Forte orientation sur l'annotation des génomes microbien et ARN.

9 serveurs avec plus de 1000 visiteurs uniques/an

The CRISPR-Cas system: a prokaryotic immune system

- The CRISPR structure carries the memory of past encounters with foreign, non-self DNA: the spacers
- Upon re-invasion by phages or plasmid sequences, an interference system attacks and destroys the invader
- Autoimmunity is prevented by a specific recognition signal in the protospacer, the Protospacer Adjacent Motif (PAM)

Structure of a CRISPR–Cas locus



- Repeat size: 24bp to 50bp
- Repeat+ spacer size: 60-75bp
- Present in the majority of archaea and 50% of bacteria
- When several CRISPRs with the same DR are present, only one group of cas genes is associated to a CRISPR
- CRISPR-Cas diversity inside a species
 - Repeat sequence
 - Number and variety of spacers
 - Number and variety of CRISPR-Cas systems

Assessing the presence and polymorphism of CRISPR-Cas systems in sequenced genomes

- Dedicated programs to identify CRISPRs
 - CRT
 - Piler CR
 - CRISPRfinder
 - Pygram
- Databases
 - CRISPRdb
 - CRISPI

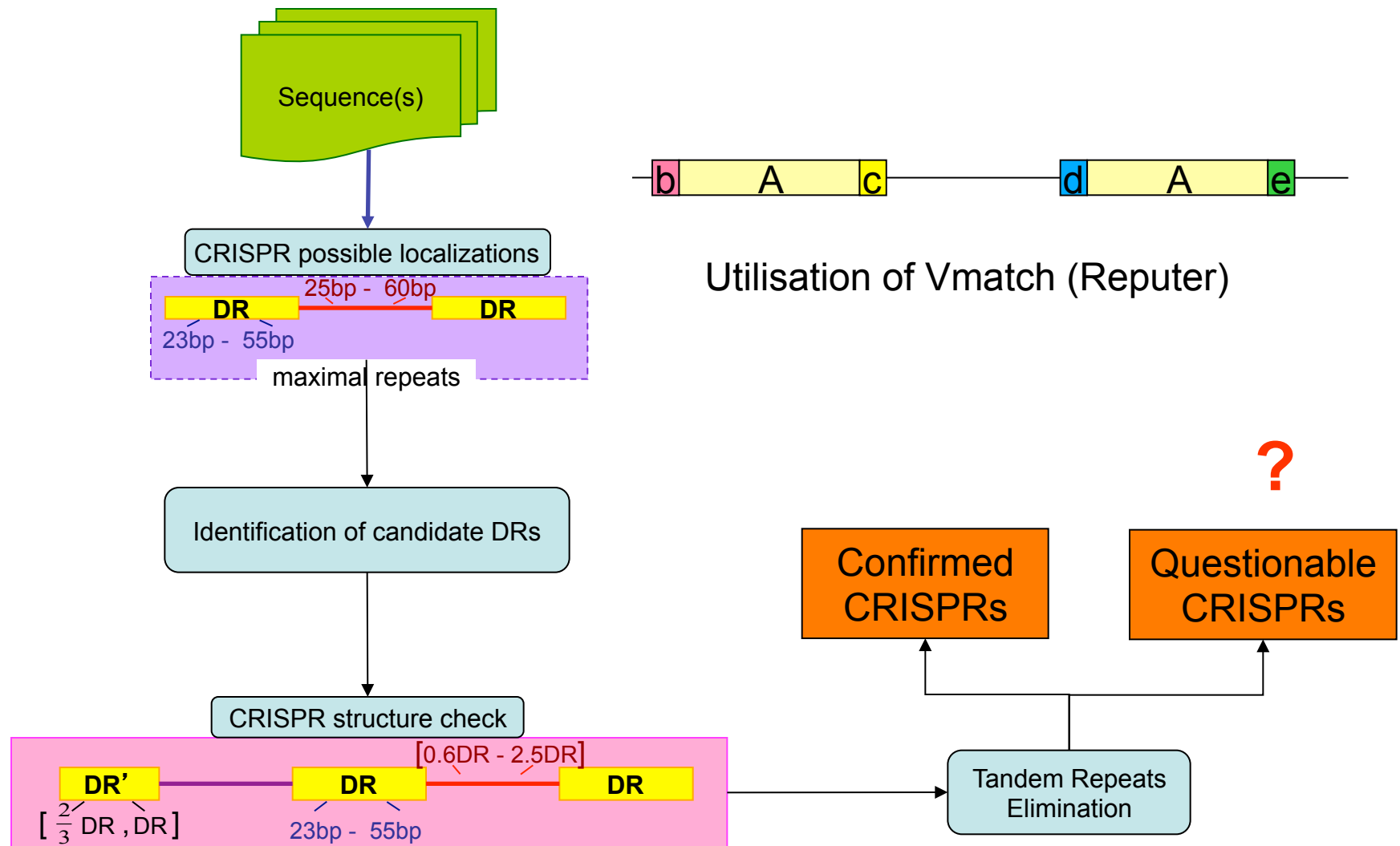
CRISPRs web-server

Université Paris-Sud

<http://crispr.u-psud.fr/>

- **CRISPRfinder**: a software and tools to identify CRISPR and analyse their components
- **CRISPRdb**: a database of CRISPRs from published genome sequences
- **CRISPRcompar**: tools to compare CRISPRs in different genomes and create a dictionary of spacers

CRISPRfinder Workflow



CRISPRfinder



Navigation

Home page
CRISPRs database
CRISPRs finder
CRISPRs comparison

Tools

MyCRISPRdb
BLAST CRISPRdb
CRISPRcompar
CRISPRtionary
FlankAlign
CRISPRs Utilities

GPMS Links

GPMS Team
Tandem Repeats DB
MLVA Web Service

External Links

Tandem Repeats Finder
GenomesOnline
Cas Genes at TIGR
Cas Genes for all genomes



CRISPRfinder program online

Wednesday January 23rd 2013

[Home](#) [About CRISPRs](#) [News](#) [FAQs](#) [Help](#) [Contact Us](#) [Examples](#) [IGM](#)

Welcome to the CRISPRs web service. This page enables the easy detection of [CRISPRs](#) in user-submitted sequence data (allows sequences up to 67,000,000 bp).

Input format :

Your data must be a DNA sequence (or many DNA sequences) in [FASTA](#) format.

Download [sample1 Aquifex VF5](#) or [sample2 P. aeruginosa PA14](#) or [samples YP1 Yersinia](#)



[Page manual](#)

Submit your Own data

Enter sequence (DNA) below using copy and paste:

Or upload a local file:

[Advanced version](#)

<http://crispr.u-psud.fr/Server/CRISPRfinder.php>

Navigation

[CRISPR Finder](#)[BLAST Spacers](#)[Flanking Sequences](#)[Identical DRs](#)[Save files](#)

Links

[Minisatellites Page](#)[GPMS Team \(french\)](#)[MLVA Web Service](#)

External Links

[GenomesOnline](#)[Tandem Repeats
Finder](#)[Cas Genes \(TIGR
Genome Properties\)](#)

CRISPRs found in the submitted genomic sequence

[FAQs](#) [Help](#) [About CRISPRFinder](#) [Contact Us](#) [NEWS](#) [Examples](#) [IGM](#)

Upload done with success

File name : **Y_pestis_C092.txt**
File size : 4786701

- Genome :
- Length (bp): 4653728
- ID : pestis

Number of CRISPRs candidates = 6

Confirmed CRISPRs = 3

Questionable CRISPRs = 3

- [tmp_1_Crispr_2](#)
- [tmp_1_Crispr_3](#)
- [tmp_1_Crispr_4](#)

- [tmp_1_PossibleCrispr_1](#)
- [tmp_1_PossibleCrispr_5](#)
- [tmp_1_PossibleCrispr_6](#)

[CRISPRs candidates localization](#)

Confirmed CRISPRs (3)

CRISPR id : tmp_1_Crispr_2

- START crispr position : 1773655 ----- END crispr position : 1773862 -----
Crispr length : 207
- DR consensus : GTTCACTGCCGCACAGGCAGCTTAGAAA DR length : 28
- Number of spacers : 3

1773655	GTTCACTGCCGCACAGGCAGCTTAGAAA	TTTTCGACGATTTATTTTATTTGTATTCAC	1773714
1773715	GTTCACTGCCGCACAGGCAGCTTAGAAA	ATTTATTAAAGATGCTGACAAAAAGAACTTAA	1773774
1773775	GTTCACTGCCGCACAGGCAGCTTAGAAA	GTCAACGCTTCACTCCCTGCCGCGGGGTATAAC	1773834
1773835	GTTCACTGCCGCACAGGCAGCCCAATAA		1773862

[Display spacers](#)[Crispr Properties](#)

CRISPR id : tmp_1_Crispr_3

- START crispr position : 2769294 ----- END crispr position : 2769802 -----
Crispr length : 508
- DR consensus : TTTCTAAGCTGCCTGTGCGGCAGTGAAC DR length : 28
- Number of spacers : 8

2769294	TTTGATTATTGCCTGTGCGGCAGTGAAC	TCAGGGGACTGGCGAACAATGTCTTTTCATGAT	2769353
2769354	TTTCTAAGCTGCCTGTGCGGCAGTGAAC	GAAAAAGGTAAAGATGGGCAAGCTTCTAGTAGTT	2769413
2769414	TTTCTAAGCTGCCTGTGCGGCAGTGAAC	TTTCTAAGCTGCCTGTGCGGCAGTGAAC	2769473

Confirmed CRISPRs

(4626121) *Shewanella* sp. ANA-3 (CRISPR_2)

1	AGGTTTTCGCTGCCTTTTCGGCGGGTATC	TCAAAGTCAACTTGTAATGACGATTTTCACG	32
	** ** *		
2	ATTTTCAGCTGCCTATTCGGCAGGTCAC	AGTTTGGGGCTGAGTTTGCCATTTTCCTAAAT	32
3	ATTTTCAGCTGCCTATTCGGCAGGTCAC	GATGAAGCAGACCACCTCGATTACCCACGCT	32
4	ATTTTCAGCTGCCTATTCGGCAGGTCAC	ACTATTTATCAAGACCTTCTTTAAATCAAAC	32
5	ATTTTCAGCTGCCTATTCGGCAGGTCAC	AGTTTGGGGCTGAGTTTGCCATTTTCCTAAAC	32
6	ATTTTCAGCTGCCTATTCGGCAGGTCAC		

(4626448)

Sulfolobus tokodaii str. 7 (CRISPR_2)

(32702)

7	GATGAATCCCAAAGGAATTGAAAG	TGATTGATCACAATGAGAAGACTGTAAAGCTGATAAAC	38
8	GATGAATCCCAAAGGAATTGAAAG	TGTTGAGGCATAAATTAATCTATCCTTAATGAAAAAT	37
9	GATGAATCCCAAAGGAATTGAAAG	TTCTTCCTCAGCCTCCATTTTGTGTTATGATTTGTAGTGCC	40
10	GATGAATCCCAAAGGAATTGAAAG	TTCAATAATCTCTATCTTTCCAAAATCTGTAAATGAAGAC	40
109	GATGAATCCCAAAGGAATTGAAAG	AAAGCACAGTCAATAACGTTATCTGGTATCATATTATCAAA	41
110	GATGAATCCCAAAGGAATTGAAAG	CTTTCTCCTTCCCTCTGATCTCTCGCTGAATTGAAAAGA	39
111	GATGAATCCCAAAGGAATTGAAAG	GTAAGTATTGATGCTAACATTGACTTCGCTGTCCAGGGGC	41
112	GATGAATCCCAAAGGAATTGAAGG	AAGTATAATAACGATAGTACTAAAATTAATTGATCC	36
	* *		
113	GATGATTCTCAAAGGAATTGATAA		

(39896)

23-55pb

0,6-2,5 X DR

False CRISPR looking true

CRISPR ID : tmp_1_Crispr_1

- Start position : 591225 ----- End position : 592255 ----- CRISPR length : 1030
- DR consensus : CTCACTGCCGGCTATGGCAGCAC DR length : 23
- Number of spacers : 21

591225	GAAATCGCCACCTATGGCAGCAC	GCTCACTGGCGCCGATCAGAGTCAG	591272
591273	CTGATCGCAGGCTATGGAAGTAC	CGAGACCGCGGGGGAACGGGCAGCGAC	591320
591321	TTGATCGCCGGCTACGGCAGCAC	CGGCGTGGCCGGATCCGACAGCACCC	591368
591369	ATCGTTGCCGGCTATGGCAGCTC	TCAAACGGCCGGAGGGCGGAAGCACG	591416
591417	CTGACTGCCGGGTATGGCAGCAC	ACAGACCGCCCGCAATGGCAGCGAG	591464
591465	CTGACGGCCGGTTACGGCAGTAC	GGAAACCGCCGGTGCAGACAGCTCG	591512
591513	CTGATCGCAGGCTACGGCAGCAC	GCAAACCTCCGGCGGCGACAGCTCG	591560
591561	TTGACTGCCGGCTATGGCAGTAC	CCAAACCGCGCAGAACGGCAGCGAC	591608
591609	CTCACC GCCGGCTACGGCAGCAC	CAGCACGGCGGGCACC GATAGCTCG	591656
591657	CTGATCGCCGGCTACGGCAGCAC	CCAGACCTCGGGCGGAGAGAGCTCG	591704
591705	CTCACTGCCGGCTATGGCAGCAC	GCAGACGGCGCAGGAAGGCAGCGAC	591752
591753	CTCACC GCCGGCTATGGCAGCAC	AGGCACCGCAGGTGCAGATAGCTCC	591800
591801	CTGATTGCTGGTTATGGCAGTAC	GCAGACATCCGGCAACGACAGCTCG	591848
591849	CTGACGGCGGGCTACGGCAGCAC	CCAAACCGCCCGTACCGGCAGCGAC	591896
591897	CTCACTGCCGGTTACGGCAGCAC	CTCCACCGCTGGTGCGGACAGCACG	591944
591945	CTGATCGCCGGTTATGGCAGCAC		591992
591993	CTCACC GCCGGCTACGGCAGCAC	CCAAACCGCGCGCAAGGGTAGCGAC	592040
592041	CTCACGACGGGGTACGGCAGCAC	CTCGACCGCGCGCGGACAGCACG	592088
592089	TTGATCGCCGGCTACGGCAGCAC	GCAGACCTCGGGTAGCGAGATTCC	592136
592137	CTCACGGCTGGCTATGGCAGTAC	GCAGACCGCACGCAAGGGCAGTGAC	592184
592185	CTCACC GCCGGCTACGGCAGCAC	CTCGACCGCCGGTGCGGACAGCAC	592232
592233	TTGATCGCCGGCTACGGCAGCAC		592255

Ice nucleated protein InaA
Xanthomonas campestris
Pantoea ananatis

SSLIAGYGSTQTS GNDSSLTAGYGSTQTA RTGSDLTAGYGSTSTA GADSTLIAGYGST

QTSGGDSSLTAGYGSTQTARKGSDLTGYGSTSTA GADSTLIAGYGSTQTS GSESSLT

True CRISPR looking false?

<input checked="" type="checkbox"/>	NC_014318_18	6436737	6437735	16	GTTGTCCCCGCACGCGCGGGGCTGGTCCC
<input type="checkbox"/>	NC_014318_19	6438560	6438832	4	GTTGTCCCCGCACGCGCGGGGCTGGTCCC
<input type="checkbox"/>	NC_014318_20	6441372	6441767	6	GTTGTCCCCGCACGCGCGGGGCTGGTCCC
<input type="checkbox"/>	NC_014318_21	6441907	6442178	4	GTTGTCCCCGCACGCGCGGGGCTGGTCCC

Strain : *Amycolatopsis mediterranei* U32

RefSeq : NC_014318 (chromosome circular)

CRISPR id : NC_014318_18

DR consensus (29 bp) : GTTGTCCCCGCACGCGCGGGGCTGGTCCC

Number of repetitions : 17

Begin Position : 6436737

End Position : 6437735

6436737	GGTTTCCCCGCGAACCCGGGGCTGGTCC	GTCGTGTAGCGGGTGTCTGTCTGTGGG
6436796	ACGTTTCCCCGCACGCGGGGCTGGTCCC	TGCCCCGTGGCGGAACCTGATCGTCACCAAGC
6436856	CGCTTCCCATCGCACGCGGGGCTGGTCCC	CGCCCTTCGTCCCGAGCTGGCGGGCTCGACGAT
6436917	ATTTTCCCCGCGCACGCGGGGCGGGTCCC	AGGGAGATCGACACGTGCGACAGCGTTTC
6436975	GCGGTTCGCCGCGCACGCGGGGCTGGTCCC	GGAAGCCGGTTGAGCTTCGCGCCGACCGTC
6437035	CGTTTGCCCCGCGCACGCGGGGCGGGTCCC	CAGGCCGACTACGAGAAGCAGAGCCCGCAG
6437094	GAGCTCCCCGTGCACGCGGGGCTGGTCCC	ATCCCGCAGTAGCGACAGGTGTGCGCGTCCG
6437155	GTTTTCGCCGCGCACGCGGGGCTGGTCCC	CACACGATCGGCGGTTTACCCCGACCTTCGCG
6437216	GTCGTCCCCGCGCACGCGGGGCTGATCCC	GCATTGGCTAGATTGTTCTGGAAGTCTGATC
6437277	GTTGTCCCCGCGCACGCTGGGCTGATCCC	GTCTCGAACGTCGTGGGCGAGACGACGTGCTC
6437338	GTTGTCCCCGCGCACGCGGGGCTGGTCCC	GACTTCAAGCTGCGCCTGGTCACCCCTCGGCGT
6437399	GTCGTCCCCGCGAACCGGGGCTGGTCCC	CCGTTTCGGGTCGCACTCGTACGTGGGCAGATC
6437460	GTCGTCCCCGCGCACGCGGGGCTGGGTC	CCGTACCTGCCGCGGGGCTACACCAGCCTGTT
6437523	GTCGTCCCCGCGCACACGAGGCTGGTCCC	GACGGCGTCAAGAAGTGTACGACGAAGTGGCA
6437585	GTCGTCCCTGCACACGCGGGGCTGGTCCC	TACGTGTACGTTGGCGCTACGGAACAGGAAC
6437646	GTCGTCCCCGCGCACGCGGGGCTGGTCCC	GCTCGGGCAGCTCAACCTCGCCACTGTCTAC
6437706	CTTGTTCACACGCGCGGGGCTGGTCCC	

CRISPR's associated proteins in reference sequence

[Home](#) [About CRISPR](#) [NEWS](#) [FAQs](#) [Help](#) [Contact Us](#) [Examples](#) [IGM](#)

Product referenced in Genbank file	Start	End	Strand
CRISPR-associated CSE3 family protein	6444925	6445575	Rev
CRISPR-associated protein CSE4	6446273	6447403	Rev
CRISPR-associated protein CSE2	6447400	6447996	Rev
CRISPR-associated protein CSE1	6447993	6449573	Rev
CRISPR-associated helicase Cas3	6449887	6452166	Rev

Type I-E system
cas1 and cas2 missing

The CRISPRs database (CRISPRdb): CRISPR, DR, spacer repertoire from published genomes

	Genomes analysed	CRISPRs found	Strains with convincing CRISPR(s)	Strains devoid of detectable CRISPRs
Archea	150	563	126	24
Bacteria	2612	3502	1176	873
Total	2762	4065	1302	897

Click on strain name to get more information.

Bacteria

- 'Nostoc azollae' 0708 (1 CRISPR, 3 questionable structures)
- Acaryochloris marina MBIC11017 (1 CRISPR, 2 questionable structures)
- Acetobacter pasteurianus 386B
- Acetobacter pasteurianus IFO 3283-01 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-01-42C (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-03 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-07 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-12 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-22 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-26 (1 CRISPR)
- Acetobacter pasteurianus IFO 3283-32 (1 CRISPR)
- Acetobacterium woodii DSM 1030 (1 CRISPR)
- Acetohalobium arabaticum DSM 5501 (1 CRISPR)
- Acholeplasma brassicae (1 questionable structure)
- Acholeplasma laidlawii PG-8A (2 questionable structures)
- Acholeplasma palmae J233 (1 CRISPR)
- Achromobacter xylosoxidans A8 (1 questionable structure)
- Achromobacter xylosoxidans NBRC 15126 = ATCC 27061

Identification of CRISPRs using the CRISPRfinder program

Selection	CRISPR_id	Start Position	End Position	Number of spacers	DR consensus
<input type="checkbox"/>	NC_013235_2	1333814	1336110	31	GTTGCATTCCACGTCACGCCGTTGGGCGTATAGAAAC
<input type="checkbox"/>	NC_013235_3	1364626	1365686	14	GTCAGAAACCACGTCACGCCGTCAGGCGTATAGAAAC
<input checked="" type="checkbox"/>	NC_013235_4	1367179	1367813	8	GTCAGAAACCACGTCACGCCGTCAGGCGTATAGAAAC
<input type="checkbox"/>	NC_013235_5	1406757	1406860	1	TTTCCTGGGCCTGGCCGCGGCTTTGATCAG
<input type="checkbox"/>	NC_013235_8	3336213	3338504	37	GGGCTCATCCCCGCGGGCGCGGGGAGCAC
<input type="checkbox"/>	NC_013235_9	3344692	3345758	17	GGGCTCATCCCCGCGGGCGCGGGGAGCAC
<input type="checkbox"/>	NC_013235_10	3348443	3355073	108	GGGCTCATCCCCGCGGGCGCGGGGAGCAC
<input type="checkbox"/>	NC_013235_11	3356560	3358846	37	GGGCTCATCCCCGCGGGCGCGGGGAGCAC

Find referenced Cas genes in the whole sequence

CRISPR's associated proteins in Taxon 479431

[Home](#) [About CRISPR](#) [NEWS](#) [FAQs](#) [Help](#) [Contact Us](#) [Examples](#) [IGM](#)

NC_013235

Product referenced in Genbank file	Start	End	Strand
CRISPR-associated protein Cas2	1364134	1364418	Fwd
CRISPR-associated protein Cas2	3358886	3359260	Rev
CRISPR-associated protein Cas1	3359248	3360204	Rev
CRISPR-associated protein, Cse3 family	3360201	3360848	Rev
CRISPR-associated protein Cas5 family	3360850	3361608	Rev
CRISPR-associated protein, Cse4 family	3361605	3362741	Rev
CRISPR-associated protein, Cse2 family	3362738	3363385	Rev
CRISPR-associated protein, Cse1 family	3363382	3365049	Rev
CRISPR-associated helicase Cas3	3365254	3368010	Rev

Searching Genbank
annotations for *cas* genes

Searching for *Cas* genes by BLAST

Selection	CRISPR_id	Start Position	End Position	Number of spacers	
<input checked="" type="checkbox"/>	NC_008463_2	2926520	2927393	14	
<input checked="" type="checkbox"/>	NC_008463_3	2935917	2937205	21	

No cas gene referenced in Genbank sequence

BLAST the sequence against the local cas bank

Searching similarities with known *cas* genes in all the sequences of the genome

using BLASTX and a local *cas* database

208963, NC_008463

Gene id	Definition	Gene family	Organisms	Start	End	Length	Strand	value	Identity
Q02ML8_PSEAB	Possible helicase	no info	<i>Pseudomonas aeruginosa</i>	2934685	2931458	3226	-	0.0	100.00
B9ZK83_9GAMM	CRISPR-associated helicase Cas3 family	no info	<i>Thioalkalivibrio</i> sp. K90mix	2934685	2931470	3214	-	0.0	52.75
C4LEF7_TOLAT	CRISPR-associated helicase Cas3 family	no info	Proteobacteria	2934685	2931473	3211	-	0.0	46.93
B2K7A0_YERPB	CRISPR-associated helicase Cas3 family	no info	Proteobacteria	2934685	2931473	3211	-	0.0	46.92
A1W3B4_ACISJ	CRISPR-associated helicase Cas3 family	no info	Proteobacteria	2934685	2931464	3220	-	0.0	47.47
A4W8Q3_ENT38	CRISPR-associated helicase Cas3 family	no info	Enterobacteriaceae	2934685	2931473	3211	-	0.0	45.51
D0FPQ9_ERWPY	CRISPR-associated helicase	Cas3_family	Enterobacteriaceae	2934685	2931464	3220	-	0.0	46.19
Q02ML9_PSEAB	Putative uncharacterized protein	no info	<i>Pseudomonas aeruginosa</i>	2931397	2930096	1300	-	0.0	100.00
C8WBP8_ZYMMN	CRISPR-associated helicase Cas3 family	no info	Proteobacteria	2934685	2931458	3226	-	0.0	44.08
A6WRC8_SHEB8	CRISPR-associated helicase Cas3 family	no info	Gammaproteobacteria	2934685	2931473	3211	-	0.0	41.92
Q02MM0_PSEAB	Putative uncharacterized protein	no info	<i>Pseudomonas aeruginosa</i>	2930106	2929126	979	-	0.0	100.00
C9LM24_9FIRM	CRISPR-associated helicase Cas3	no info	<i>Dialister invisus</i> DSM 15470	2934685	2931464	3220	-	3e-173	36.44
B7GYY5_ACIB3	CRISPR-associated helicase Cas3	no info	Gammaproteobacteria	2934685	2931509	3175	-	3e-164	35.03
A9BUF1_DELAS	CRISPR-associated protein Cas1	no info	Proteobacteria	2935656	2934706	949	-	5e-130	72.56
A9BUF6_DELAS	CRISPR-associated protein	Csy3_family	Bacteria	2929097	2928096	1000	-	9e-128	69.32

2926520	GTTCACTGCCGTATAGGCAGCTAAGAAA	ACCACCCGCTACCACCGGCAGCCGACCCGGC
2926580	GTTCACTGCCGTATAGGCAGCTAAGAAA	ATCACCGCGCGCTGGCCTTCGGCGCCAGGC
2926640	GTTCACTGCCGTATAGGCAGCTAAGAAA	TTATCAGTAAATACGTGGCGACTTGGCCACCC
2926700	GTTCACTGCCGTATAGGCAGCTAAGAAA	TGCTGCGTGTGTCAGCGACACCCGTTCAACA
2926760	GTTCACTGCCGTATAGGCAGCTAAGAAA	ACTGGAGATACGTTGCCACAGATCGGACGCACGC
2926820	GTTCACTGCCGTATAGGCAGCTAAGAAA	ATATCAGTTTGCATGGTTTGCTCCTACCAAGC
2926880	GTTCACTGCCGTATAGGCAGCTAAGAAA	TTGCATCGCACCTCTCTGCTACCGACCCAGGC
2926940	GTTCACTGCCGTATAGGCAGCTAAGAAA	TGAACGTTCCAGCATCAGCGCCCTTCCGGCC
2927000	GTTCACTGCCGTATAGGCAGCTAAGAAA	AGGCAGACGAGTACGCGCGCTCTCGCAAGTG
2927060	GTTCACTGCCGTATAGGCAGCTAAGAAA	AAGAACATGGCCCGGCTCCCGACGTTGCTGCGG
2927120	GTTCACTGCCGTATAGGCAGCTAAGAAA	ATGACGCAGACGAGCTTTGTCGCCCTTCTGGC
2927180	GTTCACTGCCGTATAGGCAGCTAAGAAA	GTTTGCAGGAAGCGCGCTATCGCACCGGAAC
2927240	GTTCACTGCCGTATAGGCAGCTAAGAAA	ATAGCTACGCCGAGCCAGTTGTAAAGCTGACGC
2927300	GTTCACTGCCGTATAGGCAGCTAAGAAA	AAAAAATGTCCCGAAGTTCATAAGCGGGCTTCGGGCGA
2927366	TTTCACTGCCACATAGGTCGTCAAGAAA	

Cas genes are searched from 10000 bp upstream to 10000 bp downstream the selected CRISPR

CRISPR's coordinates	2926520	2927393
Search region	2916520	2937393

No cas gene referenced in Genbank sequence

BLAST against the local cas bank (ouput with alignments)

NC_008463 [2916520..2937393]

Searching the local cas database ...

BLASTX 2.2.18 [Mar-02-2008]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, 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= (20,874 letters)

```
Database: casdb
      2586 sequences; 1,021,513 total letters
```

```
Searching.....done
```

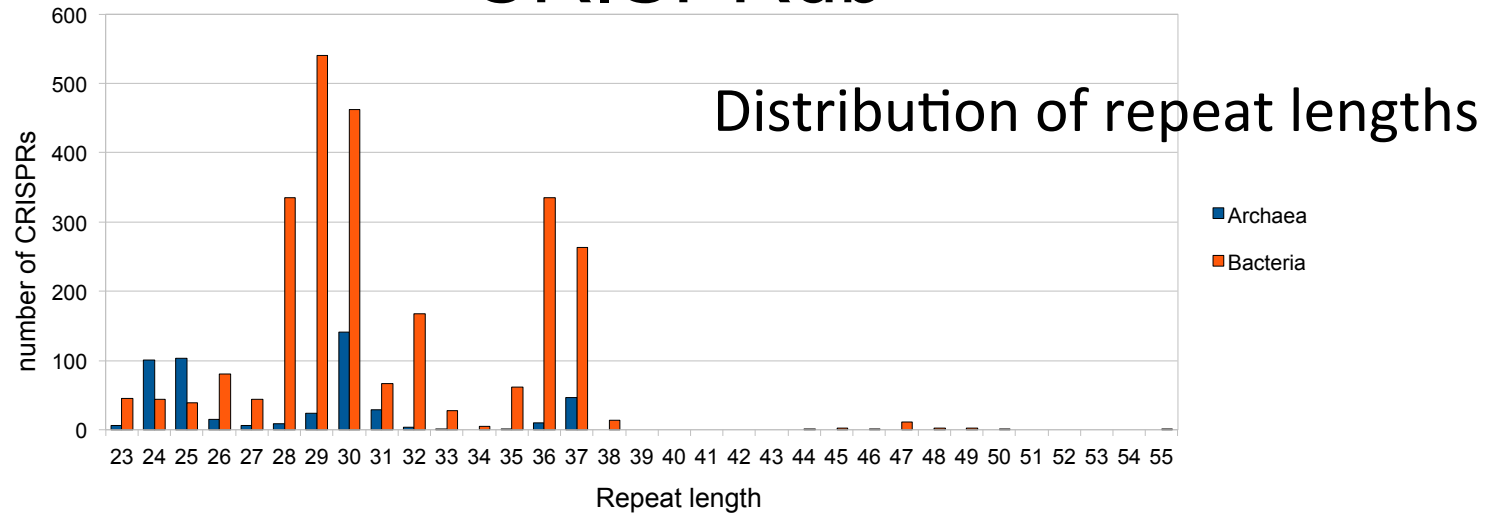
Sequences producing significant alignments:	Score	E
	(bits)	Value

Q02ML8_PSEAB	Possible helicase	tax=Pseudomonas aeruginosa	2181	0.0
B9ZK83_9GAMM	CRISPR-associated helicase	Cas3 family tax=Thioalka...	1086	0.0
C4LEF7_TOLAT	CRISPR-associated helicase	Cas3 family tax=Proteoba...	932	0.0
B2K7A0_YERPB	CRISPR-associated helicase	Cas3 family tax=Proteoba...	928	0.0
A1W3B4_AJCSJ	CRISPR-associated helicase	Cas3 family tax=Proteoba...	919	0.0
A4WQ38_ENT38	CRISPR-associated helicase	Cas3 family tax=Enteroba...	911	0.0
D7FPQ9_ERWPY	:Cas3 family CRISPR-associated helicase	tax=Enterob...	897	0.0
Q02ML8_PSEAB	Putative uncharacterized protein	tax=Pseudomonas ae...	883	0.0
C8WBP8_ZYMMN	CRISPR-associated helicase	Cas3 family tax=Proteoba...	849	0.0
A6WRC8_SHEB8	CRISPR-associated helicase	Cas3 family tax=Gammapro...	788	0.0
A1RK23_SHE5W	:Cas3 family CRISPR-associated helicase	tax=Gammapr...	671	0.0
Q02MM0_PSEAB	Putative uncharacterized protein	tax=Pseudomonas ae...	664	0.0
C9LW24_PFTRM	CRISPR-associated helicase	Cas3 tax=Dialister invis...	613	e-175
R7GMV5_ACTR3	CRISPR-associated helicase	Cas3 tax=Gammaproteobact...	583	e-167

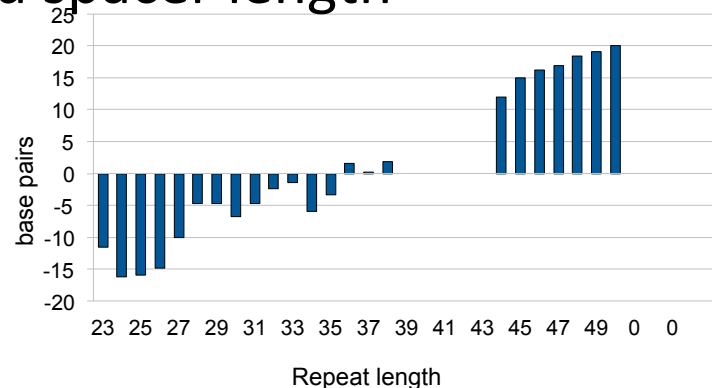
Accuracy of the database

- The parameters used to define CRISPRs are essential
- An examination of each CRISPR is necessary to eliminate false structures
- Short CRISPRs (1 or 2 spacers) cannot be easily confirmed

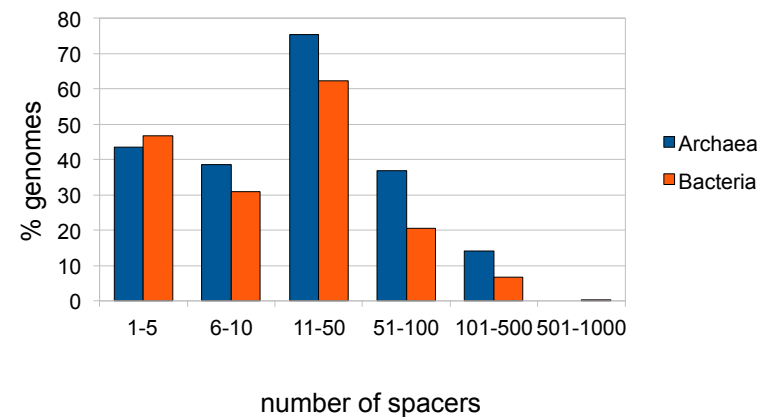
Characteristics of the DRs and spacers in CRISPRdb



Difference between repeat and spacer length



Number of spacers

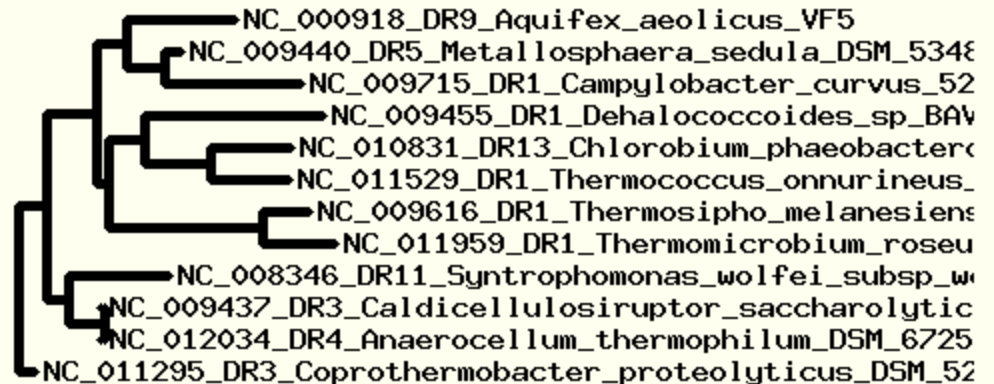


Classification of DRs


CLUSTAL 2.0.10 multiple sequence alignment

NC_010831_DR13_Chlorobium_phae	GTTTCAATTCTC	CATGAGTGCGATTAGAT--
NC_011529_DR1_Thermococcus_onn	GTTTCAATTCTC	CTAGAGTCTTATTGCAAC-
NC_009455_DR1_Dehalococcoides_	GTTTCAATTCAC	GCTCCCGCACGGGGAGCGAC
NC_009440_DR5_Metallosphaera_s	GTTTCAAACCCT	CGAAGGATCTCTACAAAC-
NC_009715_DR1_Campylobacter_cu	GTTTCAAATTCC	AACGGAATGATCTTTAAT-
NC_000918_DR9_Aquifex_aeolicus	GTTTCAACTCCA	CTACGGTACATTAGGAAC-
NC_011653_DR10_Thermosipho_afr	GTTTCAATCCCT	AATAGGTATGCTAAAAAC-
NC_011830_DR4_Desulfitobacteri	GTTTCAATCCCT	TATAGGTAAGCTAACAAC-
NC_012438_DR5_Sulfurihydrogeni	GTTTCAATTCCT	CATAGGCACTCTAAAAAC-
NC_011295_DR3_Coprothermobacte	GTTTCAATCCCT	TGTAGGTAAGCTAGAAAC-
NC_009437_DR3_Caldicellulosiru	GTTTCAATCCCC	AAAGGGAAGGCTACAAAC-
NC_012034_DR4_Anaerocellum_the	GTTTCAATCCCC	AAAGGGAAGGCTACAAAC-
NC_008346_DR11_Syntrophomonas_	GTTTCAATCCCC	TATACGCGGGTCACAAAC-
NC_009616_DR1_Thermosipho_mela	GTTTCTACCTTA	CCTTGGAGGAATTGAAAC-
NC_011959_DR1_Thermomicrobium_	GTTTCGACAGTA	CCTATGAGGGCTTGAAAC-
	***** *	

Phylogenetic relationships



The private database

**UNIVERSITÉ PARIS-SUD 11**

Navigation

- Home page
- CRISPRs database
- CRISPRs finder
- CRISPRs comparison

Tools


- MyCRISPRdb
- BLAST CRISPRdb
- CRISPRcompar
- CRISPRtionary
- FlankAlign
- CRISPRs Utilities

GPMS Links

- GPMS Team
- Tandem Repeats DB
- MLVA Web Service

External Links

- Tandem Repeats Finder
- GenomesOnline
- Cas Genes at TIGR
- Cas Genes for all genomes
- CRISPR plasmids for academic lab research at Addgene

**CRISPRfinder program online**

Thursday September 18th 2014

Home About CRISPRs News FAQs Help Contact Us Examples IGM

Private connection (christine.pourcel@u-psud.fr private data)
Disconnect from myCRISPRdb

- Consult your Private database
- make a CRISPRs comparison
- submit new data:

Welcome to the CRISPRs web service. This page enables the easy detection of CRISPRs in user-submitted sequence data (allows sequences up to 67,000,000 bp).

Input format :
Your data must be a DNA sequence (or many DNA sequences) in FASTA format.
Download [sample1 Aquifex VF5](#) or [sample2 P. aeruginosa PA14](#) or [samples YP1 Yersinia](#)

Submit your Own data
Enter sequence (DNA) below using copy and paste:

Or upload a local file: Aucun fichier sélectionné.

[Advanced version](#)

Related Works

- CRISPR evolution (Yersinia pestis)


Data Summary

Genomes analysed	CRISPRs found (*)
Archaea	503(126)
Bacteria	2812 3502(1178)
Total	2792 4095(1302)


* number of convincing CRISPR structures (number of genomes with such CRISPR)

Database status:
Last update : 2014-08-05

Contact:
Christine POURCEL

**CRISPRdb private database**

Thursday September 18th 2014

 [Page manual](#)

- View all the processed sequences
- View sequences by group
- View sequences harboring convincing CRISPRs
- myCRISPRdb home page
- Disconnect from myCRISPRdb

Click on an entry to view the CRISPRs found in the corresponding sequence

List of the processed sequences

A. baumannii_ATCC19606 6_2785	2 CRISPRs		2013-10-15 14:59:53	delete
Aquifex aeolicus VF5	5 CRISPRs	5 questionable CRISPRs	2013-10-23 12:03:11	delete
Bifidobacterium dentium 6_742	4 CRISPRs	11 questionable CRISPRs	2013-10-15 14:59:53	delete
Cjejunif458 6_12079	1 CRISPR		2013-10-15 14:59:53	delete
COH1 6_1059	1 CRISPR		2013-10-15 14:59:53	delete
E.coliB7A 6_309	2 CRISPRs		2013-10-15 14:59:53	delete
LESB58 6_753	1 CRISPR	1 questionable CRISPR	2013-10-15 14:59:53	delete
Mtub119 6_10327	1 CRISPR		2013-10-15 14:59:53	delete
pa14_ibtissem 6_266	2 CRISPRs	1 questionable CRISPR	2013-10-15 14:59:53	delete
Paeru2192 6_310	5 CRISPRs	1 questionable CRISPR	2013-10-15 14:59:53	delete
PaeruC3719 6_311	3 CRISPRs	3 questionable CRISPRs	2013-10-15 14:59:53	delete
PcyI_10_locus1	3 CRISPRs		2013-11-29 18:11:44	delete
PcyII_10_locus2	2 CRISPRs	2 questionable CRISPRs	2013-11-29 18:13:17	delete
S. agalactiae 515 6_1060	1 CRISPR		2013-10-15 14:59:53	delete

CRISPRtionary and CRISPRcompar



Navigation

Home Page
CRISPRs database
CRISPRs finder
CRISPRs comparison

Tools

My CRISPRdb
BLAST CRISPRdb
CRISPRcompar
CRISPRtionary
FlankAlign
CRISPRs Utilities

CRISPRs comparison

[Home](#) [About CRISPRs](#) [News](#) [FAQs](#) [Help](#) [Contact Us](#) [Examples](#) [IGM](#)

	RefSeq	Sequence
<input checked="" type="checkbox"/>	NC_021812	Salmonella enterica subsp. enterica serovar Heidelberg str. CFSAN002069 ,chromosome circular
<input checked="" type="checkbox"/>	NC_020307	Salmonella enterica subsp. enterica serovar Javiana str. CFSAN001992 ,chromosome circular
<input checked="" type="checkbox"/>	NC_021902	Salmonella enterica subsp. enterica serovar Newport str. USMARC-S3124.1 ,chromosome circular

Select additional genomes

Public data

'Nostoc azollae' 0708 (NC_014248: chromosome circular)
Acaryochloris marina MBIC11017 (NC_009925: chromosome circular)
Acetobacter pasteurianus IFO 3283-01 (NC_013210: plasmid pAPA01-0)
Acetobacter pasteurianus IFO 3283-01-42C (NC_017104: plasmid pAPA)
Acetobacter pasteurianus IFO 3283-03 (NC_017101: plasmid pAPA03-0)
Acetobacter pasteurianus IFO 3283-07 (NC_017143: plasmid pAPA07-0)
Acetobacter pasteurianus IFO 3283-12 (NC_017136: plasmid pAPA12-0)
Acetobacter pasteurianus IFO 3283-22 (NC_017117: plasmid pAPA22-0)

Private data

ESCH017A0_00910_009
LESB58[6_753(6_753)
Mtub119[6_10327(6_10327)
pa14_ibtissem[6_266(6_266)
Paeru219[6_310(6_310)
Paeru237[6_311(6_311)

Select a reference strain for the CRISPRs is:

CRISPR_1

- Left Flanking
- Right Flanking

CRISPR_2

- Left Flanking
- Right Flanking

CRISPR_3

CRISPR_4

CRISPR	Salmonella enterica subsp. enterica serovar Heidelberg str. CF SAN002069 (NC_021812)				Salmonella enterica subsp. enterica serovar Javiana str. CFSAN001992 (NC_020307)				Salmonella enterica subsp. enterica serovar Newport str. USMARC-S3124.1 (NC_021902)			
Label	CRISPR_id	Position	Nbr Spacers	Strand	CRISPR_id	Position	Nbr Spacers	Strand	CRISPR_id	Position	Nbr Spacers	Strand
CRISPR_1	NC_021812_2	[3651431,3652567]	18	+	NC_020307_1	[3904718,3905662]	15	+	NC_021902_2	[4098753,4099939]	19	-
CRISPR_2	NC_021812_3	[3668815,3669942]	18	+	--	--	--	--	NC_021902_1	[4080470,4082084]	26	-
CRISPR_3	NC_021812_4	[3670016,3670471]	7	+	--	--	--	--	--	--	--	--
CRISPR_4	--	--	--	--	NC_020307_2	[3915194,3915649]	7	+	--	--	--	--

dictionary1.xls [Mode de compatibilité]					A	B	C	D	E
1	key	AnnotatedSpacer	sequence						
2	1	NC_020307.1	TGCTATCCCTGCAGCTCCAGCGCCCTCTCTT						
3	2	NC_020307.2	GTGTGATCTCAITGGCGAGGGGTGCGTAAAA						
4	3	NC_020307.3	AAGAGCGTCATATGGCTCGAAGGCAAGACCA						
5	4	NC_020307.4	TTGTTAATATATACATACAAATCAATAC						
6	5	NC_020307.5	AGAGCGGCACCTTACCTTAGTCCGCGAAGGCC						
7	6	NC_020307.6	ACAAAAATCGCGCAGCGGCACTTCTACTGTGCA						
8	7	NC_020307.7	CGGTGCCCATATTGGCTATATGCTGACGCA						
9	8	NC_020307.8	CCAGTACGGTAGCGCGGGTGAGCGCGCGGCC						
10	9	NC_020307.9	TATTTCTCGTGTGGCGGCTCATTGGCGCTA						
11	10	NC_020307.10	CTCATAGCTCGGATACCGCGATATACCGGATC						
12	11	NC_020307.11	GGAGGACGCTACTATCGGATTCGCTCGGGGA						
13	12	NC_020307.12	GGGGCAATATCATCACTACGACCGGGA						
14	13	NC_020307.13	GGGTCAATATGGGGTTATGGGAGGGAGCGG						
15	14	NC_020307.14	CCCTACAGTTGCCACCGCGTACAACTCAACA						
16	15	NC_020307.15	CCATACCATTTCCACTGGCGCAACGATTTTG						
17	16	NC_021812.1	TTCCAGAACCGTTTGAATCTACTGTGGCAATTA						
18	17	NC_021812.2	GCAGCGGTTGAGTAACCTCTGTCGACGTCGA						
19	18	NC_021812.3	CTCCAGCGCTCGAATTTATTGAGGCCACCA						
20	19	NC_021812.4	TTTTGATAGTAGTATTCATTAAGCCTCTAG						
21	20	NC_021812.5	GCAGGTCATAAAAAATGGTGTGGCTTACC						
22	21	NC_021812.6	CCGGCATCAGCGCCGATCCGTTATATGCCCC						
23	22	NC_021812.7	AAAAAACAGAGAACGGCAAGCGGCACTCAA						
24	23	NC_021812.8	CGTCAGCGCGGTATTGAGCGCGGGACCGCC						
25	24	NC_021812.9	AACAGGAACAGGAAAAAAGATTGTTCGGGT						
26	25	NC_021812.10	CAGATCCTCAACGCTCAGGCTGTTAGTTTCCT						
27	26	NC_021812.11	CGAGGATGGAAATATTTCCAGGCTGGGAT						
28	27	NC_021812.12	ATGCGGAAACGCTGATGGGCTTGACATGAGC						
29	28	NC_021812.13	AATTATTTCTGTGGCTGGGGTTTCGATTGAT						
30	29	NC_021812.14	TGACGCTGGTCTATACCGGCAACGACGCGAC						
31	30	NC_021812.15	TTGATCGAGAGTGCAGAGGCGAGAACGGGCA						
32	31	NC_021812.16	CAGGTTATGCGCAAAAAATTAATTCATATTATA						
33	32	NC_021812.17	GACGAGTTCTGGAAATGGTAGCTGATAAAGA						
34	33	NC_021812.18	CGTTATCGGACGCTGACGCAATATGAGAT						
35	34	NC_021902.1	TGCTGCGGCGAAAAAAGCTCGCTAAATGGGAA						

The CRISPR webservice perspectives

- Improvements
 - database update
 - automatic quality tests
 - CRISPRfinder download
 - Access to data in the database
 - Increase the capacity of MyCRISPRdb
- New developments
 - Analysis of metagenomic data
 - Identification of protospacers

Past and future of the CRISPRwebsite

- Created by a PhD student, Ibtissem Grissa (2006-2008) on a GPMS' team server
- Transferred to a server at eBio by Christine Drevet in 2010-2011
- Regularly updated, maintained and improved until now by C. Drevet
- GPMS team cannot further maintain/develop the site due to lack of specific funding
- Help needed to keep the resource working

- Update
 - <ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>

A single family of DR >40bp, all associated to type II system (cas9/csn1)

PD	1	GTTGTGAATTGCTTCAAAATGTGTATCTTTGCAGTAGCAAGCACAAAT
WV	1	GTTGTGAATTCTTTCAAATTTTGTAGATTTACTCAGTAGAATTAC
FS	1	GTTGTGAATTGCTTTCAGATTTTCTAGTTTTACTAGTGCAATAACAA
LB	1	GTTGTGAATTCTTTCAAATTTTGTAGTAACTTTATGATTACTGTGCA
ZP	1	GTTGTGAATTGCTTTCAGAATTGTACTTTTAGTACTGGATTACACAGG
FP	1	GTTGTGAATTGCTTTCAAAATTGTATTTTAGCTTATAATTACCAAC
BF	1	GTTGTGATTTTGTCTTTCAAATTAGTATCTTTGAAACCATTGGAAACAGC
CO	1	GTTGTGAATTGCTTTCAAATTTTGTAGTTTTTGCGATTGATAACA
RA	1	GTTGTGAATTGCTTTCAAATTTTACTATCTTTGTGATAGTTCGCAAC
FT	1	GTTGTGATTCGCTTTCAAATTTGTATTTTTAGAACGATAGCACACAAC

The majority are bacteria of the Bacteroidetes phylum

Alignment with shorter DRs associated to type II system

WV	1	GTTGTGAATTCCTTTCAAATTTTGTAGATTTACTCAGTAGAATTACAAC
LB	1	GTTGTGAATTCCTTTCAAATTTTGTAGTAACTTATGATTACTGTCGCAAC
BF	1	GTTGTGATTTGCTTTCAAATTAGTATCTTTGAACCATTGGAAACAGC
FS	1	GTTGTGAATTGCTTTCAGATTTTCTAGTTTACTAGTGCAATAACAAC
ZP	1	GTTGTGAATTGCTTTCAGAATTGTACTTTTAGTACTGGATTACACAGG
FP	1	GTTGTGAATTGCTTTCAAAATTGTATTTTAGCTTATAATTACCAAC
CO	1	GTTGTGAATTGCTTTCAAATTTTGTAGTTTTCGATTGATAACA
RA	1	GTTGTGAATTGCTTTCAAATTTTACTATCTTTGTGATAGTTCGCAAC
RA ₂	1	GTTGTGAATACCTTTCAAATGAGAGCAGTCCCAAC
PM	1	GTTGTAGTTGCTCTCTCATTTCGCAGTGCTACAAT
NM	1	GTTGTAGCTCCCTTTCTCATTTCGCAGTGCTACAAT
IP	1	GTTGTACTTCCTAATTATTTTAGCTATGTTACAAT
DD	1	GTTGTAGCTCCCTCTCTCACCCCGGATAGCTACACT
AZO	1	GTTGGGGCTGGACCCCGATCCCCATCGGCTACACT
LI	1	GTTTTGTTAGCATTCAAAATAACATAGCTCTAAAAAC
SP	1	GTTTTGGGACATTCAAAACAGCATAGCTCTAAAAAC
ST	1	GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC
PR	1	GTTGTATATCATTCCTTTCCATACATCAAAACCACAAC

S. pyogenes

