W170–W175 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 16 May 2008*
doi:10.1093/nar/gkn294

QUMA: quantification tool for methylation analysis

Yuichi Kumaki, Masaaki Oda and Masaki Okano*

Laboratory for Mammalian Epigenetic Studies, Center for Developmental Biology, RIKEN,
2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan

Received January 28, 2008; Revised April 18, 2008; Accepted April 29, 2008

ABSTRACT

Bisulfite sequencing, a standard method for DNA methylation profile analysis, is widely used in basic and clinical studies. This method is limited, however, by the time-consuming data analysis processes required to obtain accurate DNA methylation profiles from the raw sequence output of the DNA sequencer, and by the fact that quality checking of the results can be influenced by a researcher's bias. We have developed an interactive and easy-to-use web-based tool, QUMA (quantification tool for methylation analysis), for the bisulfite sequencing analysis of CpG methylation. QUMA includes most of the data-processing functions necessary for the analysis of bisulfite sequences. It also provides a platform for consistent quality control of the analysis. The QUMA web server is available at http://quma.cdb.riken.jp/.

INTRODUCTION

DNA methylation at the C-5 position of cytosine is a major epigenetic silencing mechanism in many eukaryotic organisms. The sequence context of methylated cytosines in the genome depends on the organism. In mammals, the genomes are methylated almost exclusively at CpG dinucleotides, while those of flowering plants are methylated at both CpG and non-CpG sequences. In mammals, CpG methylation plays a pivotal role in genomic imprinting, development, cell growth and survival and gene regulation (1,2). Abnormal DNA methylation profiles are associated with a broad range of human diseases, especially cancers (3,4). Bisulfite sequencing analysis has been the standard method used by biological and medical researchers to detect cytosine methylation profiles in genomic DNA at the single-nucleotide level (5). This method, however, is limited by the laborious and time-consuming data analysis steps required: making the alignment, trimming raw sequences, extracting DNA methylation profiles, excluding low-quality sequences, performing statistical analysis and drawing figures that summarize the methylation patterns. Furthermore, due to the error-prone nature of the bisulfite chemical reaction and subsequent PCR amplification (6–8), a quality check of the aligned sequences is essential. Thus, a user-friendly tool for both data analysis and a quality check of the bisulfite sequencing is desirable.

Several software tools are currently available for bisulfite sequencing analysis (9–14), and they have both advantages and disadvantages. MethTools is a pioneering tool for bisulfite sequencing analysis (9), and comprehensive analysis results are quickly obtained from input data. However, it requires aligned sequences as input data, so that trimming and multiple-alignment of raw sequences are necessary prior to data submission. Similarly, CyMATE (13), a unique mapping tool for both CpG and non-CpG methylation, and CpG PatternFinder (14), a Windows-based program, also require aligned sequences as input data. MethTools and CyMATE receive input data through a web submission form and return their results by Email. MethylMapper (11) is a perl script program for bisulfite sequence analysis that is executable from the command line. It is designed for high-throughput mapping, but not for routine bisulfite sequencing. BiQ Analyzer (10) and CpGviewer (12) are local executable programs with a graphical and interactive user interface, and they accept raw bisulfite sequences as input data. One of the important features of the BiQ analyzer is a quality control function for evaluating bisulfite-unconversion, a typical experimental artifact in bisulfite sequencing, which is based on a comparison with the frequency of unconverted cytosines at non-CpG sites. This function is important in practice when evaluating results from the vertebrate genome, which is methylated primarily at CpG sites. However, the BiQ analyzer requires a significantly longer time to execute the analysis compared with the other programs, especially when a large number of sequences are used. CpGviewer has several useful features, including the ability to accept electropherogram formats as input sequence data, but it does not have the quality check function for bisulfite-unconversion like the BiQ analyzer.

Here, we report an interactive web-based bisulfite sequencing analysis tool called QUMA (quantification tool for methylation analysis) for CpG methylation

*To whom correspondence should be addressed. Tel: +81 78 306 3164; Fax: +81 78 306 3167; Email: okano@cdb.riken.jp
Correspondence may also be addressed to Yuichi Kumaki. Tel: +81 78 306 3166; Fax: +81 78 306 3167; Email: kuma@cdb.riken.jp
Present address:
Masaaki Oda, Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, UK

analysis, intended for routine use by experimental researchers studying organisms with genomes methylated primarily at CpG sites, such as those of vertebrates. QUMA has four major features. First, it is easy-to-use and needs only two types of input: a PCR target genomic sequence and raw bisulfite sequences. With its user-friendly interface, only a few clicks are needed to quickly align, visualize and quantify the bisulfite sequence data in a comprehensive manner. Almost all the displayed data are downloadable. Second, QUMA is an all-in-one tool that includes most of the data-processing functions necessary for the analysis of bisulfite sequences. In addition, many optional parameters are available to change the output style according to the user's preferences. Third, QUMA provides a helpful feature that allows the user to control the quality of aligned sequences easily, by changing the cutoff parameters; if the input data and cutoff parameters are indicated, anyone can reproduce the analysis, by using the QUMA web server. Fourth, QUMA server can be launch locally, on a personal computer connected to a local network, by using a bootable CD. This feature is especially helpful to the researcher who must analyze sensitive data.

WEBSITE USAGE

Overview

QUMA has useful functions for bisulfite sequencing analysis, including bisulfite sequence alignment, trimming raw bisulfite sequence, checking sequence quality, quantifying the methylation status, visualizing methylation profiles by various diagrams and, if necessary, calculating statistics values. The use of the basic features of QUMA is quite simple; that is, the user selects sequence files and submits them. The bisulfite alignment data, summarized analysis data, diagrams of methylation status and figures of methylation patterns can then be displayed and downloaded (Figure 1). The user's manual with detailed descriptions can be downloaded.

Input

The input data for QUMA consist of a target genomic sequence and raw bisulfite sequences. The home page of the QUMA web server contains 'file input' fields for each of these data sets (Figure 2). The acceptable sequence formats for the target genomic sequence file are FASTA, GenBank and plain sequence. The genomic sequence must be unconverted (that is, 'C' should not be converted to 'T', which mimics bisulfite conversion) in the same strand as the region between the primer pairs for the bisulfite PCR. A set of multiple bisulfite sequences may be loaded either as a multi-FASTA format file or as a zipped archive of text sequence files (FASTA, GenBank and plain sequence). Raw bisulfite sequences can be used as input data and it is not necessary to remove the plasmid vector sequences. Most recent operating systems (Mac OS X 10.3 or later and Windows ME or later) support the creation of a zipped archive by default, and it is easy to create such an archive, which can contain even hundreds of sequence files, on a personal computer. Thus, it may be convenient to use a zipped archive as an input file of multiple bisulfite sequences, especially for those who are unfamiliar with the multi-FASTA format. Optional parameter fields, which are represented as 'Show options' by default, allow users to input sequences directly into a text field by 'copy and paste', to change the cutoff values for unconverted cytosines at non-CpG sites and alignment mismatch, and to change the direction of the bisulfite conversion of the genomic sequence.

Output

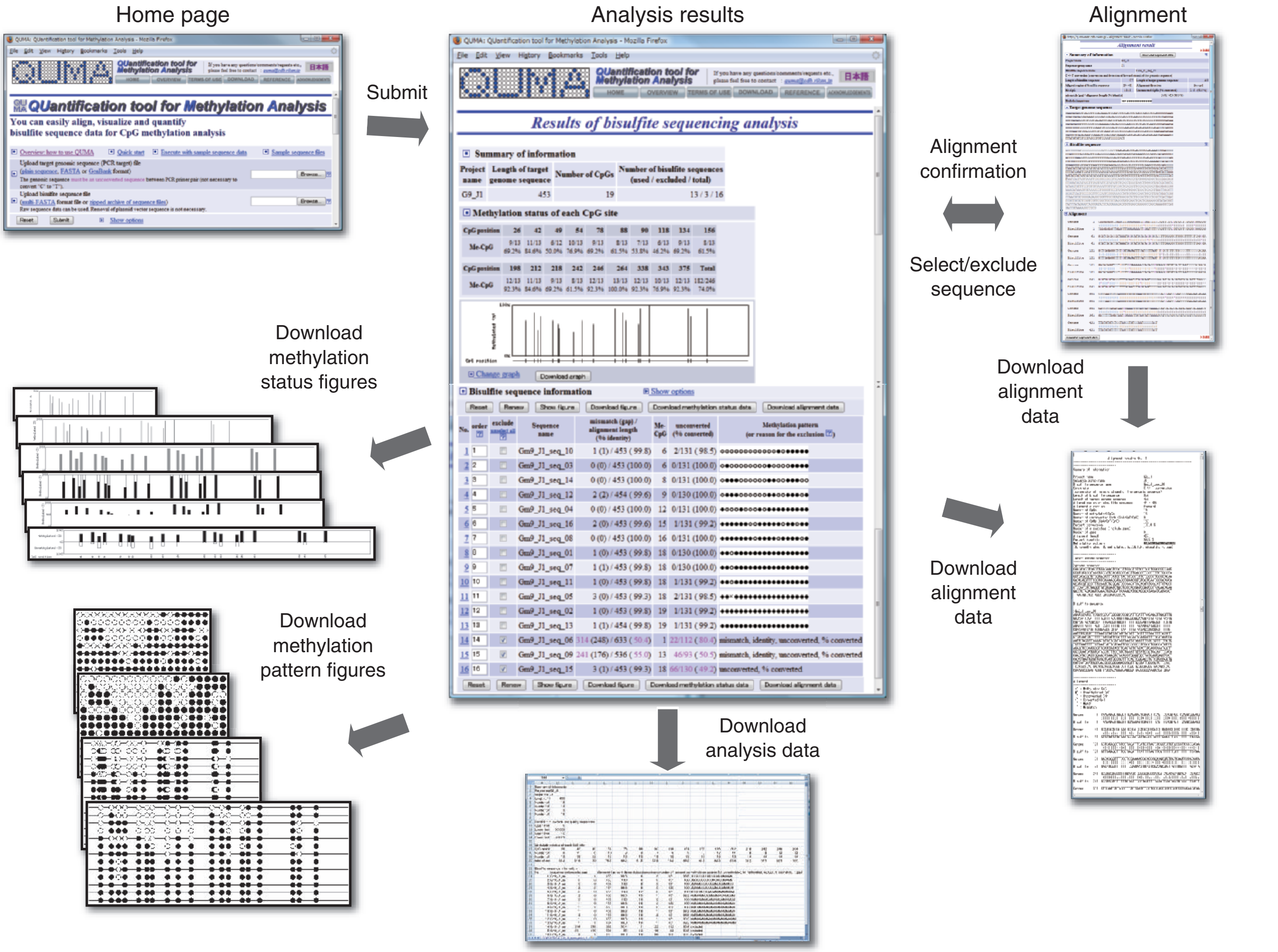
After submitting the input data, bisulfite alignment, sequence trimming, exclusion of problematic sequences and methylation status analysis are performed by the QUMA web server. A typical run takes a few seconds to process 30 bisulfite sequences and the analysis result page is then displayed (Figure 1). The positions of CpG sites and methylation status of each CpG site are shown both in a table and as a diagram, which can be switched to several different formats. The summarized information on sequence alignments and the methylation pattern of each bisulfite sequence are also indicated in a table. A detailed bisulfite alignment between the target genomic sequence and each bisulfite sequence can be displayed from links. DNA methylation patterns can be displayed at the click of a button in several types of black/white circle-style figures, which are frequently used to represent methylation patterns. Optional parameters allow users to change the disposition and resolution of the black/white circle-style figures. Almost all the data shown in the web pages are downloadable in standard file formats (text for alignment data, CSV for analysis data and PNG for graphics), which can be opened by many applications. The user can set the preferences for sorting the order of bisulfite sequence information in the table or use several parameters that can be selected in optional fields.

Quality control

To control the quality of the analysis results, we introduced two types of criteria for detecting problematic sequences: a sequencing quality check and a bisulfite conversion quality check.

Due to the low complexity of sequences after bisulfite conversion, sequencing errors sometimes occur during the bisulfite sequencing analysis. To control for sequence quality, we use the identity and number of mismatches in the local pairwise alignment of the bisulfite sequence and the target genomic sequence as parameters. We adopted a 90% lower limit for identity and 10 as the upper limit of mismatches as default values.

Incomplete bisulfite conversion of cytosines is a frequently encountered problem in bisulfite sequencing analysis (6–8). Although low-level non-CpG methylation does occur in the mammalian genome (15), the presence of a certain number of unconverted cytosines in a single bisulfite sequence clone has been a practical indicator of incomplete bisulfite conversion in analyzing mostly CpG-methylated genome of vertebrates (16). Therefore, to control for the quality of bisulfite conversion, we adopted the conversion efficiency of cytosines at non-CpG sites and



QUantification tool for Methylation Analysis

You can easily align, visualize and quantify bisulfite sequence data for CpG methylation analysis

QUMA: Quantification tool for Methylation Analysis - Mozilla Firefox

Results of bisulfite sequencing analysis

Summary of information

Project name	Length of target genome sequence	Number of CpGs	Number of bisulfite sequences (used / excluded / total)
G9_J1	453	19	13 / 3 / 16

Methylation status of each CpG site

CpG position	26	42	49	54	78	88	90	118	134	156
Me-CpG	9/13 69.2%	11/13 84.6%	6/12 50.0%	10/13 76.9%	9/13 69.2%	8/13 61.5%	7/13 53.8%	6/13 46.2%	9/13 69.2%	8/13 61.5%

CpG position	198	212	218	242	246	264	338	343	375	Total
Me-CpG	12/13 92.3%	11/13 84.6%	9/13 69.2%	8/13 61.5%	12/13 92.3%	13/13 100.0%	12/13 92.3%	10/13 76.9%	12/13 92.3%	182/246 74.0%

Bisulfite sequence information

No.	order	exclude	Sequence name	mismatch (gap) / alignment length (% identity)	Me-CpG	unconverted (% converted)	Methylation pattern (or reason for the exclusion)
1	1	□	Gm9_J1_seq_10	1 (1) / 453 (99.8)	6	2/131 (98.5)	○○○○○○○○○○○○●○●●●●●
2	2	□	Gm9_J1_seq_03	0 (0) / 453 (100.0)	6	0/131 (100.0)	○●○○○○○○○○●○○○○●●●●
3	3	□	Gm9_J1_seq_14	0 (0) / 453 (100.0)	8	0/131 (100.0)	○●●●○○○○○○●●○○●●●○○
4	4	□	Gm9_J1_seq_12	2 (2) / 454 (99.6)	9	0/130 (100.0)	○●●●○○○○○○●●○○●●●●○
5	5	□	Gm9_J1_seq_04	0 (0) / 453 (100.0)	12	0/131 (100.0)	●○○○●○○○●○●●●●●●●●●
6	6	□	Gm9_J1_seq_16	2 (0) / 453 (99.6)	15	1/131 (99.2)	●●●●●●○●●●●●●○○●○●●
7	7	□	Gm9_J1_seq_08	0 (0) / 453 (100.0)	16	0/131 (100.0)	●●●●●●●○●●●●●●○●●●○
8	8	□	Gm9_J1_seq_01	1 (0) / 453 (99.8)	18	0/130 (100.0)	●○●●●●●●●●●●●●●●●●●
9	9	□	Gm9_J1_seq_07	1 (1) / 453 (99.8)	18	0/130 (100.0)	●○●●●●●●●●●●●●●●●●●
10	10	□	Gm9_J1_seq_11	1 (0) / 453 (99.8)	18	1/131 (99.2)	●○●●●●●●●●●●●●●●●●●
11	11	□	Gm9_J1_seq_05	3 (0) / 453 (99.3)	18	2/131 (98.5)	●●×●●●●●●●●●●●●●●●●
12	12	□	Gm9_J1_seq_02	1 (0) / 453 (99.8)	19	1/131 (99.2)	●●●●●●●●●●●●●●●●●●●
13	13	□	Gm9_J1_seq_13	1 (1) / 454 (99.8)	19	1/131 (99.2)	●●●●●●●●●●●●●●●●●●●
14	14	☑	Gm9_J1_seq_06	314 (248) / 633 (50.4)	1	22/112 (80.4)	mismatch, identity, unconverted, % converted
15	15	☑	Gm9_J1_seq_09	241 (176) / 536 (55.0)	13	46/93 (50.5)	mismatch, identity, unconverted, % converted
16	16	☑	Gm9_J1_seq_15	3 (1) / 453 (99.3)	18	66/130 (49.2)	unconverted, % converted

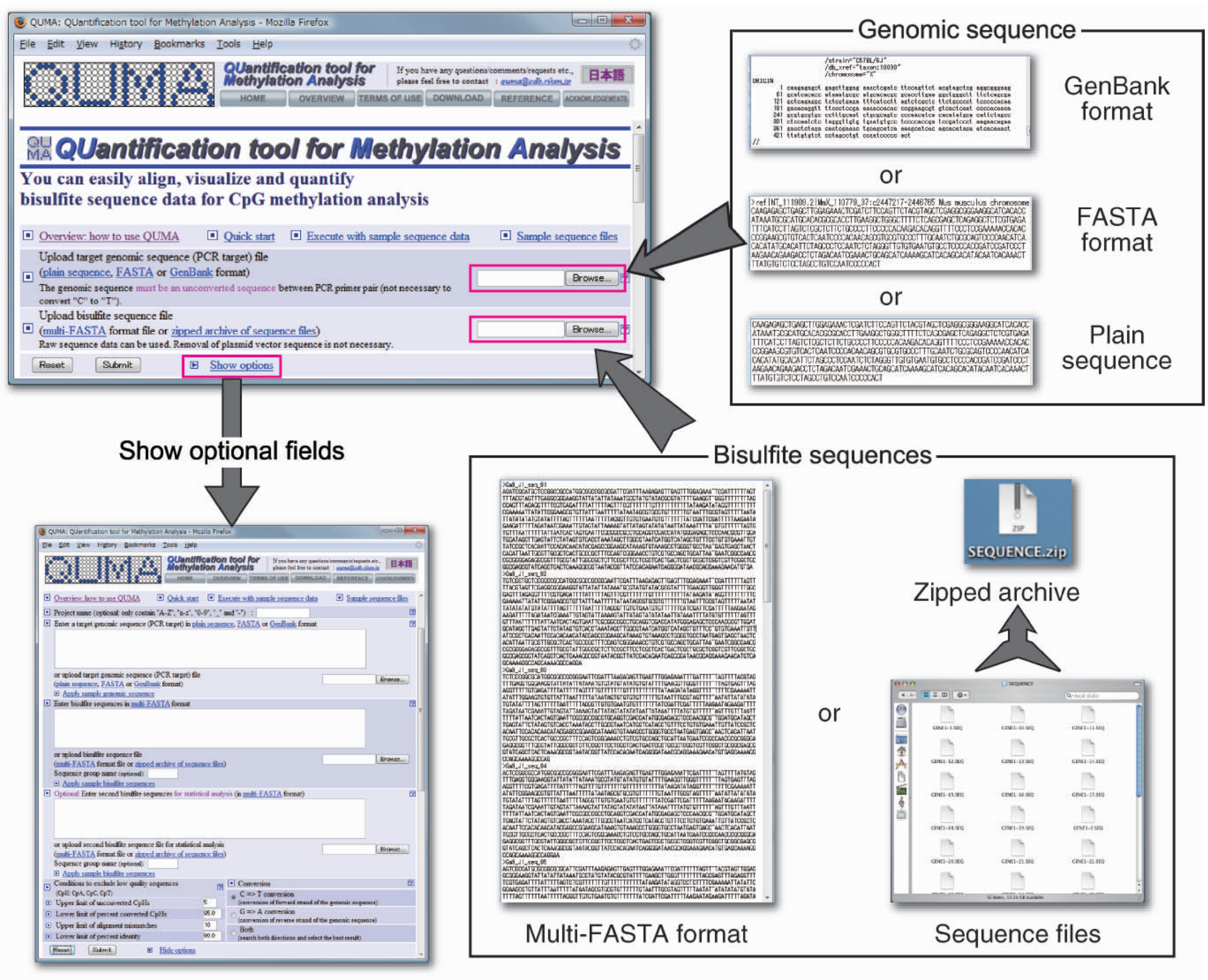
Figure 1. Flow and output of the QUMA web server. After data submission, an analytical results page is displayed. Methylation status data for each CpG site, a summary of the sequence alignment of the bisulfite sequences, and the methylation pattern of each sequence are shown on this page. The alignment of the genomic sequence with each bisulfite sequence can be shown in a different window. The analysis results data (CSV format, which can be opened from Excel or other spread-sheet software), alignment data (text format), several types of methylation status figures and methylation pattern figures (PNG format) can be downloaded.

the number of unconverted cytosines at non-CpG sites as parameters. We set 95% as the lower limit for the conversion efficiency and 5 for the upper limit of the number of unconverted cytosines as default values. The default values of the parameters were determined based on our experience and usually work well for analyzing mammalian CpG methylation.

QUMA shows these parameters for all bisulfite sequences in a summary table, and displays a diagram of DNA methylation patterns excluding sequences that do not match the above criteria. Users can check the sites of the mismatches and unconverted non-CpG cytosines in detailed alignments of these sequences, decide on their inclusion or exclusion, and, if necessary, change the cutoff values of the parameters through QUMA's interactive interface. This option is useful in analyzing polymorphism-containing sequences (e.g. polymorphic alleles of imprinted genes and repetitive sequences) or sequences from organisms with genomes that contain significant amounts of non-CpG methylation.

Statistical analysis

In cases in which two groups of bisulfite sequences are fed into the optional fields on the home page of the QUMA web server (Figure 2), QUMA performs a statistical analysis between the methylation profiles of the two groups. In addition to the standard analysis results, statistical significances (P-values) and a diagram of comparative methylation status are shown (Figure 3). The statistical significance of the difference between two bisulfite sequence groups at each CpG site is evaluated with Fisher's exact test (17), while that of the entire set of CpG sites is evaluated with the Mann–Whitney U-test (18). For a statistical analysis at each CpG site, the two-tailed P-value of Fisher's exact test is calculated from the 2×2 tables at each CpG site. This P-value is used to show the independence of CpG methylation between two groups at a given CpG site. For statistical analysis of the entire set of CpG sites, the two-tailed P-value of the Mann–Whitney U-test is determined from ranks of the ratio of methylated CpGs to all


QUMA: QUantification tool for Methylation Analysis - Mozilla Firefox

QUantification tool for Methylation Analysis

If you have any questions/comments/requests etc., please feel free to contact : quma@cdb.riken.jp

QUantification tool for Methylation Analysis

You can easily align, visualize and quantify bisulfite sequence data for CpG methylation analysis

▣ Overview: how to use QUMA ▣ Quick start ▣ Execute with sample sequence data ▣ Sample sequence files

▣ Upload target genomic sequence (PCR target) file (plain sequence, FASTA or GenBank format)
The genomic sequence must be an unconverted sequence between PCR primer pair (not necessary to convert "C" to "T").

▣ Upload bisulfite sequence file (multi-FASTA format file or zipped archive of sequence files)
Raw sequence data can be used. Removal of plasmid vector sequence is not necessary.

Reset Submit ▣ Show options

▣ Project name (optional: only contain "A-Z", "a-z", "0-9", "_" and "-") :

▣ Enter a target genomic sequence (PCR target) in plain sequence, FASTA or GenBank format

or upload target genomic sequence (PCR target) file (plain sequence, FASTA or GenBank format)

▣ Apply sample genomic sequence

▣ Enter bisulfite sequences in multi-FASTA format

or upload bisulfite sequence file (multi-FASTA format file or zipped archive of sequence files)

Sequence group name (optional):

▣ Apply sample bisulfite sequences

▣ Optional: Enter second bisulfite sequences for statistical analysis (in multi-FASTA format)

or upload second bisulfite sequence file for statistical analysis (multi-FASTA format file or zipped archive of sequence files)

Sequence group name (optional):

▣ Apply sample bisulfite sequences

▣ Conditions to exclude low quality sequences (CpH: CpA, CpC, CpT)

▣ Upper limit of unconverted CpHs: 5

▣ Lower limit of percent converted CpHs: 95.0

▣ Upper limit of alignment mismatches: 10

▣ Lower limit of percent identity: 90.0

▣ Conversion

● C => T conversion (conversion of forward strand of the genomic sequence)

○ G => A conversion (conversion of reverse strand of the genomic sequence)

○ Both (search both directions and select the best result)

Reset Submit ▣ Hide options

Figure 2. The home page and input data formats of the QUMA web server. QUMA needs two types of input files: a PCR target genomic sequence and bisulfite sequences. Acceptable file formats are plain sequence, FASTA or GenBank format for the genomic sequence, and a multi-FASTA format or a zipped archive of sequence files for bisulfite sequences. Optional fields can be seen from the 'Show options' link.

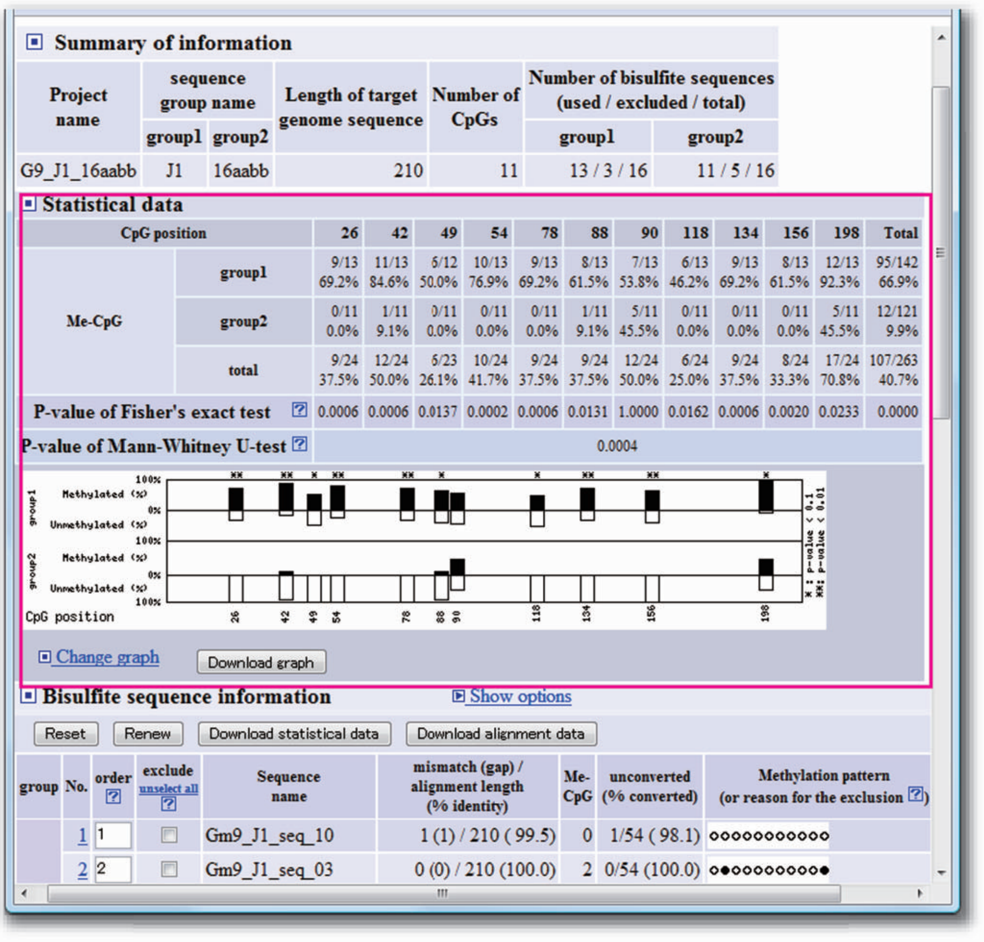
the CpGs at each bisulfite sequence. This *P*-value indicates the independence of distribution of the ratio of CpG methylation to all CpGs.

One limitation is that the statistical analysis by both tests does not take the CpG methylation pattern into account.

IMPLEMENTATION

QUMA is implemented in HTML, JavaScript and Perl-CGI script. The QUMA web server is a Linux server (2.66 GHz Intel Core2Quad Processor, 8 GB RAM) running on CentOS 5. The QUMA web server limits the maximum number of bisulfite sequences per request to 400 to prevent the server from crashing during periods of high use. This should take <60 s (35 s for 450 bp genomic sequence and 384 bisulfite sequences that contain a total of 370 kb). The download version does not limit the maximum number of sequences per analysis request.

To align bisulfite sequences with a genomic sequence, we took a pairwise-alignment approach for QUMA, instead of the multiple-alignment approach used by some other programs (6,10,13,14). QUMA aligns a target genomic sequence with each bisulfite sequence, and therefore the alignment of a certain bisulfite sequence is independent of other bisulfite sequences. However, an alignment between the target sequence and a given bisulfite sequence in a multiple-alignment could be affected by the qualities of other bisulfite sequences. In addition, a pairwise-alignment approach is suitable for prompt interactive changes to the inclusion or exclusion of bisulfite sequences during a quality-check process, because of the independence of each bisulfite sequence result. An output style of pairwise alignment may be suitable for checking mismatches and unconverted cytosines in individual bisulfite sequence clones.



▣ Summary of information

Project name	sequence group name		Length of target genome sequence	Number of CpGs	Number of bisulfite sequences (used / excluded / total)	
	group1	group2			group1	group2
G9_J1_16aabb	J1	16aabb	210	11	13 / 3 / 16	11 / 5 / 16

▣ Statistical data

CpG position		26	42	49	54	78	88	90	118	134	156	198	Total
Me-CpG	group1	9/13 69.2%	11/13 84.6%	6/12 50.0%	10/13 76.9%	9/13 69.2%	8/13 61.5%	7/13 53.8%	6/13 46.2%	9/13 69.2%	8/13 61.5%	12/13 92.3%	95/142 66.9%
	group2	0/11 0.0%	1/11 9.1%	0/11 0.0%	0/11 0.0%	0/11 0.0%	1/11 9.1%	5/11 45.5%	0/11 0.0%	0/11 0.0%	0/11 0.0%	5/11 45.5%	12/121 9.9%
	total	9/24 37.5%	12/24 50.0%	6/23 26.1%	10/24 41.7%	9/24 37.5%	9/24 37.5%	12/24 50.0%	6/24 25.0%	9/24 37.5%	8/24 33.3%	17/24 70.8%	107/263 40.7%
P-value of Fisher's exact test ?		0.0006	0.0006	0.0137	0.0002	0.0006	0.0131	1.0000	0.0162	0.0006	0.0020	0.0233	0.0000
P-value of Mann-Whitney U-test ?		0.0004											

▣ Bisulfite sequence information ▣ Show options

group	No.	order ?	exclude unselect all ?	Sequence name	mismatch (gap) / alignment length (% identity)	Me-CpG	unconverted (% converted)	Methylation pattern (or reason for the exclusion ?)
	1	1	□	Gm9_J1_seq_10	1 (1) / 210 (99.5)	0	1/54 (98.1)	○○○○○○○○○○○
	2	2	□	Gm9_J1_seq_03	0 (0) / 210 (100.0)	2	0/54 (100.0)	○●○○○○○○○○●

Figure 3. An example of statistical analysis output at the QUMA web server. In the statistical data section (magenta rectangle), the position of CpG sites, methylation status of each CpG site, and statistical significance (*P*-values) of differences between two bisulfite sequence groups are shown. A diagram representing the comparative methylation status can be displayed in several different formats.

To perform pairwise alignment, we adopted the needle program of the EMBOSS package (19) because of its accurate production of Needleman–Wunsch pairwise alignments (20). A modified score matrix was used for the needle, so that 'C' bases in the genomic sequence may match with both 'C' and 'T' bases in the bisulfite sequence, since unmethylated cytosine is detected as 'T' in bisulfite sequencing.

The software source code of the QUMA is freely available on the download section of the QUMA web server (http://quma.cdb.riken.jp/) under the GNU General Public License (GPL). Also, a CD-image file of a 'LiveCD' Linux version of QUMA is freely available at the download section. LiveCD is a computer operating system that easily boots a personal computer from a single CD without installation, set-up or changes to the hard disk. After rebooting and ejecting the CD, a computer can be rebooted from its original operating system. The LiveCD version of QUMA enables the user to launch the QUMA server easily on a personal computer connected to a local network, and the server will be accessible from other personal computers connected to the local network. This option may be useful for analyzing sensitive data without submitting sequences to a site outside the user's institution.

CONCLUSION

QUMA is an easy-to-use, all-in-one, interactive web-based tool for the bisulfite-sequencing analysis of CpG methylation that aligns and trims raw sequences, analyzes CpG methylation profiles, performs statistical comparisons, checks the quality of sequencing data and displays the results. QUMA is designed to be useful and understandable for experimental researchers, even if they are unfamiliar with bisulfite sequencing analysis. This tool may improve the speed of the entire analysis of bisulfite sequencing data, and may help standardize the analysis, communication and quality of results across different computing environments. This website is free and open to all users and there is no login requirement.

ACKNOWLEDGEMENTS

We thank Akiko Yamagiwa for sample bisulfite sequences of the mouse Gm9 region (21), Morito Sakaue, Masahumi

Kawaguchi and Yoko Dote for useful comments, Hazuki S. Hiraga for checking the website, and the Information Networks Office of RIKEN Kobe Institute for helpful suggestions in setting up the Internet connection for the server. This work was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan to M. Okano. Funding to pay the Open Access publication charges for this article was provided by RIKEN.

Conflict of interest statement. None declared.

REFERENCES

1. Bernstein,B.E., Meissner,A. and Lander,E.S. (2007) The mammalian epigenome. *Cell*, **128**, 669–681.
2. Weber,M. and Schubeler,D. (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr. Opin. Cell. Biol.*, **19**, 273–280.
3. Esteller,M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, **8**, 286–298.
4. Jones,P.A. and Baylin,S.B. (2007) The epigenomics of cancer. *Cell*, **128**, 683–692.
5. Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
6. Grunau,C., Clark,S.J. and Rosenthal,A. (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.*, **29**, e65.
7. Hajkova,P., el-Maarri,O., Engemann,S., Oswald,J., Olek,A. and Walter,J. (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol. Biol.*, **200**, 143–154.
8. Liu,L., Wylie,R.C., Hansen,N.J., Andrews,L.G. and Tollefsbol,T.O. (2004) Profiling DNA methylation by bisulfite genomic sequencing: problems and solutions. *Methods Mol. Biol.*, **287**, 169–179.
9. Grunau,C., Schattevoy,R., Mache,N. and Rosenthal,A. (2000) MethTools—a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Res.*, **28**, 1053–1058.
10. Bock,C., Reither,S., Mikeska,T., Paulsen,M., Walter,J. and Lengauer,T. (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**, 4067–4068.
11. Ordway,J.M., Bedell,J.A., Citek,R.W., Nunberg,A.N. and Jeddeloh,J.A. (2005) MethylMapper: a method for high-throughput, multilocus bisulfite sequence analysis and reporting. *BioTechniques*, **39**, 464–472.
12. Carr,I.M., Valleley,E.M., Cordery,S.F., Markham,A.F. and Bonthron,D.T. (2007) Sequence analysis and editing for bisulphite genomic sequencing projects. *Nucleic Acids Res.*, **35**, e79.
13. Hetzl,J., Foerster,A.M., Raidl,G. and Mittelsten Scheid,O. (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.*, **51**, 526–536.
14. Xu,Y.H., Manoharan,H.T. and Pitot,H.C. (2007) CpG PatternFinder: a Windows-based utility program for easy and rapid identification of the CpG methylation status of DNA. *BioTechniques*, **43**, 334–342.
15. Ramsahoye,B.H., Biniszkiewicz,D., Lyko,F., Clark,V., Bird,A.P. and Jaenisch,R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl Acad. Sci. USA*, **97**, 5237–5242.
16. Lane,N., Dean,W., Erhardt,S., Hajkova,P., Surani,A., Walter,J. and Reik,W. (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis*, **35**, 88–93.
17. Alan,A. (2002) *Categorical Data Analysis*, 2nd edn. Wiley, Hoboken, NJ.
18. Ewens,W. and Grant,G. (2001) *Statistical Methods in Bioinformatics: An Introduction* (*Statistics for Biology and Health*). Springer, New York, NY.
19. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
20. Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
21. Oda,M., Yamagiwa,A., Yamamoto,S., Nakayama,T., Tsumura,A., Sasaki,H., Nakao,K., Li,E. and Okano,M. (2006) DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev.*, **20**, 3382–3394.