



Novel Methyltransferase Recognition Motif Identified in *Chania multitudinisentens* RB-25^T gen. nov., sp. nov.

Robson Ee¹, Yan-Lue Lim^{1*}, Wai-Fong Yin¹, Wah-Seng See-Too¹, Richard J. Roberts² and Kok-Gan Chan¹

¹ Division of Genetics and Molecular Biology, Faculty of Science, Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia, ² New England Biolabs, Ipswich, MA, USA

Keywords: SMRT sequencing, methylome, REBASE, DNA methylation, prophage, DNA methyltransferase (MTase), restriction-modification (RM) system, restriction endonuclease (REase)

INTRODUCTION

DNA methylation, defined by the addition of a methyl group to adenine or cytosine bases in DNA catalyzed by DNA methyltransferases (MTases), is one of the most studied post-replicative DNA modification mechanism in bacteria (Roberts et al., 2003b). The three forms of nucleotide methylation identified to date are: N6-methyladenine (^{m6}A), N4-methylcytosine (^{m4}C), and 5-methylcytosine (^{m5}C) (Gromova and Khoroshaev, 2003).

Generally, MTases can be classified into two main groups: as part of a restriction modification (RM) system in which the MTase is associated with a cognate restriction endonuclease (REase) or as a solitary MTase, which as the name suggests, serves the role of an independent MTase (Murphy et al., 2013; Roberts et al., 2015). The RM systems, which are found to occur exclusively in unicellular organisms, can be further classified into four groups based on their subunit composition, recognition sequence specificity, substrate specificity, cofactor requirements, and DNA cleavage positions (Wilson and Murray, 1991; Casadesus and Low, 2006). In brief, Type I RM systems comprise three polypeptides which form a hetero-oligomeric protein complex, namely the R (restriction), M (modification), and S (specificity) subunits, that recognizes asymmetric, and bipartite recognition sequences (Murray, 2000). Type II systems, the most ubiquitous, and simplest systems, are composed of two functionally-independent R and M genes, which are responsible for restriction and methylation activity respectively. The recognition sequences of the Type II systems are most often symmetrical, but can also be asymmetric (Wilson and Murray, 1991). Type III system also consist of two subunits, named res, and mod. The mod subunit can function independently as a MTase, but only methylates one strand of the DNA. The res subunit must form a complex with the mod subunit to express its DNA restriction activity, because the recognition specificity is encoded in the mod subunit (Wilson and Murray, 1991). The recognition sequences of the known Type III systems are asymmetric and four to six bases in length. Furthermore, when known, the res enzyme requires the presence of two unmethylated recognition sites for efficient DNA cleavage (Rao et al., 2013). Lastly, the Type IV systems are restriction enzymes that recognize and cleave only methylated DNA (Vasu and Nagaraja, 2013). The sequence specificities of the Type IV systems are not well studied. MTases can be further sub-classified according to the order of their conserved amino acid motifs, which represent the DNA binding domain, the target recognition domain (TRD) and the catalytic domain (Malone et al., 1995; Jeltsch, 2002).

RM systems are often considered as the most primitive defense mechanism of prokaryotes against the invasion of extraneous DNA elements (Wilson and Murray, 1991; Bickle, 2004).

OPEN ACCESS

Edited by:

Wendy Schluchter,
University of New Orleans, USA

Reviewed by:

Jing-Sheng Cheng,
Tianjin University, China
Lukasz Dziejewit,
University of Warsaw, Poland

*Correspondence:

Yan-Lue Lim
yanluelim@hotmail.my

Specialty section:

This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

Received: 01 July 2016

Accepted: 17 August 2016

Published: 31 August 2016

Citation:

Ee R, Lim Y-L, Yin W-F, See-Too W-S,
Roberts RJ and Chan K-G (2016)
Novel Methyltransferase Recognition
Motif Identified in *Chania
multitudinisentens* RB-25^T gen. nov.,
sp. nov. *Front. Microbiol.* 7:1362.
doi: 10.3389/fmicb.2016.01362

However, more recent studies have offered new insights into their additional biological roles, including genomic island stabilization, species identity maintenance, generating, and enhancing genomic diversity for host fitness and adaptability as well as the regulation of gene expression (Vasu and Nagaraja, 2013).

The development of Single Molecule Real-Time (SMRT) sequencing technology which enables the simultaneous genome-wide detection of MTase activity during genome sequencing has permitted the rapid detection of novel MTase recognition motifs in various prokaryotes including *Helicobacter pylori*, *Salmonella enterica*, *Escherichia coli*, and *Campylobacter coli* (Krebes et al., 2014; Forde et al., 2015; Lee et al., 2015; Pirone-Davies et al., 2015; Zautner et al., 2015). These data are important not only for the identification of novel types of MTases, but they also lay the groundwork for the discovery of new biological roles for DNA methylation. Furthermore, these new data on previously characterized methylomes can further extend our understanding of these RM systems.

Chania multitudinisentens RB-25^T gen. nov., sp. nov. is a fully characterized newly proposed novel genus in the family of *Enterobacteriaceae*. Isolated in a soil sample collected from a former municipal landfill site, *C. multitudinisentens* RB-25^T was initially misidentified as a member of the *Serratia* genus (Ee et al., 2014a). Further in-depth investigation later reclassified *C. multitudinisentens* RB-25^T as a novel genus (Ee et al., 2016). To date, *C. multitudinisentens* RB-25^T is only characterized for its quorum sensing properties (the production of C4-HSL, C6-HSL, and 3-oxo-C6-HSL) and its potential chitinolytic activity (Ee et al., 2014a; Lim et al., 2015b). In this study, we report a novel methyltransferase recognition motif in *C. multitudinisentens* RB-25^T, which will serve as the foundation for future investigation of the role of methylation in the genus *Chania*.

MATERIALS AND METHODS

Genome Sequencing and Assembly

Genomic DNA extraction was performed using a MasterpureTM DNA purification kit (Epicentre) on pelleted bacterial cells from an overnight culture. Briefly, the bacterial cells were first transferred into a mixture of Tissue and Cell Lysis Solution with proteinase K, and subsequently treated with RNase A prior to the addition of MPC protein precipitation reagent. Pure isopropanol was used to precipitate DNA and the precipitated DNA pellet was washed twice with 75% ethanol and resuspended in Buffer EB (Qiagen). The extracted gDNA was quantified using a Qubit fluorometer (Thermo Fisher Scientific) and was subsequently examined for its integrity using 0.8% agarose gel electrophoresis. A NanodropTM spectrophotometer (Thermo Fisher Scientific) was used to quantify the extracted gDNA and to determine its purity. Following the “Procedure and Checklist-20 kb Template Preparation Using BluePippinTM Size-selection system” protocol (Pacific Biosciences), a size-selected SMRTbellTM library (7 kb as size-selection cut-off length) was prepared. After annealing and polymerase binding, the library was subsequently sequenced in a PacBio RSII sequencer (Pacific Biosciences) using P5-C3 chemistry on 2 SMRT cells.

De novo assembly of this genome was performed as described previously in which the sequenced polymerase reads underwent three major steps within the Hierarchical Genome Assembly Process (HGAP) assembly pipeline (Pacific Biosciences), namely sub-read filtering, preassembly, assembly, and consensus polishing to generate a highly accurate polished assembly (Chin et al., 2013; Ee et al., 2014b, 2015; Lim et al., 2015a). Circularity of the assembled genome, denoted by the presence of self-overlapping ends, was determined and visualized using Contiguity (Sullivan et al., 2015) and Gepard (Krumstiek et al., 2007). The overlapping ends were finally circularized using the Minimus2 pipeline included in the AMOS software package (Treangen et al., 2011).

Genome Annotation

Genome annotation was performed using three different server-based genome annotation pipelines namely: Rapid Annotation using Subsystem Technology, RAST version 2.0 (Aziz et al., 2008); Rapid Prokaryotic Genome Annotation, PROKKA (Seemann, 2014); and NCBI Prokaryotic Genome Annotation Pipeline, PGAP. Genes of interest were subsequently curated manually by comparison between the annotation results of different pipelines followed by BLAST comparison against NCBI nucleotide (nr/nt) database and NCBI Reference Sequence, RefSeq database. Conserved domain analyses and functional protein predictions were subsequently performed using a combination of various protein model databases including NCBI-curated domains; Simple Modular Architecture Research Tool, SMART (Letunic et al., 2015); Protein Families, Pfam database (Finn et al., 2014); Clusters of Orthologous Groups of proteins, COGs database (Tatusov et al., 2003); The Institute for Genomic Research's database of protein families, TIGRFAM (Haft et al., 2003); and InterPro protein families database (Mitchell et al., 2015). PHAST analysis was also performed to identify the presence of prophages in the genome (Zhou et al., 2011).

Base Modification Analysis

The complete genome sequence was uploaded into the SMRT portal and was processed into an *in silico* kinetic reference. Subsequently, SMRT analysis RS_Modification_and_Motif_Analysis.1 protocol was used to perform epigenome analysis using the imported reference sequence. A default modification quality value (QV) score of 30 (correspond to a *p*-value of 0.001) was used to call the modified bases.

RM System Genes Annotation

By using the SEQWARE computer resources, which constitutes BLAST-based software modules in combination with REBASE curated internal databases, the genome sequence was scanned for homologs of RM system genes (Murray et al., 2012). As described previously, the prediction were supported by sequence similarity, presence, and order of predictive functional motifs, and the known genomic context and characteristics of previously characterized RM system genes (Murray et al., 2012; Pirone-Davies et al., 2015). The predicted RM system genes were named

following the proposed nomenclature as described by Roberts et al. (2003a). Assignment of the predicted RM system genes to the identified recognition motifs was performed based on putative methyltransferases sequence homology and RM system Type pairing.

RESULTS

Genome Characteristics

The first draft genome sequence of *C. multitudinisentens* RB-25^T which was deposited in GenBank with the accession number CP007044.1 was a non-contiguous single contig assembly (35-fold coverage) constructed from sequence data generated using a single SMRT cell. Here we report the updated version of the genome sequence (CP007044.2) which with the sequence data generated from an additional SMRT cell, enabled successful assembly into a single circular chromosome of 5,488,183 bp (85.4-fold coverage) with 50.90% GC content. A total of 4883 genes, 4540 coding DNA sequences, 243 pseudo genes, 3 CRISPR arrays, 22 rRNAs, 77 tRNAs, and 1 ncRNA were annotated in the genome. Genome sequence data are available at GenBank in the format of FASTA, annotated GenBank flat file, graphical, and ASN.1 formats.

Following annotation using a prophage search tool (PHAST), two intact prophages, an incomplete prophage and a questionable prophage were identified from the genome (Supplementary Tables 1–5).

Base Modification Analysis

Genome-wide base modification analysis revealed a distinct population of adenine bases with high modification QV (modQV) value, as demonstrated by the distinct separation of the red adenine cloud from the background in the modQV vs coverage scatterplot (Supplementary Figure 1). Motif analysis detected a total of two methyltransferase recognition motifs. The first motif is 5'-G^{m6}ATC-3', a common motif pattern corresponding to the well-known specificity pattern for many Type II alpha subtype MTases (Barras and Marinus, 1989; Henaut

et al., 1996; Low et al., 2001). Methylation of the adenine residues in this motif is known to influence the regulation of various cellular events for instance gene expression, virulence, DNA replication coordination, as well as normal cellular processes (Low et al., 2001; Murphy et al., 2013). The second motif detected is a novel recognition motif, 5'-GC^{m6}AGNNNNNTCC-3' with its partner motif 5'-GG^{m6}ANNNNNCTGC-3'. This motif matched the asymmetric, and bipartite structure of a typical Type I recognition sequence with a 3 nucleotides and a 4 nucleotides component separated by a nonspecific spacer sequence (represented by N). Across the genome, more than 99% of all genomic positions, which correspond to both methyltransferase recognition motifs were found to be methylated (Supplementary Table 6).

Restriction Modification System Analysis

Annotation of the *C. multitudinisentens* RB-25^T genome identified seven putative DNA methyltransferase-encoding genes that might be associated with Type II RM systems and one MTase gene that would be part of a Type I RM system. Among the eight annotated MTases, three were assigned to the MTase recognition motifs detected in base modification analysis, whereas another three MTases were predicted to be potentially inactive and two of the MTases had unknown specificity. Each of the identified MTases and the predicted specificities are summarized in **Table 1**.

Firstly, based on the matching RM system types, the CmuRB25ORF16340P system, the only Type I RM system annotated in the genome was assigned to be responsible for the novel Type I motif, 5'-GC^{m6}AGNNNNNTCC-3'. The S subunit responsible for recognizing this sequence was renamed S.CmuRB25I and its partner MTase was renamed M.CmuRB25I (Supplementary Table 7). S.CmuRB25I was a typical Type I S subunit and its N-terminal target recognition domain (TRD1) showed a reasonable similarity to the C-terminal TRD (TRD2) of S.Eco3609I (TGHAYNNNNCTNC). M.CmuRB25I was a typical N⁶ DNA methylase of the gamma subtype.

Secondly, two Type II solitary methyltransferases, M.CmuRB25DamP, and M.CmuRB25ORF1485P, were predicted

TABLE 1 | MTases and their predicted specificities identified in *Chania multitudinisentens* RB-25^T.

RM system genes annotated	Type (subtype)	Locus tag (bp start/stop) ^a	Predicted recognition motif ^b	Methylation type
M.CmuRB25I	I (gamma)	Z042_16340 (5441947/5444100)	GG ^{m6} AGNNNNNTCC	6 mA
S.CmuRB25I	I	Z042_16335 (5444100/5445641)	GCAGNNNNNTCC	
M.CmuRB25DamP	II (alpha)	Z042_10170 (1345643/1346455)	G ^{m6} ATC	6 mA
M.CmuRB25ORF1485P	II (alpha)	Z042_01485 (3271527/3272399)	G ^{m6} ATC	6 mA
M.CmuRB25ORF14775P	II	(319513/320601)	GTCGAC	5 mC
CmuRB25ORF14775P	II (P)	(318782/319516)	GTCGAC	
M.CmuRB25ORF23095P	II	Z042_23095 (4027210/4029192)	–	–
M.CmuRB25ORF23090P	II (beta)	Z042_23090 (4029189/4030583)	AAGCTT	–
M.CmuRB25ORF23015P	II (gamma)	Z042_23015 (4038239/4038991)	–	–
M.CmuRB25ORF22915P	II (alpha)	(4049231/4050046)	TGGCCA	–
CmuRB25McrBCP	IV	Z042_16355 (5438103/5440178)	–	–

^aFor predicted MTases which do not have a GenBank locus tag, genes coordinates are provided in the parenthesis (bp start/stop).

^bModified bases are highlighted in bold.

to be the candidate enzymes responsible for formation of the G^{m6}ATC motif due to their sequence similarity to various MTases which were previously predicted or documented to recognize G^{m6}ATC (Supplementary Tables 8–9). Both MTases contained the characteristic D12 class N⁶-adenine-specific DNA methyltransferase pfam domain (pfam02086) and were grouped into COG0338 (site specific DNA-adenine methylase). M.CmuRB25DamP is potentially the constitutive G^{m6}ATC methyltransferase of *C. multitudinisentens* as it showed higher sequence similarity (93%) to G^{m6}ATC methyltransferases of *Serratia* species, including one of the REBASE gold standard MTases, M.SmaII in *Serratia marcescens* strain Sb. On the other hand, M.CmuRB25ORF1485P shows only 54% of amino acid identity to various G^{m6}ATC methyltransferases from *Dickeya* and *Yersinia* species and was identified to be in close proximity to a predicted intact prophage sequence which led us to speculate that this gene could be a prophage-encoded MTase. Even more interesting, the MTases which have sequence similarity to M.CmuRB25ORF1485P are also MTases which were found to be located within intact prophage sequences, hence further supporting our hypothesis (data not shown). Prophage-derived Dam homologs are prevalent, examples of prophages found to encode Dam homologs include phage T4, phage P1, Shiga toxin-encoding 933 W phage, and *Pseudomonas* phage B3 (Sternberg and Coulby, 1990; Miller et al., 2003; Braid et al., 2004; Murphy et al., 2008). These Dam homologs were found to serve various function in the phage genomes, including phage genome protection role from host REases, phage packaging, and progeny release, maintenance of phage lysogeny, and regulation of phage gene expression (Sternberg et al., 1986; Sternberg and Coulby, 1990; Lobočka et al., 2004; Murphy et al., 2013). Although the majority of MTases encoded by prophages were observed to be active only during the lytic stage (Trautner et al., 1980; Citron et al., 1989), we cannot rule out the possibility that M.CmuRB25ORF1485P could also contribute to formation of G^{m6}ATC methylation. Therefore, due to the presence of two possible MTase candidates for motif G^{m6}ATC, no confirmed assignment was made.

Another three MTases were also predicted to recognize specific target sites based on their sequence similarity to characterized MTases. The first is M.CmuRB25ORF14775P, which is predicted to encode a ^{m5}C DNA MTase that recognizes 5'-GTCGAC-3' based on matching conserved amino acid motif pattern and 44% amino acid sequence identity to M.BbrUII, which is known to recognize GTCGAC (O'Connell Motherway et al., 2009). Other methylases of similar specificity can be found in Supplementary Table 10. Adjacent to this MTase, in a tandem transcriptional orientation with a 3 bp overlap, is predicted to be its cognate REase, CmuRB25ORF14775P. The second example is M.CmuRB25ORF23090P, a Type II, subtype beta solitary MTase which shares sequence similarity (60%) to MTases that recognize 5'-AAGCTT-3' (Supplementary Table 11). The third one, M.CmuRB25ORF22915P, is a Type II orphan methyltransferase (subtype alpha), and

is predicted to recognize 5'-TGGCCA-C' due to its high sequence similarity (75%) to a large number of MTases (>30) predicted to recognize this target motif (Supplementary Table 12).

Lastly, two MTases, M.CmuRB25ORF23095P and M.CmuRB25ORF23015P, were annotated, but with unknown recognition sequences. Both MTases are Type II solitary MTases, in which, based on their conserved amino acid motif arrangement, M.CmuRB25ORF23095P is predicted to be a ^{m5}C MTase whereas M.CmuRB25ORF23015P is an amino MTase of the gamma subtype.

Data Access

The methylome data of *C. multitudinisentens* strain RB-25^T are accessible through organism number 12049 in REBASE, in which the full profile of annotated RM system genes in the form of a summary table can be accessed. Through the complete genome hyperlink, various information which could aid in understanding the complete methylome of *C. multitudinisentens* is available, including: visualization of the annotated RM system genes location on a circular genome map (Supplementary Figure 2), schematics of the conserved amino acid motifs on each annotated RM system genes as well as the genomic arrangement of the associated specificity subunits or REases, summary of the detailed annotation information of each RM system genes, and the annotation report. The detailed characteristics and sequence data of all of the RM system genes detected in the genome can be viewed through the hyperlink of each gene listed on the summary table. Cross reference information through other database such as Expasy, InterPro, Pfam, ProDom, ProtoMap, and PRESAGE can also be accessed through the “sequence data” hyperlink. Details of prophages are available in the PHAST database with the NC number of CP007044.2 (Web link: <http://phast.wishartlab.com/cgi-bin/Results.cgi?num=CP007044.2&multi=1>).

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This research was supported by the University of Malaya Grants (A-000001-50001, H-50001-A000027, and PG080-2015B) awarded to K-GC. and by the Small Business Innovation Research Program (National Institute of General Medical Sciences) of the National Institutes of Health award R44GM105125 awarded to RJR.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01362>

REFERENCES

- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Barras, F., and Marinus, M. G. (1989). The great GATC: DNA methylation in *E. coli*. *Trends Genet.* 5, 139–143. doi: 10.1016/0168-9525(89)90054-1
- Bickle, T. A. (2004). Restricting restriction. *Mol. Microbiol.* 51, 3–5. doi: 10.1046/j.1365-2958.2003.03846.x
- Braid, M. D., Silhavy, J. L., Kitts, C. L., Cano, R. J., and Howe, M. M. (2004). Complete genomic sequence of bacteriophage B3, a Mu-Like phage of *Pseudomonas aeruginosa*. *J. Bacteriol.* 186, 6560–6574. doi: 10.1128/jb.186.19.6560-6574.2004
- Casadesus, J., and Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* 70, 830–856. doi: 10.1128/mmb.00016-06
- Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Citron, M., Velleman, M., and Schuster, H. (1989). Three additional operators, Op21, Op68, and Op88, of bacteriophage P1. Evidence for control of the P1 dam methylase by Op68. *J. Biol. Chem.* 264, 3611–3617.
- Ee, R., Lim, Y. L., Tee, K. K., Yin, W. F., and Chan, K. G. (2014a). Quorum sensing activity of *Serratia fonticola* strain RB-25 isolated from an ex-landfill site. *Sensors* 14, 5136–5146. doi: 10.3390/s140305136
- Ee, R., Lim, Y. L., Yin, W. F., and Chan, K. G. (2014b). *De novo* assembly of the quorum-sensing *Pandoraea* sp. strain RB-44 complete genome sequence using PacBio single-molecule real-time sequencing technology. *Genome Announc.* 2:e00245-14. doi: 10.1128/genomeA.00245-14
- Ee, R., Madhaiyan, M., Ji, L., Lim, Y. L., Nor, N. M., Tee, K. K., et al. (2016). *Chania multitudinisentens* gen. nov., sp. nov., a N-acyl-homoserine lactone-producing bacterium in the family of Enterobacteriaceae isolated from a former municipal landfill site soil. *Int. J. Syst. Evol. Microbiol.* 66, 2297–2304. doi: 10.1099/ijsem.0.001025
- Ee, R., Yong, D., Lim, Y. L., Yin, W. F., and Chan, K. G. (2015). Complete genome sequence of oxalate-degrading bacterium *Pandoraea vervacti* DSM 23571. *J. Biotechnol.* 204, 5–6. doi: 10.1016/j.jbiotec.2015.03.020
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Forde, B. M., Phan, M.-D., Gawthorne, J. A., Ashcroft, M. M., Stanton-Cook, M., Sarkar, S., et al. (2015). Lineage-specific methyltransferases define the methylome of the globally disseminated *Escherichia coli* ST131 Clone. *MBio* 6, e01602–01615. doi: 10.1128/mBio.01602-15
- Gromova, E. S., and Khoroshaev, A. V. (2003). Prokaryotic DNA methyltransferases: the structure and the mechanism of interaction with DNA. *Mol. Biol.* 37, 300–314. doi: 10.1023/A:1023301923025
- Haft, D. H., Selengut, J. D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. doi: 10.1093/nar/gkg128
- Henaut, A., Rouxel, T., Gleizes, A., Moszer, I., and Danchin, A. (1996). Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. *J. Mol. Biol.* 257, 574–585. doi: 10.1006/jmbi.1996.0186
- Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBiochem* 3, 274–293. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S
- Krebs, J., Morgan, R. D., Bunk, B., Sproer, C., Luong, K., Parusel, R., et al. (2014). The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* 42, 2415–2432. doi: 10.1093/nar/gkt1201
- Krumsiek, J., Arnold, R., and Rattei, T. (2007). Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* 23, 1026–1028. doi: 10.1093/bioinformatics/btm039
- Lee, W. C., Anton, B. P., Wang, S., Baybayan, P., Singh, S., Ashby, M., et al. (2015). The complete methylome of *Helicobacter pylori* UM032. *BMC Genomics* 16:424. doi: 10.1186/s12864-015-1585-2
- Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949
- Lim, Y. L., Ee, R., How, K. Y., Lee, S. K., Yong, D., Tee, K. K., et al. (2015a). Complete genome sequencing of *Pandoraea pnomenusa* RB38 and molecular characterization of its N-acyl homoserine lactone synthase gene *ppnI*. *PeerJ* 3:e1225. doi: 10.7717/peerj.1225
- Lim, Y. L., Yong, D., Ee, R., Tee, K. K., Yin, W. F., and Chan, K. G. (2015b). Complete genome sequence of *Serratia multitudinisentens* RB-25(T), a novel chitinolytic bacterium. *J. Biotechnol.* 207, 32–33. doi: 10.1016/j.jbiotec.2015.04.027
- Lobocka, M. B., Rose, D. J., Plunkett, G. III, Rusin, M., Samojedny, A., Lehnher, H., et al. (2004). Genome of bacteriophage P1. *J. Bacteriol.* 186, 7032–7068. doi: 10.1128/jb.186.21.7032-7068.2004
- Low, D. A., Weyand, N. J., and Mahan, M. J. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect. Immun.* 69, 7197–7204. doi: 10.1128/iai.69.12.7197-7204.2001
- Malone, T., Blumenthal, R. M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA aminomethyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* 253, 618–632. doi: 10.1006/jmbi.1995.0577
- Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003). Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86–156. doi: 10.1128/MMBR.67.1.86-156.2003
- Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, D213–D221. doi: 10.1093/nar/gku1243
- Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., and van Sinderen, D. (2013). Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl. Environ. Microbiol.* 79, 7547–7555. doi: 10.1128/aem.02229-13
- Murphy, K. C., Ritchie, J. M., Waldor, M. K., Löbner-Olesen, A., and Marinus, M. G. (2008). Dam methyltransferase is required for stable lysogeny of the Shiga toxin (Stx2)-encoding bacteriophage 933W of enterohemorrhagic *Escherichia coli* O157:H7. *J. Bacteriol.* 190, 438–441. doi: 10.1128/JB.01373-07
- Murray, I. A., Clark, T. A., Morgan, R. D., Boitano, M., Anton, B. P., Luong, K., et al. (2012). The methylomes of six bacteria. *Nucleic Acids Res.* 40, 11450–11462. doi: 10.1093/nar/gks891
- Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of bertani and weigle). *Microbiol. Mol. Biol. Rev.* 64, 412–434. doi: 10.1128/mmb.64.2.412-434.2000
- O’Connell Matherway, M., O’Driscoll, J., Fitzgerald, G. F., and Van Sinderen, D. (2009). Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb. Biotechnol.* 2, 321–332. doi: 10.1111/j.1751-7915.2008.00071.x
- Pirone-Davies, C., Hoffmann, M., Roberts, R. J., Muruvanda, T., Timme, R. E., Strain, E., et al. (2015). Genome-wide methylation patterns in *Salmonella enterica* subsp. *enterica* serovars. *PLoS ONE* 10:e0123639. doi: 10.1371/journal.pone.0123639
- Rao, D. N., Dryden, D. T. F., and Bheemanaiik, S. (2013). Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* 42, 45–55. doi: 10.1093/nar/gkt616
- Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., and Bitinaite, J. (2003a). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812. doi: 10.1093/nar/gkg274
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2003b). REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.* 31, 418–420. doi: 10.1093/nar/gkg069
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43, D298–D299. doi: 10.1093/nar/gku1046
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Sternberg, N., and Coulby, J. (1990). Cleavage of the bacteriophage P1 packaging site (pac) is regulated by adenine methylation. *Proc. Natl. Acad. Sci. U.S.A.* 87, 8070–8074. doi: 10.1073/pnas.87.20.8070
- Sternberg, N., Sauer, B., Hoess, R., and Abremski, K. (1986). Bacteriophage P1 cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. *J. Mol. Biol.* 187, 197–212. doi: 10.1016/0022-2836(86)90228-7
- Sullivan, M. J., Ben Zakour, N. L., Forde, B. M., Stanton-Cook, M., and Beatson, S. A. (2015). Contiguity: contig adjacency graph construction and visualisation. *PeerJ PrePrints* 3:e1273. doi: 10.7287/peerj.preprints.1037v1

- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Trautner, T. A., Pawlek, B., Gunthert, U., Canosi, U., Jentsch, S., and Freund, M. (1980). Restriction and modification in *Bacillus subtilis*: identification of a gene in the temperate phage SP beta coding for a BsuR specific modification methyltransferase. *Mol. Gen. Genet.* 180, 361–367. doi: 10.1007/BF00425849
- Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S., and Pop, M. (2011). Next generation sequence assembly with AMOS. *Curr. Protoc. Bioinformatics* 11:18. doi: 10.1002/0471250953.bi1108s33
- Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72. doi: 10.1128/mmb.00044-12
- Wilson, G. G., and Murray, N. E. (1991). Restriction and modification systems. *Annu. Rev. Genet.* 25, 585–627. doi: 10.1146/annurev.ge.25.120191.003101
- Zautner, A. E., Goldschmidt, A.-M., Thürmer, A., Schuldes, J., Bader, O., Lugert, R., et al. (2015). SMRT sequencing of the *Campylobacter coli* BfR-CA-9557 genome sequence reveals unique methylation motifs. *BMC Genomics* 16:1088. doi: 10.1186/s12864-015-2317-3
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl. 2), W347–W352. doi: 10.1093/nar/gkr485

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

RJR is a full-time employee of New England Biolabs, a company that sells research reagents such as DNA MTases.

Copyright © 2016 Ee, Lim, Yin, See-Too, Roberts and Chan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.