Global Veterinaria 20 (1): 06-10, 2018
ISSN 1992-6197
© IDOSI Publications, 2018
DOI: 10.5829/idosi.gv.2018.06.10

The Plasmid Praw Extends the Spectrum of Defense Genes of *Mycobacterium marinum*

[1,2]*Salma A. Shoulah,* [2]*Abdelfattah Selim,*
[2]*Faysal Hamouda,* [2]*Elsayed Galila and* [1]*Astrid Lewin*

[1]Division 16, Mycotic and parasitic Agents and Mycobacteria,
Robert Koch Institute, Berlin, Germany
[2]Department of Animal Medicine (Infectious Diseases),
Faculty of Veterinary Medicine, Benha University, Egypt

Abstract: Only a few conjugative plasmids from Mycobacteria have so far been identified and the assessment of their contribution to mycobacterial virulence is hampered by restricted knowledge of the functions of the plasmid genes. That also applies to the conjugative plasmid pRAW from *Mycobacterium marinum*. Bioinformatic tools nowadays facilitate prediction of gene functions. In this study we assigned COG (Clusters of Orthologous Groups of protein) groups to all 98 genes from pRAW including those that were annotated as "hypothetical" to contribute to functional characterization of genes from this conjugative plasmid.

Key words: Mycobacterium · Conjugative Plasmid · Gene Function Analysis · COG Categories · Defense Genes

INTRODUCTION

Many bacteria contain one or more plasmids, genetic units capable of replicating independently of the chromosome. Plasmids are vectors of horizontal gene transfer and may provide genes involved in virulence and antibiotic resistance to the recipients. They are also important engineering tools for genetic analysis of any organism. Studies of plasmids can help to understand the evolution of genes involved in virulence and to trace the spread of drug-resistant bacteria [1].

In general, mycobacterial plasmids have only been partially characterized, while the main focus has been to their development as vectors [2].

Although, nucleotide sequencing nowadays has enabled a broader understanding of mycobacterial plasmids, there are very few reports that have assigned functions to the genes on the plasmids [3].

Improved analytical tools are available to make intelligent predication about function, localization, secondary structure and other important protein properties as for example the Uniprot server (Universal Protein Resource) (http://www.uniprot.org/) [4] or the COG database (Clusters of Orthologous Groups of proteins), which are mainly based on sequence similarity (http://eggnogdb.embl.de/#/app/home) [5]. Other predication tools such as I Tasser (Iterative Threading Assembly refinement) (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) or Dali (http://ekhidna.biocenter.helsinki.fi/dali_server/start) are based on comparison of 3-D protein structures with PDB (Protein Data Bank).

The plasmid pRAW from *Mycobacterium* (*M.*) *marinum* strain E11 is a conjugative plasmid of 114.229 bp and can be conjugated to other slow growing mycobacteria including *M. tuberculosis* and *M. avium* [6, 7] with the potential of modifying virulence- and resistance-associated traits of causative agents of diseases such as tuberculosis. Assessment of this risk is, however, hampered by the fact that 60 of the 98 genes (61.22%) from pRAW are annotated as "hypothetical". Consequently, we aimed at filling this knowledge gap by applying different bioinformatics tools to predict functions for the "hypothetical" genes from pRAW.

Corresponding Author: Salma A. Shoulah, Division 16, Mycotic and parasitic Agents and Mycobacteria, Robert Koch Institute, Berlin, Germany, Department of Animal Medicine (Infectious Diseases), Faculty of Veterinary Medicine, Benha University, Egypt. E-mail: salmashoulah@gmail.com, salma.soulah@fvtm.bu.edu.eg.

MATERIALS AND METHODS

The clusters of orthologous groups (COG) category of genes were assigned by screening each gene in a COG database (http://eggnogdb.embl.de/#/app/home) [8]. In order to decipher putative gene functions, the nucleotide sequences were subjected to the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [9] and the putative function prediction using DALI server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [10] and UniProt server (http://www.uniprot.org/) [4]. Nucleotide sequence homologies were determined by using Geneious 10.0.5 software program. Coding sequences exhibiting at least 70% query coverage and at least 70% identity were considered to represent the same gene.

RESULTS

Function predication was performed by sequential application of different bioinformatics tools. First eggnog analysis (evolutionary genealogy of genes: Non-supervised Orthologous Groups) was performed, which provided COG category functions for the products of 36 out of 98 genes from the plasmid pRAW (Table 1). The remaining 62 genes were subjected to homology protein modeling using the I-TASSER server and protein putative function predication using the DALI server as well as to the UniProt server. The annotations provided by UniProt were employed for assignment of COG categories to these 62 genes from pRAW by using the COGs - Phylogenetic classification of proteins encoded in complete genomes – database from NCBI (Table 2).

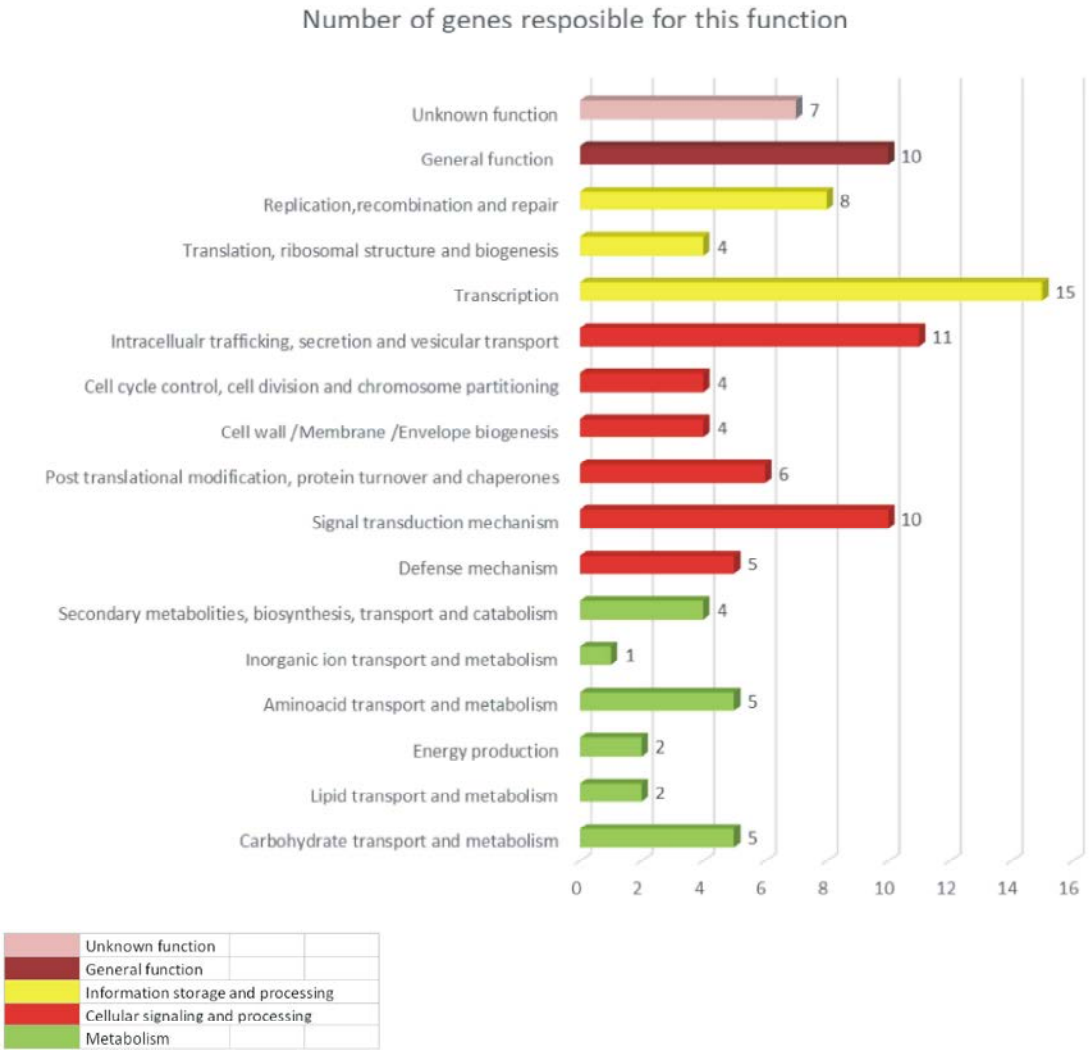

Fig. 1: COG functions of annotated genes of the plasmid pRAW

Table 1: COG categories assigned to 36 genes of the plasmid pRAW by Eggnog analysis (http://eggnogdb.embl.de/#/app/home)

Locus Tag	Annotation according to accession NZ_HG917973	COG categories
MMARE11_RS26350	hypothetical protein CDS	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS26355	NrdH-redoxin CDS	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS26360	nuclease CDS	(K) Transcription
MMARE11_RS26410	hypothetical protein CDS	(L) Replication, recombination and repair
MMARE11_RS26430	IS110 family transposase CDS	(L) Replication, recombination and repair
MMARE11_RS26435	recombinase CDS	(L) Replication, recombination and repair*
MMARE11_RS26440	hypothetical protein CDS	(J) Translation, ribosomal structure and biogenesis
MMARE11_RS26445	hypothetical protein CDS	(T) Signal transduction mechanisms
MMARE11_RS26455	recombinase CDS	(L) Replication, recombination and repair
MMARE11_RS26470	serine/threonine protein kinase CDS	(T) Signal transduction mechanisms
MMARE11_RS26475	serine/threonine protein kinase CDS	(T) Signal transduction mechanisms
MMARE11_RS26485	IS3 family transposase CDS	(L) Replication, recombination and repair
MMARE11_RS27965	resolvase CDS	(L) Replication, recombination and repair
MMARE11_RS26495	cytochrome P450 CDS	(Q) Secondary metabolites biosynthesis, transport and catabolism
MMARE11_RS26500	TetR family transcriptional regulator CDS	(K) Transcription
MMARE11_RS27970	TetR family transcriptional regulator CDS	(K) Transcription
MMARE11_RS26520	histidine kinase CDS	(T) Signal transduction mechanisms
MMARE11_RS26535	hypothetical protein CDS	(T) Signal transduction mechanisms
MMARE11_RS26540	sensor domain-containing protein CDS	(T) Signal transduction mechanisms
MMARE11_RS27975	hypothetical protein CDS	(T) Signal transduction mechanisms
MMARE11_RS26565	WhiB family transcriptional regulator CDS	(K) Transcription
MMARE11_RS26575	peptide transporter CDS	(K) Transcription
MMARE11_RS26580	ParA family protein CDS	(D) Cell cycle control, cell division and chromosome partition
MMARE11_RS26585	ParA family protein CDS	(D) Cell cycle control, cell division and chromosome partition
MMARE11_RS26600	WhiB family transcriptional regulator CDS	(K) Transcription
MMARE11_RS26605	transcriptional regulator CDS	(K) Transcription
MMARE11_RS26635	hypothetical protein CDS	(M) Cell wall/membrane/envelope biogenesis
MMARE11_RS26645	type VII secretion protein EccC CDS	(D) Cell cycle control, cell division and chromosome partition
MMARE11_RS26675	hypothetical protein CDS	(D) Cell cycle control, cell division and chromosome partition
MMARE11_RS26685	type VII secretion-associated serine protease mycosin CDS	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS26695	type VII secretion AAA-ATPase EccA CDS	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS26705	hypothetical protein CDS	(T) Signal transduction mechanisms
MMARE11_RS26710	hypothetical protein CDS	(T) Signal transduction mechanisms
MMARE11_RS26725	hypothetical protein CDS	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26765	type VI secretion protein CDS	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26785	prolyl aminopeptidase CDS	(Q) Secondary metabolites biosynthesis, transport and catabolism

Fig. 1 shows the representation of the genes from pRAW in the COG categories. Out of the 23 COGS groups 17 were represented on pRAW. The functional cluster "Cellular signaling and processing" was most frequently represented (45 genes) followed by "Information storage and processing" (27 genes) and "Metabolism" (19 genes). Only seven genes remained in the COG group "Function unknown". Five genes were assigned to the COG group "Defense". These were annotated by Uniprot Blast as Methyl-accepting chemotaxis protein, ESAT-6-like protein, EsxI, ESX secretion-associated protein EspG and Multidrug efflux system subunit C.

With exception of one gene (WXG100, Locus tag MMARE11_RS26660) none of the genes from pRAW were present elsewhere on the chromosome of *M. marinum* E11 meaning that pRAW extends the spectrum of COG group "Defense" genes of *M. marinum* by at least five genes.

Table 2: COG categories assigned to the 62 genes of pRAW that could not be categorized by EGGNOG analysis by comparing annotations obtained with Uniprot BLAST to the NCBI COG lists (https://ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/listCOGs.html

Locus tag	Annotation according to accession NZ_HG917973	Annotation according to Uniprot Blast	Assigned COG categories
MMARE11_RS26365	hypothetical protein CDS	Acyl-CoA dehydrogenase	(I) Lipid transport and metabolism
MMARE11_RS26370	hypothetical protein CDS	Antitoxin Phd	(K) Transcription
MMARE11_RS26375	death-on-curing protein CDS	Phage killer protein	(R) General function predication only
MMARE11_RS26380	hypothetical protein CDS	Phage excisionase	(R) General function predication only
MMARE11_RS26385	hypothetical protein CDS	Biofilm regulator BssS	(S) Function unknown
MMARE11_RS26390	hypothetical protein CDS	Nucleoid-associated protein ME5_00606	(L) Replication, recombination and repair
MMARE11_RS26395	hypothetical protein CDS	ESX-1 secretion-associated protein	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26400	hypothetical protein CDS	PE family protein	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26405	hypothetical protein CDS	Type I secretion protein TolC	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26415	hypothetical protein CDS	Outer membrane protein, nutrient binding	(M) Cell wall/ membrane/ envelope biogenesis
MMARE11_RS26420	hypothetical protein CDS	Chemotaxis protein	(T) Signal transduction mechanism
MMARE11_RS26425	hypothetical protein CDS	UPF0056 membrane protein	(M) Cell wall/ membrane/ envelope biogenesis
MMARE11_RS27960	hypothetical protein CDS	Polyprenol-phosphate-mannose-dependent alpha-(1-2)-phosphatidylinositol	(I) Lipid transport and metabolism
MMARE11_RS26450	hypothetical protein CDS	TetR family transcriptional regulator	(K) Transcription
MMARE11_RS26460	MazF family transcriptional regulator CDS	Transcriptional modulator of MazE/toxin MazF	(K) Transcription
MMARE11_RS26465	hypothetical protein CDS	Predicted membrane protein involved in D-alanine export	(S) Function unknown
MMARE11_RS	hypothetical protein CDS RS26490	Transmembrane protein	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS26510	hypothetical protein CDS	Glycosyl transferase 2 family protein	(C) Energy production and conversion
MMARE11_RS26515	hypothetical protein CDS	Riboflavin biosynthesis protein RibD	(P) Inorganic ions transport and metabolism
MMARE11_RS26525	hypothetical protein CDS	Uncharacterized protein	(S) Function unknown
MMARE11_RS26530	hypothetical protein CDS	TetR family transcriptional regulator	(K) Transcription
MMARE11_RS26555	hypothetical protein CDS	Possible conserved transmembrane protein	(S) Function unknown
MMARE11_RS26560	hypothetical protein CDS	Conjugal transfer protein TrbG/VirB9/CagX	(L) Replication, recombination and repair
MMARE11_RS26570	hypothetical protein CDS	Isoleucine--tRNA ligase	(O) Post translational modification, protein turnover and chaperones
MMARE11_RS27980	hypothetical protein CDS	Translation initiation factor IF-2	(J) Translation, including ribosome structure and biogenesis
MMARE11_RS26590	hypothetical protein CDS	MarR family transcriptional regulator	(K) Transcription
MMARE11_RS26595	hypothetical protein CDS	Transcription factor WhiB family protein	(K) Transcription
MMARE11_RS26610	hypothetical protein CDS	6-phosphogluconate dehydrogenase	(G) Carbohydrate transport and metabolism
MMARE11_RS26615	hypothetical protein CDS	Oxidoreductase	(R) General function predication only
MMARE11_RS26620	hypothetical protein CDS	Methyl-accepting chemotaxis protein	(V) Defense mechanism
MMARE11_RS26625	hypothetical protein CDS	MFS	(E, G) Amino acid transport and metabolism Carbohydrate transport and metabolism
MMARE11_RS26630	hypothetical protein CDS	Glycoside hydrolase	(G) Carbohydrate transport and metabolism
MMARE11_RS26640	type VII secretion protein EccB CDS	MPSS family PPE41 protein	(R) General function predication only
MMARE11_RS26650	PE family protein CDS	PE family protein	(U) Intacellular trafficking , secretion and vesicular transport
MMARE11_RS26655	PPE family protein CDS	PEE family protein	(U) Intacellular trafficking , secretion and vesicular transport
MMARE11_RS26660	WXG100 family type VII secretion target CDS	ESAT-6-like protein	(V) Defense mechanism
MMARE11_RS26665	type VII secretion protein EsxI CDS	EsxI	(V) Defense mechanism
MMARE11_RS26670	ESX secretion-associated protein EspG CDS	ESX secretion-associated protein EspG	(V) Defense mechanism
MMARE11_RS26680	type VII secretion integral membrane protein EccD CDS	Type VII secretion integral membrane protein EccD	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26690	type VII secretion protein EccE CDS	Type VII secretion protein EccE	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26700	hypothetical protein CDS	Cation/calcium exchanger 4	(S) Function unknown
MMARE11_RS26715	hypothetical protein CDS	Histidinol phosphatase	(E) Amino acid transport and metabolism
MMARE11_RS26720	hypothetical protein CDS	Transmembrane transport	(S) Function unknown
MMARE11_RS26730	hypothetical protein CDS	Histidinol phosphatase	(E) Amino acid transport and metabolism
MMARE11_RS26735	hypothetical protein CDS	MFS transporter	(E, G) Amino acid transport and metabolism Carbohydrate transport and metabolism
MMARE11_RS26740	hypothetical protein CDS	Conjugative transposon TcpC family protein	(R) General function prediction only
MMARE11_RS26745	hypothetical protein CDS	F5/8 type C domain-containing protein	(M) Cell wall/ membrane/ envelope biogenesis
MMARE11_RS26750	hypothetical protein CDS	Acetyltransferase, gnat family	(R) General function prediction only
MMARE11_RS26755	hypothetical protein CDS	Transcriptional regulator	(K) Transcription
MMARE11_RS26760	hypothetical protein CDS	Type IV secretory pathway, VirD4 component	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26770	hypothetical protein CDS	Fur family transcriptional regulator	(K) Transcription
MMARE11_RS26775	hypothetical protein CDS	Type VI secretion protein	(U) Intracellular trafficking, secretion and vesicular transport
MMARE11_RS26780	hypothetical protein CDS	Multidrug efflux system, subunit C	(V) Defense mechanism
MMARE11_RS26790	exonuclease V subunit alpha CDS	Polyprenol-phosphate-mannose-dependent alpha-(1-2)-phosphatidylinositol pentamannoside mannosyltransferase	(C, G) Energy production and conversion Carbohydrate transport and metabolism

Table 2: Continued

MMARE11_RS26795	hypothetical protein CDS	Putative large Ala/Glu-rich protein	(S) Function unknown
MMARE11_RS26800	hypothetical protein CDS	Putative Bacteriophage T7-related protein	(R) General function predication only
MMARE11_RS26805	DUF945 domain-containing protein CDS	Phage plasmid related protein	(R) General function predication only
MMARE11_RS26810	hypothetical protein CDS	X-Pro aminopeptidase	(E) Amino acid transport and metabolism
MMARE11_RS26815	hypothetical protein CDS	Ribosomal protein S5 domain 2-type fold, subgroup	(J) Translation, including ribosome structure and biogenesis
MMARE11_RS26820	IS1380 family transposase CDS	Putative transposase	(R) General function predication only
MMARE11_RS26825	hypothetical protein CDS	Ribosomal protein S5 domain 2-type fold, subgroup	(J) Translation, including ribosome structure and biogenesis
MMARE11_RS26830	hypothetical protein CDS	ATPase component of ABC transporters with duplicated ATPase domain-containing protein	(R) General function predication only

DISSCUSION

There are limit knowledge about function of the plasmid genes in spite of some plasmid genes play an important role in the gene transfer [3] especially for the mycobacterial plasmid. So, in our result we showed that by using different bioinformatic tools we can reach to the function of these plasmid genes as which done in our work.

CONCLUSIONS

Our analysis assigned functional COG groups to 91 of the 98 genes from the *M. marinum* plasmid pRAW for the first time, which will facilitate further analysis of the role of this plasmid for phenotypic features of this pathogen such as virulence or resistance.

ACKNOWLEDGEMENT

We thank Andrea Sanchini for his help with gene function analysis, Salma Shoulah was supported by Ministry of Higher Education and Scientific Research in Egypt.

REFERENCES

1. Antipov, D., N. Hartwick, M. Shen, M. Raiko, A. Lapidus and P.A. Pevzner, 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics, 32: 3380-3387.
2. Movahedzadeh, F. and W. Bitter, 2009. Ins and Outs of Mycobacterial Plasmids. Methods in Molecular Biology (Clifton, N.J.), pp: 217-228.
3. Stinear, T.P., M.J. Pryor, J.L. Porter and S.T. Cole, 2005. Functional analysis and annotation of the virulence plasmid pMUM001 from Mycobacterium ulcerans. Microbiology, 151: 683-692.
4. Wu, C.H., R. Apweiler, A. Bairoch, D. A. Natale, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, R. Mazumder, C. O'Donovan, N. Redaschi and B. Suzek, 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res., 34: D187-D191.
5. Tatusov, R.L., D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova and E.V. Koonin, 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res., 29: 22-28.
6. Ummels, R., A.M. Abdallah, V. Kuiper, A. Aâjoud, M. Sparrius, R. Naeem, H.P. Spaink, D. Van Soolingen, A. Pain and W. Bitter, 2014. Identification of a novel conjugative plasmid in mycobacteria that requires both type IV and type VII secretion. MBio 5, e01744-e01714.
7. Shoulah, S.A., A.M. Oschmann, A. Selim, T. Semmler, C. Schwarz, E. Kamal, F. Hamouda, E. Galila, W. Bitter and A. Lewin, 2018. Environmental Mycobacterium avium subsp. hominissuis have a higher probability to act as a recipient in conjugation than clinical strains. Plasmid, 95: 28-35.
8. Huerta-Cepas, J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M.C. Walter, T. Rattei, D.R. Mende, S. Sunagawa, M. Kuhn, L.J. Jensen, C. Von Mering and P. Bork, 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res., 44: D286-D293.
9. Roy, A., A. Kucukural and Y. Zhang, 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc., 5: 725-738.
10. Holm, L. and P. Rosenström, 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res., 38: W545-W549.