

Supplementary Table and Figures

“The odd “RB” phage – Identification of arabinosylation as a new epigenetic modification of DNA”

Julie A. Thomas, Jared Orwenyo, Lai-Xi Wang, and Lindsay W. Black

Supplementary Table S1. Genome similarity between T4 phage, “RB” phages and other phages with identified glucosyltransferases (a-gt, b-gt or ba-gt proteins listed in Table 1) as determined by Blastn. NIS, indicates no identifiable similarity by Blastn.

Phage	Query cover, %	E value	Identity, %	Accession
T4	100	0.0	100	AF158101.6
T2 ¹	87	0.0	98	
T6 ¹	89	0.0	97	
Enterobacteria phage RB3	87	0.0	97	KM606994.1
Enterobacteria phage RB5	87	0.0	97	KM606995.1
Enterobacteria phage RB6	87	0.0	97	KM606996.1
Enterobacteria phage RB7	87	0.0	97	KM606997.1
Enterobacteria phage RB9	87	0.0	97	KM606998.1
Enterobacteria phage RB10	87	0.0	97	KM606999.1
Enterobacteria phage RB14	86	0.0	95	FJ839692.1
Enterobacteria phage RB27	86	0.0	96	KM607000.1
Bacteriophage RB32	86	0.0	95	DQ904452.1
Enterobacteria phage RB33	86	0.0	95	KM607001.1
Enterobacteria phage RB51	85	0.0	95	FJ839693.1
Enterobacteria phage RB55	99	0.0	99	KM607002.1
Enterobacteria phage RB59	99	0.0	99	KM607003.1
Enterobacteria phage RB68	85	0.0	95	KM607004.1
Enterobacteria phage RB69	77	0.0	70	NC_004928
<i>Other T4-related phages with glucosyltransferase homologs</i>				
<i>Yersinia</i> phage PST	88	0.0	97	KF208315.1
<i>Shigella</i> phage Shfl2	87	0.0	96	HM035025.1
<i>Escherichia</i> phage wV7	86	0.0	95	HM997020.1
Enterobacteria phage AR1	87	0.0	95	AP011113.1
<i>Escherichia</i> phage HY01	87	0.0	95	KF925357.1
<i>Escherichia</i> phage e11/2	89	0.0	97	KJ668714.2
Enterobacteria phage ACG-C40	87	0.0	96	JN986846.1
<i>Other phages with diverged glucosyltransferase homologs</i>				
Enterobacteria phage CC31	52	0.0	72	NC_014662.1
Enterobacter phage PG7	54	0.0	72	NC_023561.1
<i>Salmonella</i> phage S16	56	0.0	72	NC_020416.1
<i>Salmonella</i> phage STML-198	56	0.0	72	NC_027344.1
<i>Citrobacter</i> phage Moon	54	0.0	72	NC_027331.1
<i>Serratia</i> phage PS2	25	0.0	68	NC_024121.1
<i>Salmonella</i> phage STP4-a	57	0.0	72	NC_026607.2
<i>Citrobacter</i> phage Merlin	55	0.0	73	NC_028857.1
Bacillus virus G	2	5e-07	69	NC_023719.1

¹ T2 and T6 nucleotide sequences were obtained from Dr. James Nolan and Dr. Jim Karam

Supplementary Table 2. Homologs to RB69 ORF003C (NP_861693.1) in other phages identified using Psi-Blast.

Phage	Protein name	Protein accession	Query coverage, %	Identity, %	E value
<i>Escherichia</i> phage vB_EcoM_JS09	hypothetical protein JS09_0177	YP_009037500.1	99	100	0.00
<i>Shigella</i> phage SHSML-52-1	hypothetical protein	YP_009289016.1	99	100	0.00
<i>Escherichia</i> phage APCEc01	hypothetical protein APCEc01_026	YP_009224986.1	99	100	0.00
<i>Escherichia</i> phage phiE142	hypothetical protein phiE142_23	ALY07828.1	99	100	0.00
<i>Escherichia</i> phage vB_EcoM_PhAPEC2	hypothetical protein PhAPEC2_3	YP_009056595.1	99	100	0.00
<i>Shigella</i> phage Shf125875	hypothetical protein	YP_009100545.1	99	99	0.00
<i>Escherichia</i> phage ST0	hypothetical protein	ASD53821.1	98	100	0.00
Acinetobacter phage Acj61	hypothetical protein Acj61p077	YP_004009694.1	41	99	9.00E-99
Acinetobacter phage Acj61	hypothetical protein Acj61p076	YP_004009693.1	42	98	3.00E-94
Acinetobacter phage Acj9	hypothetical protein Acj9p080	YP_004010217.1	41	98	8.00E-85
Acinetobacter phage Acj9	hypothetical protein Acj9p081	YP_004010218.1	52	35	6.00E-36
Acinetobacter phage Acj9	hypothetical protein Acj9p082	YP_004010219.1	35	50	4.00E-29

Supplementary Figure S1. Alignment of T4 a-gt (gp59) with homologs in related T4-like phages and phage G (gp306 and gp313). The N-terminal domain of T4 a-gt includes residues 1-177 and the C-terminal domain residues 176-386 (shaded blue). T4 a-gt residues with important interactions by Lariviere et al., 2005 (JMB 352:139-150) (Table 1) are highlighted. Structural elements from 1xv5_A (T4 a-gt) were added manually to the alignment (S, beta sheet; H, helix).

	S1	H1	S2	S3	H2	S4	H3
dssp	CEEEEEETTCCSSHHHHHHHHHHHHHHHTTCEEEEEECSSCCTTTTSSSCTTTTCEEECTTTCHHHHHHHHTSCSEEEEEECBTTSCHHHHHHHHHH						
T4	MRICIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRTSSHDHKSFSIPVILAKEYDKALKLVNDCDILIINSVPATSVQEATINNYKKL						
RB14	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
RB32	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
AR1	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
RB51	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
AR1	-----CGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
T2	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRNAHDYKSFSIPVLLAKEYDKTLKLVNDCDILIINSVPATSVVEEDTINNYKKI						
T6	MKVCIFMARGLEGCGVTKFSLEQRDWFIKNGHEVTLVYAKDKSFTRTSSHDHKSFSIPVILAKEYDKALKLVNDCDILIINSVPATSVQEATINNYKKL						
G306	-KIALIMGQSIIEGCGVTRNAEMFQWSKKAGVKFDIYSYDERMYNRRDAHMDFI SFT---RENINSTVEKLNQYDIVMFNSYPSNKFQAIIDFYEK						
G313	-KLALIFGLSIEGAGATRNGSEMQHWCCKNGVQFKIFS YDESKFTREESHK---ISYTKFNKYNLKEVVDELNTYDIVMFNTYFPFKVGQEAYENFYHN						
	T4 Gly-15 interacts beta phosphate			T4 Arg-46 interacts with base			
	S5	H4	S6	H5	S7	H6	S8
dssp	HHHSCTTSEEEEEECSSHHHHHTTBSSHHHHHHHCS	EEEE	SCTTSHHHHTHHHHHSCSSCCSSSSCCCCCCCCCEEECCCCBCHHHHHHHHCCCGGGCEEEE				
T4	LDNIKPSIRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIVKVRSTYWKDVSEINMNI						
RB14	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
AR1	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
RB32	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
RB51	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
AR1	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
T2	IDNIKPSVRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
T6	LDNIKPSIRVVVYQHDHSSLSLRNLGLEETVRRADVIFSHSDNGDFNKVLMKEWYPETVSLFDDIEEAPT VYNFQPPMDIAKVRSTYWKDVSEINMNI						
G306	VIKGVTTIK-VGFMHELNKTNIDKIPYLVGIMNEMDVIYNFGEETWFSQTI-----SDLLPSKEIGKRTKKFTMWFNFEDLENNYRNKYSLDDKSK						
G313	FVKKITALK-VGFMHEINKVIIDKIPYILGLMNEMDMIYTFGVDTWFSK-----NTSTLLPSKKPNERIKKFTMWFNFELLEEYRKTIQLDDKEKK						
	T4 His114, His116 and His140 interact with glucose						

Supplementary Fig. S1 cont.

	S8	H7	S9	S10	S11	H8	S12
dssp	EE.EECCSCGGGGCHHHHHHHHHHT	TTTTTCEEEEECCCCSHHHHHHHHTT	CCCCCEGGGGGGCCCCSSSC	EEEEESCCCHHHHHHHHTTE	EEEEEECCC		
T4	NR.WIGRTTTWKGFYQMFDHEKF	LKPAGKSTVMEGLERSPAFIAI	KEKGIPIEYYGNREIDKMN	LAPNQPAQILD	CYINSEMLERMSKSGFGYQLSK		
RB14	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
AR1	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
RB32	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
RB51	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
AR1	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
T2	NR.WIGRTTTWKGFYQMFDHEKH	LKPAGLSTIMEGLERSPAFI	PIKEKGIPIEYYRLHQVDQ	IKIAPNLPTQILDR	YVNSEMLERMSKSGFGYQLSK		
T6	NR.WIGRTTTWKGFYQMFDHEKF	LKPAGKSTVMEGLERSPAFIAI	KEKGIPIEYYGNREIDKMN	LAPNQPAQILD	CYINSEMLERMSKSGFGYQLSK		
G306	KLvYCSRWTTMKGPRRVLDLAPMLa	VKDPEFKAELKGIERS---	IGAKFDIFDHPNTLDCTGRTP	DPNSTGTVPVYGP	YIREEGIDFMAKNLFIA	SYR	
G313	ML.YLGRFVSTKDVPRLNLGAEI	LKK--DSNFKTIIRGIDTSIGAKV	SVLDHPNAINYIPKEPKVNS	NGCVPVYGP	YVRDEGMNEMAHS	SLFGVSFWN	
	T4 Arg204 and Lys209 interact with beta phosphate					T4 Tyr275 and Asn277, interact with ribose Tyr275 also interacts with base	

	H9	S13	S14	S15	S16	H10	H11	H12
dssp	CCGGGCSS...CCCHHHHHHHHT	SEEEEEHHHHHHSBCTTT	CCBGGGSCCSEEECTT	CHHHHHHHHHHHHTCH	HHHHHHHHHHHHHHHHHH	HHHHHHHHHHHHHHHHHH	BHHHHH	
T4	LNQKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
RB14	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
AR1	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
RB32	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
RB51	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
AR1	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSAR-----				
T2	LDKKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
T6	LNQKYLQR...SLEYTHLELGACGT	IPVFWKSTGENLKFRVDNTPLT	SHDSGIIWFDENDMESTFERIKEL	SSDRALYDREREKAYEFLYQH	QDSSFCF			
G306	MPKAIQDYgd.RMEYSQIESIAVGS	IPVFDTNWGENNR-TLDG	QRYIDVPYSAIYSDEKDLEGT	VEKLI	EVANNKELQEKYRETSYKIAKQ	EFDANIVL		
G313	MQKRGIEEygdrMEYTIETI	AVGTIPVFDKHWGENNR-TLDG	RRYIDIPYSAIYCDREN	LN	NETVDKLI	EVANNKELQEKYRNTSYN	LAKQEF	DA
	T4 Tyr307 Glu-311-interacts with ribose and conserved in all GT-B glycosyltransferases and Thr308-both interact with alpha phosphate							

	H12
dssp	HHHHHHHTC
T4	KEQFDIITK
RB14	KEQFDIITK
AR1	KEQFDIITK
RB32	KEQFDIITK
RB51	KEQFDIITK
AR1	-----
T2	KEQFDIITK
T6	KEQFDIITK
G306	PKMFNEI--
G313	PKMFNE---

Supplementary Figure S2. HHpred alignment of phage G gp306 (AE093565.1) and RB69 ORF003c (32350306) HHMs. Residues in phage G gp306 that are conserved with residues of known functional importance in T4 a-gt (see Supplementary Figure 1) are shaded yellow and residues which align with them in RB69 ORF003c (or an appropriate residue close by) are also shaded yellow.

>gi|32350306|gb|AAP75905.1| RB69ORF003c hypothetical protein [Enterobacteria phage RB69]
 Probab=93.22 E-value=2.9e-07 Score=55.85 Aligned cols=343 Identities=10% Similarity=0.043 Sum probs=180.4

Q AE093565.1	21	-----GVTRNAEMFQWSKAGVKFDIYSYDERMYNRDAHEMDFISF-----TREN	67	(469)
Q Consensus	2	MkI~~~t~~~p-----GG~~~~v~~l~~l~~GheV~vi~~~~~ +++-.+~ . . -+.+. .+. ++-++..+...+....+ +..+....+ -.....	74	(452)
T Consensus	1	MkIlyiPcR~VpFn~dri~GGLEaVqlN~ik~l~S~~~~IdYv~FgdenFG~kvn----id~pi~~KfT~~h~~ki~~	76	(366)
T gi 32350306 gb	1	MKVMFIPSRAPVPFPNPDRVQGGLLEAVHLNLVKYLVSIGADIDYIGFDNDTFGDWKVNH---HPVGHLTKFSLGMSYTMAR	76	(366)
Confidence		999999766666 359999999999999999999999999988876665541 1112211111 1112334		
		T4 Gly-15 interacts beta phosphate		
			T4 Arg-46 interacts with base	

Q AE093565.1	68	INSTVEKLNQYDIVMFNSYPSNKFQAIIDFYEkvIKGVTTIKVGFMHELNKTNIkdIPYLVGIMNEMDVIYNFGEETW	147	(469)
Q Consensus	75	~~~~~DiIh~~~~~..ip~V~t~h~~~~~..~~~~~d~ii~ +.+.+.-.+ . .-.+.-+.+-+-. . .+-+.+..... ...+.-.-.+.-+.	150	(452)
T Consensus	77	ki~el~~~~YDavVt~Eg~K~~~n~l~~G1--..~kvrN~mATPld--P~vRg..iVq-i-W~ai~iHknGGKnlv	146	(366)
T gi 32350306 gb	77	KIVELAGIHEYDFVVTMEPTKLTVQAIKDAGL--..SKVHKNFMATPF--PVSRG..IVQ-I-WDQTIQIHKNNGGKSya	146	(366)
Confidence		4444556679999988777665443222111 1111111111000 00000 000 1 1112334556666778		
		T4 His114, His116 and His140 interact with glucose		

Q AE093565.1	148	FSQTISDLLPSKEI--GKRTKKFTMWFNFE	DLSE-----NNYRNKYS	LD-----DKSKKL	LVYCSRW-TTMKGPPRRVLDL	212	(469)
Q Consensus	151	~S~~~~~l~~~~~ki~vIpnGvD~~F~~~~~	~~~~~r~~~~~	~~~~~iIl~vGRl~	~~~KGid~LI~A	221	(452)
		+++..++.-.+.+.--+..++	-- +++.+.-.-	+++...-+.	+..-. +~+~-+..++		
T Consensus	147	PT~tFkdl~~vyskmn~~~~~	~vIDfDYWa~ndiis~f~p~li~kp~vl~s~g~ivg	AQRyD~kfRktnV	Alea	222	(366)
T gi 32350306 gb	147	PTKAFREFERKYCYMTSGLT----	DKIDYDYWRANPLFEAEDYPVICLNEKPEVLPATDLII	SAQRDYDTKMR	RTDVALEA	222	(366)
Confidence		88888887777776233344	5567766653222	122222222	233344444453 222334566777		
					T4 Arg204 and Lys209		
					interact with beta phosphate		

Supplementary Figure S3. Fasta sequence of RB69 ORF53_52c

```
>RB69_ORF53_52c
MKKAVILGAGLATRLYPITHHIPKVLVNYKQDTILSNLYTIYSDLGADEIIVVVHSKFAE
TVRAYCEQEGFNVTIRTVD EAYGSAYALAKLYKDLDGHNVIIVNWCDIIPDFGSWSWNVNA
IYVKGDECRYNFDGENITNVGSTGGNVVGIYQFKDWEFYMGSTDEEIHEYCKGRDFVEFL
YGSFANKSELMNLIDLGDMPKLEKAHEVRELNRSFNAVEIGEETVTKFALTEHGRALQKD
EISWYLKVKSDSVPQLVSVNENNFEMERIKGKPAFEYIKSKSSLARPQIVDAILDALKFS
TDTYFVSPETVRRDFTKEFYTKVIDRCESIQPLIDSF GKITHVNYTKIGRLKPMLKQALE
HLIRYHNRSQGGQYSVIHGDPNFSNTMITDNGEVKFIDPRGYFGETKIYGPKLYDEAKVLY
AVSGYDEFNANPTWGQFTIDETTCNVTININPLVYKYGKMSSFNEYHHLAVAIIWIALGG
YFKNNPLKAVAAYYKGMELLTKQLRNMGRVLQDGSISYDVAEPVTATLITKNPGKWVLT
KETGVSYRPIGGDITHQWERI
```

13. [1XV5_A](#) DNA alpha-glucosyltransferase (E.C.2.4.1.26); Transferase; HET: CME, EDO, GOL, UDP; 1.73A
 {Enterobacteria phage T4}; Related PDB entries: 1Y6F_A 1Y6F_B 1Y6G_A 1Y6G_B 1YA6_A 1YA6_B 1Y8Z_A 1Y8Z_B
 Probability: 99.9 E-value: 1.9E-24 Score: 193.53 Aligned Cols: 323 Identities: 9% Similarity:
 0.016

8


```

Q ss_pred          C-CCEEEeCCCCchcccCC-----CeEEEEcCCCCCcHHHHHHHHHHHHHHhCCHHHHH--HHH
Q NP_861693.1 286 G-VPVIQLVFKDYPHATFEYDP-----DTVRVEIDSSTPKKEVVALYIKAVLEFTDTYEARV---KRA 344 (366)
Q Consensus      286 G-~pvI~-----~g~-----i~l~-----~ 344 (366)
                   | +|||+++..++..+++..+++      +|+++ ++.|. .+++++|.+++++++.+.   +++
T Consensus      317 G~~pvi~-----e~-----g~-----~l~~i~l~-----~ 378 (401)
T 1XV5_A         317 GTIPVFWKSTGENL-KFRVDNTPLTSHDSGIWF--DEND----MESTFERIKELSSDRALYDREREKA 378 (401)
T ss_dssp        TEEEEEEHHHHHHS-BCTTTCCBGGGSCCSEEE--CTTC----HHHHHHHHHHHHTCHHHHHHHHHHH
T ss_pred        CCCCEEEEeCCccch-hhcccCCCCCcCCCCcEEE--cCCc----hhHHHHHHHHHhCCHHHHHHHHHH

Q ss_pred          HH-HHHHcCHHHHHHHHHHHHhC
Q NP_861693.1 345 EA-VWKKYNRDAVVAMWDKIFTA 366 (366)
Q Consensus      345 ~~~~~~s~~~~~ 366 (366)
                   ++ ..+.|+|+.+++++.+++++
T Consensus      379 ~~~~~~ 401 (401)
T 1XV5_A         379 YEFLYQHQDSSFCFKEQFDIITK 401 (401)
T ss_dssp        HHHHHHHHBHHHHHHHHHHHTC
T ss_pred        HHHHHHcCCHHHHHHHHHHhhC

```


Q ss_pred		HcCCCCccccHHHHHHHHHHhcccc--ccceeeccc hhhcCCCCCcccCCccccCCCCCCC--CCeEEE	
Q NP_861693.1	139	KNGGKSYAPT KAFRE FERKYCYMTSG-- LT DKIDYDYWRAN PLFE AEDYP VIC LN EKPE VL P --AT DLII	204 (366)
Q Consensus	139	~~~~~s~~~~~-----~~~~~i~	204 (366)
		...+.++++. +++++...+. +...++	
T Consensus	132	~~~d~vi~~s~~~~~ip~~~~~-----~~~~~i~	188 (413)
T 30Y2_A	132	PKVVGVMAMSKCWISD IC NYGCKVPINIV SHF VD T K TI Y-----DARK LVGL SEY ND D VL FL	188 (413)
T ss_dssp		TT EEEEEE SS THHHHHHH TTCC SEEE CCCCCCCCCT----- THHHH TCGGGT TSEEEE	
T ss_pred		cCCC EEEE c CHHHHHHH hCCCC EE E CC CC hHh ----- hHHHH hCC C CC EEEE	
Q ss_pred		EeeCCChhhcCHHHHHHHHHHccC---c EEEE CCCCCCCCc-----CC--- EE E C ---	
Q NP_861693.1	205	SA QRY DT KMR RT DVA LE AIK AL GEN ---GAGY CP SK WAP PA K -----YP--- VI IDA--	250 (366)
Q Consensus	205	~~g~~~~~k~~~~li~a~~l~~~~~l~i~G~~~~~-----v~~g--	250 (366)
		++ +++. . +++.+++++.+. . +++++ . + +... .. +++	
T Consensus	189	~~g~~~~~k~~~~li~~~~l~~~~~l~i~G~~~~~-----l~~~~~v~~g~v	255 (413)
T 30Y2_A	189	NMNR TA- RKRL D IY VLAA RF ISKYP DAK V RL CNS-- HHES K FD L HSI AL REL VAS GVD N V F TH L NKI	255 (413)
T ss_dssp		CCSCSSG-GGT HHHHHHHHHHHHHH CTTCC EEEEEC --CTTCC HHHHHHHHHHHH TCSC HHHHH TT E	
T ss_pred		E cCCC h - h c CHHHHHHHHHHHHH hCCC EE EE ee CC--CCcccc HHHHHHHHHH CC C ce EE ec hh	
Q ss_pred		-----ch--HHHHHHHHhCC EEEE cCCCCCCCC HHHHHH hCC CEEE ecCCCC hhh cccCC C EE EE Ec	
Q NP_861693.1	251	-----PH-- SEI ME RLK TAKAL INT CPDTGT VEN SS IEA ISK GV P VI QL V FK DY PHAT F E Y DP DT V R VE I	313 (366)
Q Consensus	251	-----~~~~~adi~i~ps~~~e~~~~~Ea~a~G~pvI~~~~~g~~~~~	313 (366)
		.+ ++++.++++. ++++ + .. ++ . +++ ++ + +++.. +++ .+. + +++	
T Consensus	256	~~~~~~~~~~adi~l~~s~~~e~~~~~Ea~~~G~pvi~~~~~-----g~~~~~	320 (413)
T 30Y2_A	256	MIN RT VL T DER V DM YN ACD V IV N C S-SG E GF GL C SA E GAV L GK PL I I SA V G G AD - DY F S - G D CV Y KI --	320 (413)
T ss_dssp		EE EC SC CC HHHHHHHH HC EEEE CC-SCC SS CH HHHHHH TTTCC EEEE CC HHHH - HH SC-TTT SE EE--	
T ss_pred		ccccCCCC HHHHHHHH hCC EEEE cC-CCCC HHHHHH hCC CEEE ecCCC h - hh cc-CC ce EE ee --	

```

Q ss_pred          CC-----CCCcHHHHHHHHHHHHHHhCCHHHHH---HHHHH-HHHHcCHHHHHHHHH
Q NP_861693.1 314 DS-----STPKKEVVALYIKAVLEFTDTYEARV--KRAEA-VWKKYNRDAVVAMWD 361 (366)
Q Consensus 314 ~~~~~~i~l~~~~~s~~~~~
      ++          .|      .+++++|. ++.+++.+.   +++++ .+.|+|+.+++++.
T Consensus 321 ~~~~~~l~i~~~~~~s~~~~~
T 3OY2_A 321 KPSAWISVDDRDIIGGIEGIID----VDDLVEAFT-FFKDEKNRKEYGKRVQDFVKTPTWDDISSDII 384 (413)
T ss_dssp          CCCEEEECTTTCSSCCEEEECC----HHHHHHHHH-HTTSHHHHHHHHHHHHHHTTSCCHHHHHHHHH
T ss_pred          CccceccCCCCCCCCcccccC----HHHHHHHHH-HhcCHHHHHHHHHHHHHHHhCCCHHHHHHHHH

Q ss_pred          HHHhC
Q NP_861693.1 362 KIFTA 366 (366)
Q Consensus 362 ~~~~~ 366 (366)
      +++.+
T Consensus 385 ~~~~~ 389 (413)
T 3OY2_A 385 DFFNS 389 (413)
T ss_dssp          HHHHH
T ss_pred          HHHHH

```



```

Q ss_pred          HHHHHHH---hCCEEEEcCCCCCCCCHHHHHHHHcCCCEEEcCCCCchhhcccCCCeEEEEcCCCCCc
Q NP_861693.1 254  EIMERLK---TAKALINTCPDTGTVENSSIEAISKGVPIQLVFKDYPHATFEYDPDTRVVEIDSSTPK 319 (366)
Q Consensus 254  ~~~~~--~adi~i~ps~~~e~~~~~Ea~a~G~pvI~~~~~g~~~~~
      ++.++++ .|++++|+ ..|++|.+++|||++|+|||+++.. +++..++|+++ ++.|
T Consensus 347  ~~~~~~di~i~ps-~~e~~~~~Ea~a~G~PvI~~~~~-e~i~~~~g~~~~~ 410 (499)
T 2R60_A 347  ELAGCYAYLASKGSVFALTS-FYEPFGLAPVEAMASGLPAVVTRNGGPA-EILDGGKYGVIV--DPED-- 410 (499)
T ss_dssp          HHHHHHHHHHHTTCEEEECC-SCBCCCSHHHHHHHHTTCEEEESSBHHH-HHTGGGTSSEEE--CTTC--
T ss_pred          HHHHHHHHHHhCcCEEEEcC-CCCCChHHHHHHHcCCCEEEcCCChH-HhccCCCCEEEe--CCCC--

Q ss_pred          hHHHHHHHHHHHHHHhCCHHHHH---HHHHH-HHHHcCHHHHHHHHHHHHHhC
Q NP_861693.1 320  KEVVALYIKAVLEFTDTYEARV---KRAEA-VWKKYNRDVAVVAMWDKIFTA 366 (366)
Q Consensus 320  ~~~~~~i~l~~~~~s~~~~~
      .+++++|.+++++.+. +++++ .+.|+|.++++.++++
T Consensus 411  ---~l~i~l~~~~~s~~~~~ 458 (499)
T 2R60_A 411  ---PEDIARGLLKAFESEETWSAYQEKQKQVVEERYTWQETARGYLEVIQE 458 (499)
T ss_dssp          ---HHHHHHHHHHHSCHHHHHHHHHHHHHHHHHHHSHHHHHHHHHHHHHH
T ss_pred          ---HHHHHHHHHHhCCHHHHHHHHHHHHHHHHHhCCHHHHHHHHHHHHHH

```

Supplementary Figure S7. SDS-PAGE of *E. coli* DH10B containing in Lane (1) pHERD20T plasmid with the region of the RB69 genome encompassing ORF53c and ORF52c and (2) pHERD20T plasmid alone. Both cultures were induced with 1% arabinose for 1 hour at 37°C. Black arrow in indicates a protein band with a mass consistent to the predicted mass of the full length RB69 ORF53_2c (63.9 kDa).

