

DoRead putative gene (54759- 54884 bp) has significant coding potential in both GeneMark\_smeg & TB and is part of an operon with the upstream and downstream genes as is common in F1 phages. Its aa sequence is MTYTIGVVAHTKHADQLIHGPQVATVFKANERNTWSWWRHKZ

Whereas PhagesDB & NCBI BLASTp yields multiple hits to glycosyltransferase including with phage Madiba gp 100, I note that it is Q4:S441 alignment albeit with 89% coverage and HHPred shows no significant hits. Would you still call this a glycosyltransferase, or NKF?

## PhagesDB

>Madiba\_100, glycosyltransferase, 478  
Length = 478

Score = 80.5 bits (197), Expect = 2e-15  
Identities = 34/38 (89%), Positives = 36/38 (94%)

Query: 4 TIGVVAHTKHADQLIHGPQVATVFKANERNTWSWWRHK 41  
T+ VA+TKHADQLIHGPQVATVFKANERNTWSWWRHK  
Sbjct: 441 TLNVVAYTKHADQLIHGPQVATVFKANERNTWSWWRHK 478

>Juniper1\_92, glycosyltransferase, 536  
Length = 536

Score = 80.5 bits (197), Expect = 2e-15  
Identities = 34/38 (89%), Positives = 36/38 (94%)

Query: 4 TIGVVAHTKHADQLIHGPQVATVFKANERNTWSWWRHK 41  
T+ VA+TKHADQLIHGPQVATVFKANERNTWSWWRHK  
Sbjct: 499 TLNVVAYTKHADQLIHGPQVATVFKANERNTWSWWRHK 536

## HHPred

Template alignment | Template 3D Structure | PDBe

1. 3P3V\_B PTS system, N-acetylgalactosamine-specific IIB component; PTS IIB COMPONENT, PHOSPHOTRANSFERASE, SUGAR TRANSPORT, STRUCTURAL GENOMICS, JOINT CENTER FOR STRUCTURAL GENOMICS, JCSG, PROTEIN STRUCTURE; HET: PGE, MSE, PEG; 1.65A (Streptococcus pyogenes serotype M1) SCOP: c.38.1.0, I.1.1.1  
Probability: 50.54%, E-value: 14, Score: 23.35, Aligned cols: 23, Identities: 17%, Similarity: 0.209, Template Neff: 7.3

Q ss_pred		CchhcceeEehhhcceecccc	
Q Q_2855369	1	MTYTIGVVAHTKHADQLIHGPQVAT	25 (42)
Q Consensus	1	~~~~fH~Vaym~f~p~]-hGm~V~T	25 (42)
		++...+...+... -.+  + +.++	
T Consensus	1	~~~~~]~~~~R~VDhRLIHGQ~~~~~	23 (163)
T 3P3V_B	1	GHTQPHLIMTR~VDERLIHGQ~GQL	23 (163)
T ss_dssp		CCCCCEEEEE~EETTGGGTT~HHH	
T ss_pred		CCCCCceeEEE~ECCCCchHH~HHH	

Template alignment | Template 3D Structure | PDBe

2. 3EYE\_A PTS system N-acetylgalactosamine-specific IIB component 1; STRUCTURAL GENOMICS, PHOSPHOTRANSFERASE, PSI-2, Protein Structure Initiative, New York SGX Research Center for Structural Genomics, NYSGXRC; 1.45A (Escherichia coli O157:H7) SCOP: c.38.1.0  
Probability: 37.87%, E-value: 31, Score: 21.83, Aligned cols: 24, Identities: 17%, Similarity: 0.23, Template Neff: 7.5

Q ss_pred		CchhcceeEehhhcceecccc	
Q Q_2855369	1	MTYTIGVVAHTKHADQLIHGPQVAT	25 (42)
Q Consensus	1	~~~~fH~Vaym~f~p~]-hGm~V~T	25 (42)
		++...+...+... -.+  + +.++	
T Consensus	2	~~~~~]~~~~R~VDhRLIHGQ~~~~~	25 (168)
T 3EYE_A	2	SLSSPHILLTR~IDhRLVHGQGVGT	25 (168)
T ss_dssp		-----CEEEEE~ECTTTTTT#####	
T ss_pred		CCCCCceeEEE~ECCCCchHHHHH	