**Table S2.** Protein homology analysis of gene products of the *S. thermophilus* 05-34 *eps* cluster

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gene\_locus tag** | **Gene name (BPGNa name)** | **Protein size (aa)** | **Protein accession number** | **Top PDBb match** | **Probability(%)c** | **Annotation of the matched PDB protein** |
| DIS31\_02520 | *epsA (wzr)* | 486 | PWG83868.1 | 2XXP\_A | 100 | LytR family transcriptional regulator CPS2A {*Streptococcus pneumoniae*} |
| DIS31\_02525 | *epsB (wzb)* | 243 | PWG83869.1 | 2WJE\_A | 100 | Tyrosine protein phosphatase CpsB (E.C.3.1.3.48) {*Streptococcus pneumoniae*} |
| DIS31\_02530 | *epsC (wzd)* | 230 | PWG83870.1 | 4WL1\_W | 99.90 | Polysaccharide Co-Polymerase WzzE {*Escherichia coli*} |
| DIS31\_02535 | *epsD (wze)* | 232 | PWG83871.1 | 3CIO\_D | 99.95 | Tyrosine-protein kinase etk (E.C.2.7.10.2) {*Escherichia coli*} |
| DIS31\_02540 | *epsE* | 455 | PWG83872.1 | 5W7L\_A | 100 | UDP-galactose phosphate transferase {*Campylobacter concisus*} |
| DIS31\_02545 | *epsFN* | 87 | -d | - | - | incomplete, only the N-teminal 87 amino acids of rhamnosyltransferases (WP\_014634443.1) |
| DIS31\_02550 | - | - | - | - | - | Pseudogene, no start codon, encoding only the C-terminal 80 amino acids of UDP-galactose phosphate transferase (ARE12406.1) |
| DIS31\_02555 | IS985 | 143 | - | - | - | Transposase, internal stop at position 144 |
| DIS31\_02560 | *epsG* | 259 | PWG83873.1 | - | - | DUF4422 domain-containing protein (WP\_014568768.1) |
| DIS31\_02565 | *epsH (wzy)* | 417 | PWG83874.1 | - | - | oligosaccharide repeat unit polymerase (WP\_024704134.1) |
| DIS31\_02570 | *epsI* | 322 | PWG83875.1 | 5HEA\_B | 99.95 | Putative glycosyltransferase (GalT1){*Streptococcus parasanguinis*} |
| DIS31\_02575 | *eps2C* | 59 | - | 4JLV\_A | 97.03 | C-terminal fragment of Membrane protein; Rossmann fold, tyrosine kinase, ATP-binding {*Staphylococcus aureus*} |
| DIS31\_02580 | *eps2D* | 85 | - | 3CIO\_D | 99.21 | Tyrosine-protein kinase etk (E.C.2.7.10.2) {*Escherichia coli*} |
| DIS31\_02585 | *epsJ* | 316 | PWG83876.1 | 3BCV\_A | 99.94 | Putative glycosyltransferase protein {*Bacteroides fragilis*} |
| DIS31\_02590 | *epsK (wzx)* | 471 | PWG83877.1 | 6NC9\_A | 100 | Lipid II flippase MurJ {*Thermosipho africanus*} |
| DIS31\_02595 | *(wzx)* | - | - | 6NC9\_A | 100 | Pseudogene, no start codon, lacking the 1-87bp of gene encoding WP\_014621629.  Lipid II flippase MurJ {*Thermosipho africanus*} |
| DIS31\_02600 | *glf* | 365 | PWG83878.1 | 4MO2\_A | 100 | UDP-galactopyranose mutase (E.C.5.4.99.9) {*Campylobacter jejuni* subsp. *jejuni*} |
| DIS31\_02605 | ISL3 | - | - | - | - | frameshifted; too many ambiguous residues; internal stop. (WP\_002263494.1) |

a Bacterial polysaccharide gene nomenclature system (Reeves et al., 1996)

b PDB, Protein Data Bank

c Protein structural similarity analyzed by HHpred tools (Soding, Biegert & Lupas, 2005)

d No match