

Supplementary Figure 1. Alignment of a single TBLASTN hit (Sbjct) of the WashingtonLake metagenome against the protein sequence of *wbdI* (A), and alignment of a single TBLASTN hit (Sbjct) of the GOS metagenome against the protein sequence of *pqaB* (B). Although the sequence identity of the environmental homologue to *wbdI* was relatively low (46%), the GDP-mannose binding site (grey shading) was present. In the case of *pqaB*, almost the entire sequence is conserved. Thus, the high identity (77%) indicate that the environmental homologue is functional.

A

```

WbdI      2      FITSDKFR ERIIKLVPLVSIDLLIENENGEYLFGLRNNRPAKNYFFVPGGRIRKNESIKNA 61
Sbjct    1      FITSDKFR +++ PLVSIDLL++N G+ L GLR NRPA+ +FVPGGR+ K+E++ A 60

WbdI     62      FKRISSMELGKEYGISGSVFNGVWEHFYDDGFFSEGE---ATHYIVLCYTLKVLKSEL-N 117
Sbjct   61      FRRLAHAEELGTPLERSQARLLGVYEHFYSDSVFGSGETNPGTHYVVLGYQLKISEPDLQD 120

WbdI    118      LPDDQHREYLWLTKHQINAKQDVHNYSKNYF 148
Sbjct  121      LPQSQHEKYRWWSSSTDMRASAMVHENTRAYL 151

```

Query coverage: 98%; Sequence identity: 46%

B

```

PqaB     9      LAFTAFIALYVYVIPVNSRLLWQPDETRYAEISREMLASGDWIVPHFLGLRYPFEKPIAGYW 68
Sbjct    1      LA AFIALY+++P+NS LLWQPDETRYAEISREML SGDWIVPHFLGLRYPFEKPIAGYW 60

PqaB    69      INSLGQWLFGATNFGVRAGAILTLLAAALVAVLTFRLWRDKRTALLASVIFLSLFAVYS 128
Sbjct   61      LNSIGQWLFGANFNGVRAGAIFATLLTALLVAVLAMRLWQDKRTAVLSSAIFLTFFSVYS 120

PqaB   129      IGTAVLDPMIALWLTAGMCCFWQGMQATTRTGKIGMFLLLGATCGLGVLTKGFLALAVP 188
Sbjct  121      IGTAVLDP+IALWL AGMCCFW+GMQATTRT KI FLLLG TCG+GV+TKGFLALAVP 180

PqaB   189      VVSVLPWVIVQKRWKDFLLYGWLAVLSCFVVVLPWAIARREADFWHYFFWVEHIQRFA 248
Sbjct  181      VLSVLPWVFAQKRWKDLILYGLLAVIGCVVIVLPWGIAIARREPDFWHYFFWVEHIQRFA 240

PqaB   249      MSDAQHKAPFWYYLPVLLAGSLPWLGLLPGALKLGRERN---GAFYLLGWTIMPLFFS 305
Sbjct  241      QDDAQHKAPIWYYIPVFAVAGTLPWLGLLPGALRTGWRERDVPRESALYLLSWIVMPILFFS 300

PqaB   306      IAKGKLPTYVLSCFAPIAILMARFVLHNVKEGVAALRVNNGGINLVFGIIGIVAAFVVSSW 365
Sbjct  301      IAKGKLPTY+LSCFAP+AIL ARF L ++G AALRVNG IN++FG+ GIVA +VS+ 360

PqaB   366      GPLKSPVWTHIET 378
Sbjct  361      GPLKTPVWVHIET 373

```

Query coverage: 66%; Sequence identity: 77%