

Table 1. Summary of relationships between a set of control genes and MaxQ* for the 573-genome sample, proteobacteria sample and γ -proteobacteria sample.

gene	Description	Direction	p_value1	p_value2	p_value3
<i>aroE</i>	shikimate dehydrogenase	+	0.001	6.31E-07	0.050
<i>leuB</i>	3-isopropylmalate dehydrogenase	+	0.025	0.001	0.048
<i>mmuM</i>	homocysteine S-methyltransferase	+	0.840	0.496	0.344
<i>astA</i>	arginine N-succinyltransferase	+	3.96E-06	0.001	0.013
<i>hisG</i>	ATP phosphoribosyltransferase	+	0.032	0.002	0.170
<i>aspB</i>	aspartate aminotransferase	-	0.015	0.001	0.003
<i>serC</i>	PSAT1; phosphoserine aminotransferase	+	2.15E-06	0.008	NA
<i>tyrB</i>	aromatic-amino-acid transaminase	+	0.010	0.396	0.298
<i>lysC</i>	aspartate kinase	+	0.193	0.069	0.004
<i>ureA</i>	urease subunit gamma	-	0.959	0.058	0.113
<i>paaF</i>	echA; enoyl-CoA hydratase	-	0.040	0.001	0.796
<i>argH</i>	ASL; argininosuccinate lyase	+	0.174	0.019	0.125
<i>hemH</i>	FECH; ferrochelatase	+	0.003	0.483	0.024
<i>paaK</i>	phenylacetate-CoA ligase	-	0.565	0.077	0.658
<i>glnA</i>	glutamine synthetase	+	0.435	0.227	0.217
<i>bioD</i>	dethiobiotin synthetase	+	4.19E-06	0.002	0.123
<i>cobF</i>	precorrin-6A synthase	-	0.416	0.293	0.435
<i>mecR1</i>	methicillin resistance protein	-	0.025	NA	NA
<i>ribB</i>	RIB3; 3,4-dihydroxy 2-butanone 4-phosphate synthase	+	0.004	0.002	0.007
<i>thiC</i>	thiamine biosynthesis protein ThiC	+	0.407	0.514	0.668
<i>moaD</i>	molybdopterin synthase sulfur carrier subunit	+	0.0002	7.29E-06	4.83E-05
<i>cefD</i>	isopenicillin-N epimerase	-	0.982	0.309	NA
<i>acdH</i>	short/branched chain acyl-CoA dehydrogenase	-	0.987	NA	NA
<i>spuC</i>	putrescine aminotransferase	-	0.516	0.061	0.104
<i>nicC</i>	6-hydroxynicotinate 3-monooxygenase	-	0.017	0.003	0.480

‘Direction’ indicates a positive or negative relationship between presence of the gene and MaxQ* (‘+’ means that genomes with the gene have higher mean MaxQ* than genomes without the gene); P_value1 refers to the significance of the MaxQ* differences between genomes that possess a homolog of the gene and those where the homolog is absent in the 573-genome sample, p-value2 in the proteobacteria sample, and p-value3 in the γ -proteobacteria sample. NA indicates that statistical tests cannot be performed due to lack of data.