

Word	Corrected P-value
transposase	2.19E-112
putative	5.23E-68
fimbrial	4.26E-47
iscro1	9.12E-40
protein	3.96E-36
island	7.46E-36
pathogenicity	7.46E-36
hypothetical	4.34E-25
membrane	1.37E-22
prophage	1.02E-21
outer	1.12E-21
effector	5.66E-20
curli	4.49E-17
is200	4.49E-17
pilin	1.16E-14
invasion	1.39E-14
secretory	2.38E-13
secreted	2.38E-13
sequence	4.75E-13
element	4.75E-13
virulence	2.04E-12
fibre	6.60E-12
usher	1.86E-11
adhesin	3.40E-11
secretion	7.22E-11
general	3.53E-10
t3ss	5.09E-10
fima	7.03E-10
pilus	1.88E-09
insertion	2.09E-09
fimbriae	2.48E-09
tail	5.63E-09
leader	1.28E-08
exported	2.77E-08
chaperone	4.64E-08
iscro3	5.94E-08
type	5.94E-08
assembly	1.02E-07
fragment	1.75E-07
triphosphohydrolase	2.53E-07
iscro4	2.53E-07
isec14	2.53E-07
unknown	0.000000327
fimbrial-like	5.58E-07
possible	7.24E-07
curlin	0.000000768
is102	1.11E-06
trap-type	0.000002512
ferrous	2.58E-06
dt dp-4-dehydrorhamnose	2.81E-06
pathway	3.22E-06
phosphoribulokinase	3.22E-06
poly-beta-1	3.22E-06
function	3.33E-06
lipoprotein	3.74E-06
large-conductance	4.39E-06
deoxyguanosinetriphosphate	4.39E-06
cadc	4.39E-06
fkbp-type	0.00000491
cryptic	6.56E-06
luxr	8.50E-06
production	8.70E-06
autotransporter	8.99E-06
galactitol-specific	1.83E-05
bactoprenol-linked	1.89E-05
euts	1.89E-05
mechanosensitive	3.88E-05
during	4.66E-05
porin	0.000049676
cellobiose-specific	7.15E-05
haemolysin	9.50E-05
dihydrooorotate	0.0001220835
tnpa	0.000146769
adherence	0.000146769
ferritin	0.000146769
stylti	0.000146769
associated	0.0001662196
ferritin-like	0.0002200841
trap	0.0002338847
potassium-efflux	0.0003132541
5-dehydrogenase	0.0003534953
caif	0.0003592484
rcsf	0.0003592484
glutathione-regulated	0.0005656823
conserved	0.0005656823
cadaverine	0.0005656823
luxr-family	0.0005656823
modulating	0.0005656823
5-epimerase	0.0006125402
lipopolysaccharide	0.0006574058
core	0.0007029668
universal	0.0007029668
-diacetylchitobiose-specific	0.0008812555
dna-binding	0.0011105287
tight	0.0011952509
yad-type	0.0014736815
yiam	0.0014736815
within	0.0014736815
cell-division	0.0014736815
minor	0.0017372251
stimulation	0.001881871
phage	0.0019646081
virulence-associated	0.0020811247
eutp	0.0020811247
gifsy-2	0.0020811247
stpa	0.0020811247
6-n-acetyl-d-glucosamine	0.0020811247
sugar	0.0020811247
stress	0.0024941973
fermentation	0.0024941973
trans-activator	0.0024941973
alkaline	0.0029902177
inducible	0.0040059426
integrase	0.0043434469

**Table S1.** List of words enriched in EMBL file description of beneficial losses. P-values are calculated by running Fisher's exact test on the number of repetitions of each word in the description of beneficial losses compared to other genes and the number of other words in the description of these two groups of genes. The P-values are then adjusted using the Benjamini, Hochberg, and Yekutieli method.