Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome

Ying and Seno et al

Supplementary figures and figure legends

Supplementary figure legends

Figure S1 Chromosomal locations of the DEGs. The positions of the DEGs_gr and DEGs_hs on the genome are indicated. The black line indicates the MG1655 genome, and the deleted segments are indicated by gray boxes. The up- and downregulated DEGs_gr are indicated in orange and purple, respectively (**A**). The up- and downregulated DEGs_hs are indicated in yellow and cyan, respectively (**B**). The right- and left-pointing arrowheads represent the genes encoded on the forward and complementary strands, respectively.

Figure S2 Statistical analysis. **A**. Significance test of the correlation of the DEGs_gr. To test for significance, 159 genes were randomly selected, and their correlation coefficients were calculated. The histogram shows the results of 10,000 permutations of random sampling, and the red points indicate the correlation coefficients of 159 DEGs_gr between MG1655 and MG1655_heatshock. **B**. Significance test of the correlation of the DEGs_hs. To test for significance, 95 genes were randomly selected, and their correlation coefficients were calculated. The histogram shows the results of 10,000 permutations of random sampling, and the red points indicate the correlation coefficients of the gene expression level of the 95 DEGs_hs between MG1655 and MDS42. The *P*-values are indicated. **C**. Significance test of the overall correlation of the gene expression levels between MG1655 and MDS42. To test for significance, a dummy vector of expression values was randomly generated while maintaining a certain level of correlation in the gene expression values with MG1655. This figure demonstrates that the effects of genetic interruption and external perturbation are not independent. **D**. The significance test of the overall correlation between MG1655 and MG1655_heatshock. A dummy vector was randomly generated as above and evaluated. The *P* values are indicated.

Figure S3 Genes of unknown function. All 3710 genes were divided into two categories, “Unknown” and “not Unknown,” according to Reliy *et al.* [45]. The distances from each gene to the genomic scars are plotted. The “Not Unknown” and “Unknown” categories are shown in the upper panel. The “Unknown” genes were classified as DEGs or nonDEGs and are plotted in the bottom panel. Statistical significance was calculated with the Wilcoxon-Mann-Whitney test, and the *P*-values are indicated.

Figure S4 A close-up view of the transition in major peaks. The vertical, broken lines in red indicate the major peaks at 663 and 773 kb, respectively.

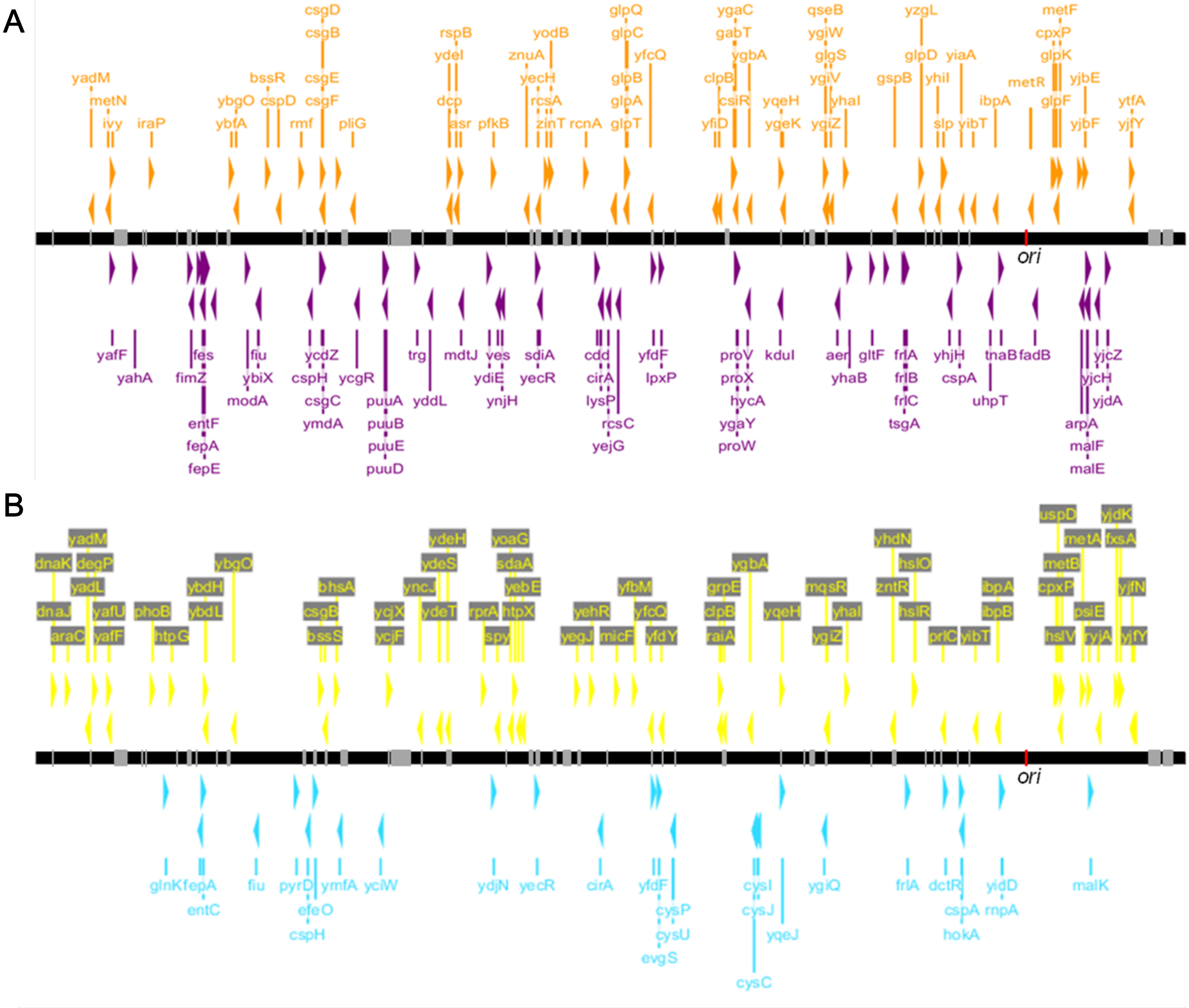
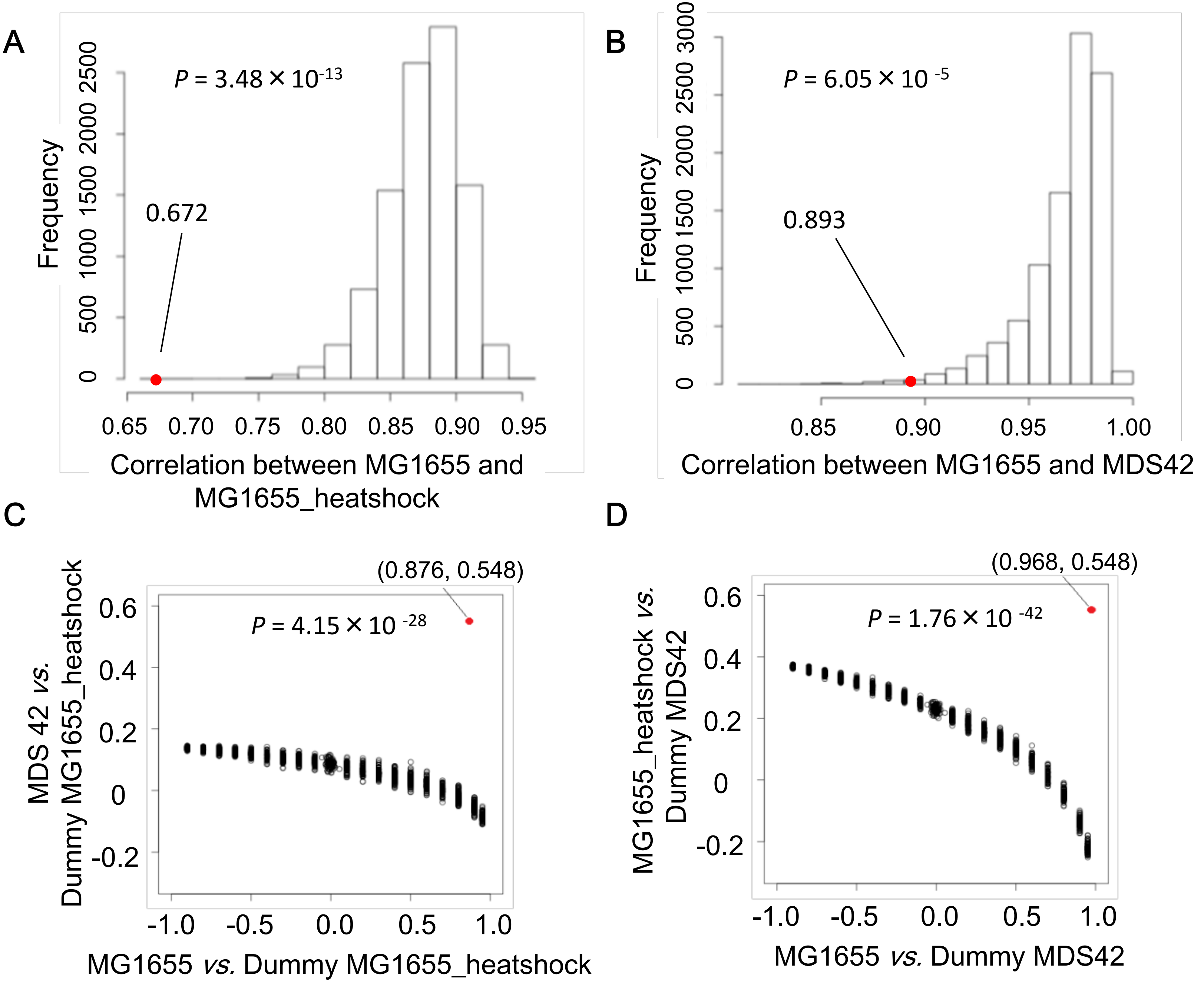

A
csgD
glpQ
ygaC
qseB
yzgL
metF
csgB
rspB
yodB
glpC
gabT
ygiW
cpxP
ydeI
znuA
yfcQ
ygbA
glgS
glpD
yiaA
glpK
yadM
bssR
csgE
yecH
glpB
clpB
ygiV
gspB
yhiI
metR
yjbE
metN
ybgO
cspD
csgF
dcp
rcsA
glpA
csiR
yqeH
yhaI
ibpA
glpF
ytfA
ivy
iraP
ybfA
rmf
pliG
asr
pfkB
zinT
rcnA
glpT
yfiD
ygeK
ygiZ
slp
yibT
yjbF
yjfY
ori
yafF
fes
fiu
ycdZ
trg
mdtJ
ves
sdiA
cdd
yfdF
proV
kduI
aer
gltF
frlA
yhjH
tnaB
fadB
yjcZ
yahA
fimZ
ybiX
cspH
ycgR
yddL
ydiE
yecR
cirA
lpxP
proX
yhaB
frlB
cspA
yjcH
modA
csgC
puuA
ynjH
lysP
hycA
frlC
uhpT
yjdA
entF
ymdA
puuB
rcsC
ygaY
tsgA
arpA
fepA
puuE
yejG
proW
malF
fepE
puuD
malE
B
uspD
yjdK
yadM
ydeH
yoaG
yhdN
metA
fxsA
dnaK
degP
ybgO
ydeS
sdaA
ygbA
hslO
metB
yadL
ybdH
bhsA
yncJ
yebE
yfbM
grpE
mqsR
zntR
ibpA
cpxP
yjfN
dnaJ
yafU
phoB
ybdL
csgB
ycjX
ydeT
rprA
htpX
yehR
yfcQ
clpB
yqeH
yhaI
hslR
ibpB
psiE
araC
yafF
htpG
bssS
ycjF
spy
yegJ
micF
yfdY
raiA
ygiZ
prlC
yibT
hslV
ryjA
yjfY
ori
glnK
fepA
fiu
pyrD
ymfA
yciW
ydjN
yecR
cirA
yfdF
cysI
ygiQ
frlA
dctR
yidD
malK
entC
efeO
cysP
cysJ
cspA
rnpA
cspH
cysU
yqeJ
hokA
evgS
cysC

Figure S1



A
P = 3.48 × 10-13
Frequency
0
500
1000
1500
2000
2500
0.672
0.65
0.70
0.75
0.80
0.85
0.90
0.95
Correlation between MG1655 and MG1655_heatshock
B
P = 6.05 × 10 -5
Frequency
0
500
1000
1500
2000
2500
3000
0.893
0.85
0.90
0.95
1.00
Correlation between MG1655 and MDS42
C
(0.876, 0.548)
P = 4.15 × 10 -28
MDS 42 vs. Dummy MG1655_heatshock
0.6
0.4
0.2
0
-0.2
-1.0
-0.5
0.0
0.5
1.0
MG1655 vs. Dummy MG1655_heatshock
D
(0.968, 0.548)
P = 1.76 × 10 -42
MG1655_heatshock vs. Dummy MDS42
0.6
0.4
0.2
0
-0.2
-1.0
-0.5
0.0
0.5
1.0
MG1655 vs. Dummy MDS42

Figure S2

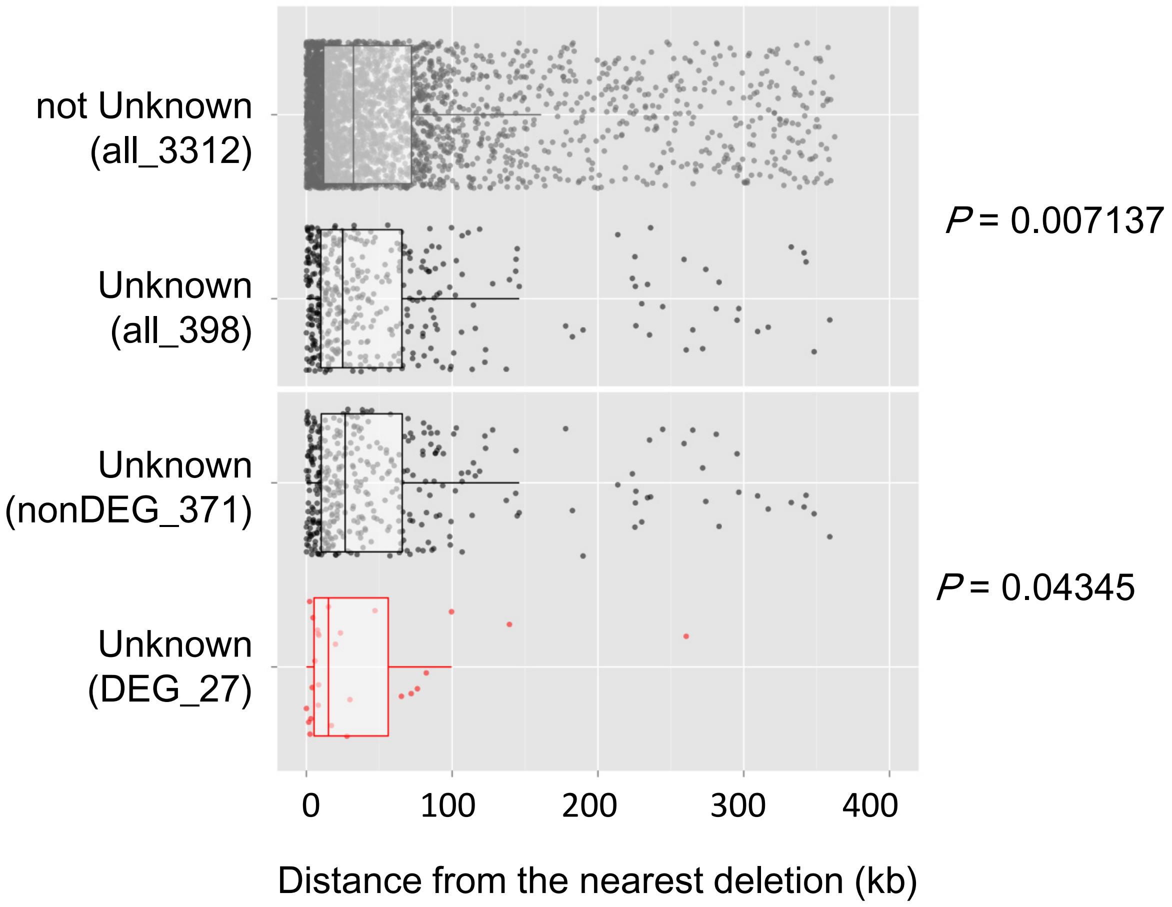

not Unknown (all_3312)
Unknown (all_398)
Unknown (nonDEG_371)
Unknown (DEG_27)
P = 0.007137
P = 0.04345
0
100
200
300
400
Distance from the nearest deletion (kb)

Figure S3

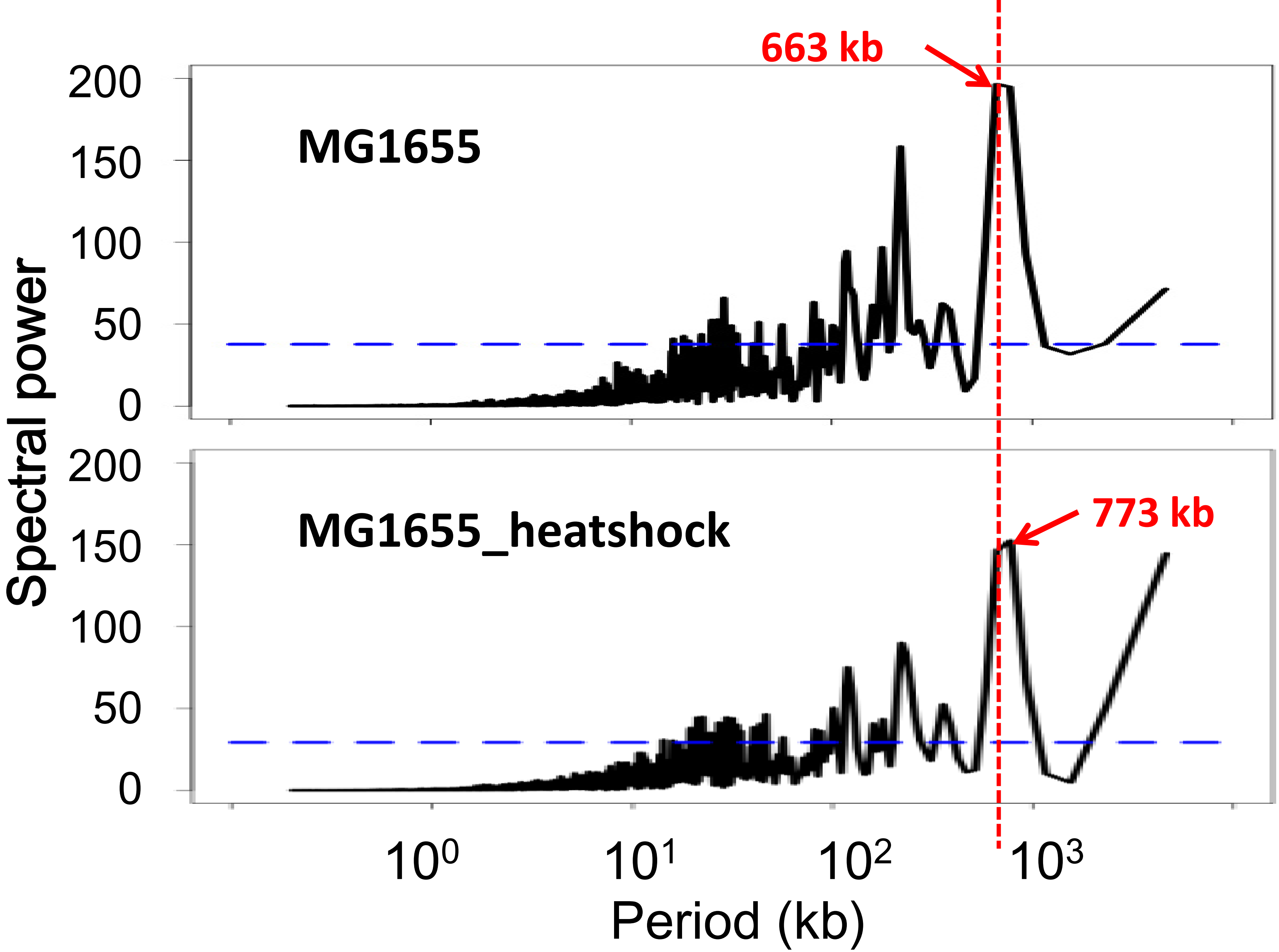

663 kb
MG1655
MG1655_heatshock
773 kb
Spectral power
0
50
100
150
200
0
50
100
150
200
10^0
10^1
10^2
10^3
Period (kb)

Figure S4