**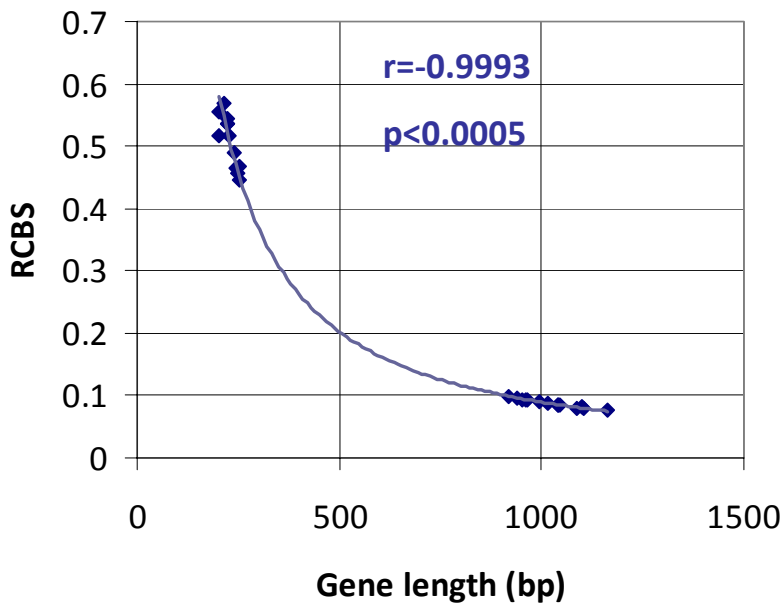Figure S1 –** Distribution of average RCBS values vs. length for 14 group lengths of arbitrary nucleotide sequences randomized 500 times (A) and 26 *E. coli* gene sequences, each randomized 500 times. Data from Roymondal *et al.* Tables SIIA and SIIB .
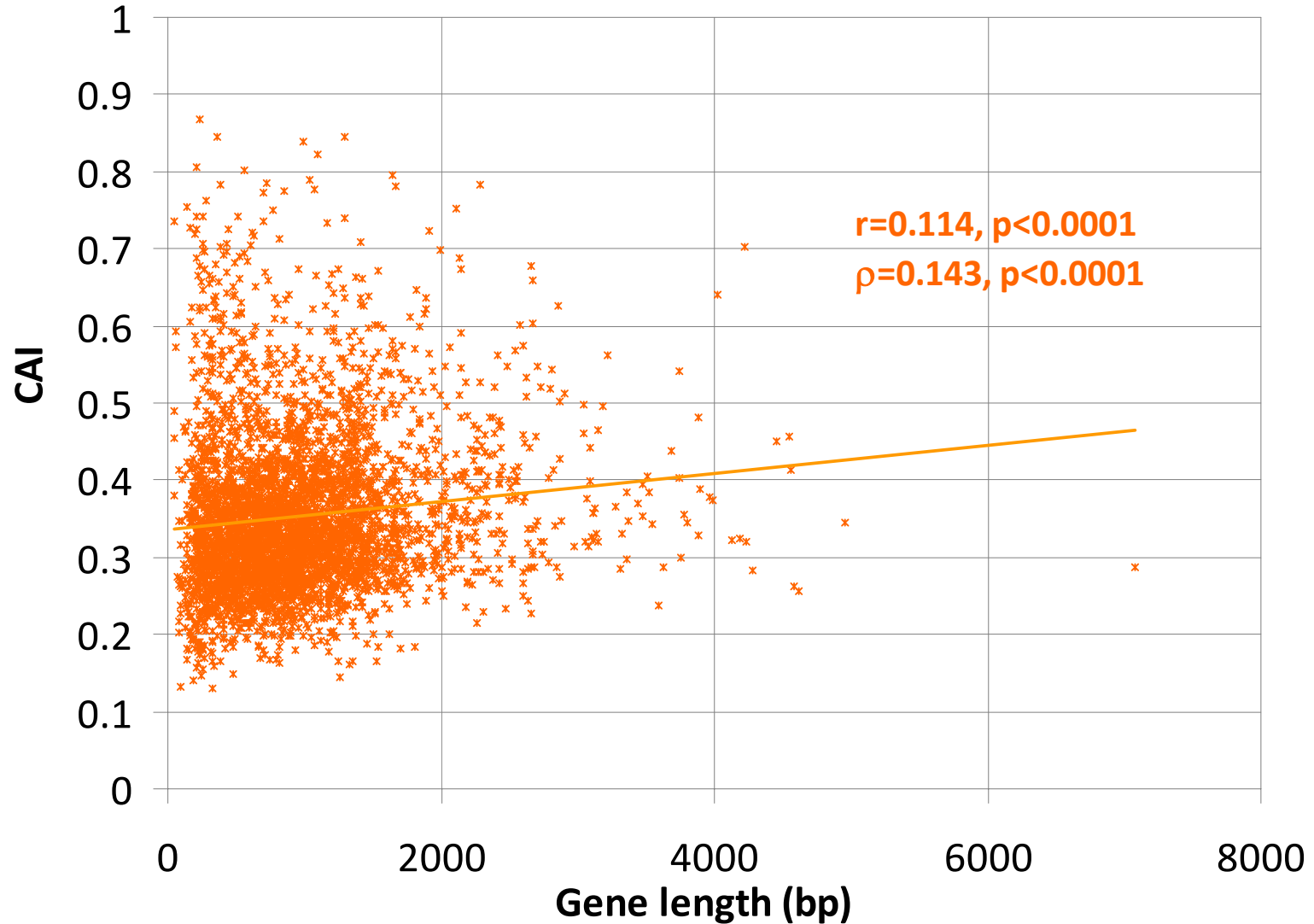
**Figure S2 –** Distribution of CAI values vs. length for the 4,029 *E. coli* genes in the M³ᴰ-derived expression dataset.
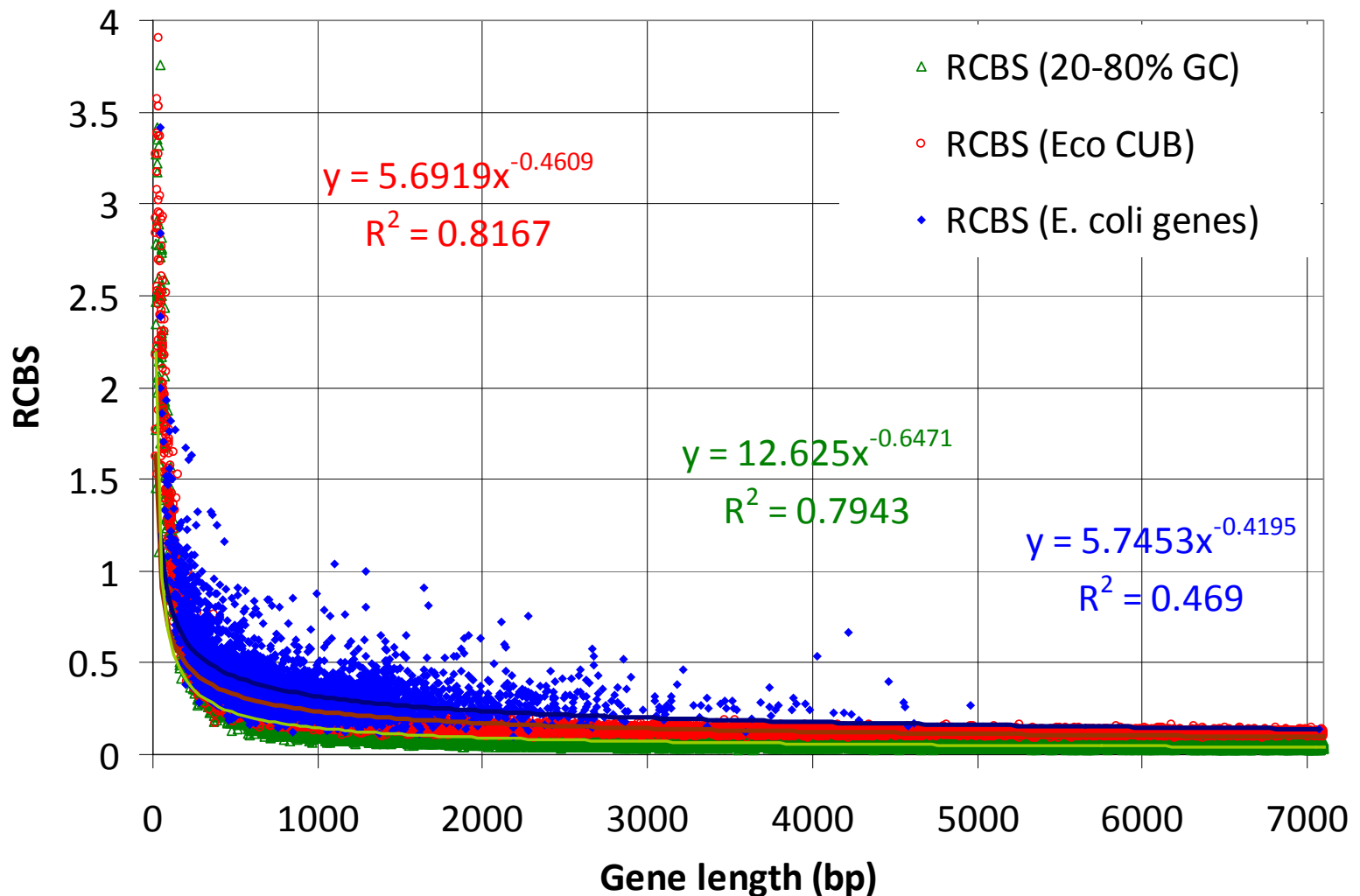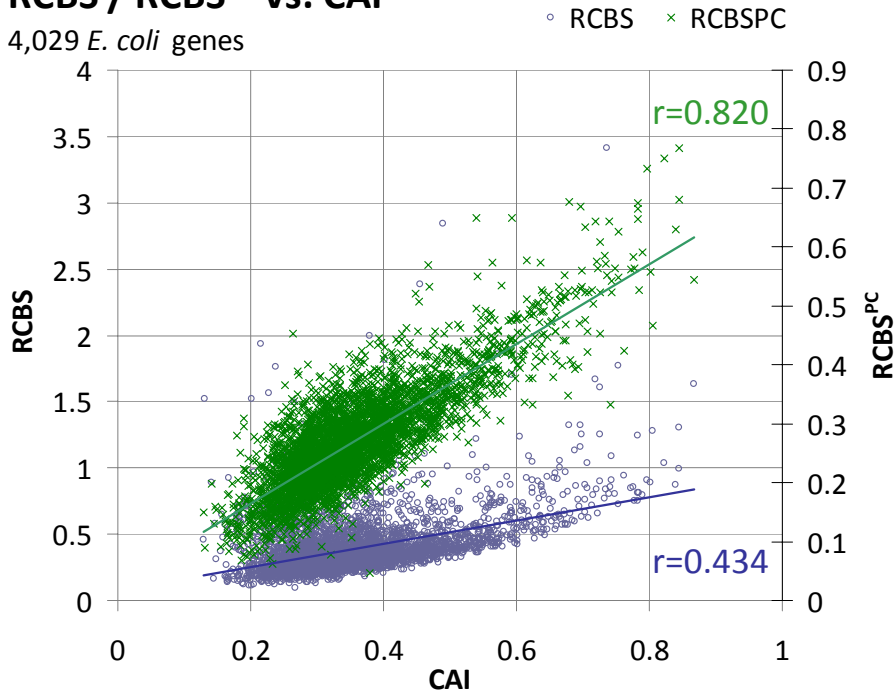
**Figure S3** – Distribution of RCBS values vs. length for the 4,029 *E. coli* genes in the M³ᴰ-derived expression dataset, 30,000 randomly generated sequences of random length with random %GC content in the 20%-80% range and 30,000 randomly generated sequences of random length following the *E. coli* codon usage bias (CUB). The power law fit equations are shown, together with their coefficients of determination ($r^2$). All correlation coefficients have $p<0.000001$.

(A)

RCBS / RCBS$^{PC}$ vs. CAI
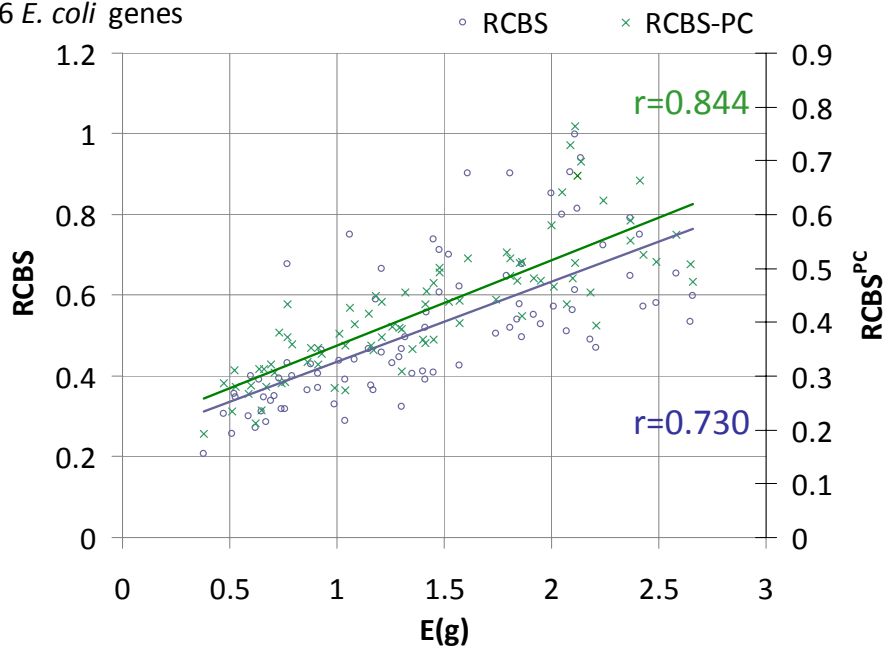
4,029 *E. coli* genes

(B)

RCBS & RCBS$^{PC}$ vs. E(g)

96 *E. coli* genes

**Figure S4 – (A)** Distribution of RCBS and RCBS$^{PC}$ values vs. CAI for the 4,029 *E. coli* genes in the M$^{3D}$-derived expression dataset. Correlation coefficients: $r$=0.433[***], $\rho$=0.448[***] for RCBS; $r$=0.820[***], $\rho$=0.753[***] for RCBS$^{PC}$. **(B)** Distribution of RCBS and RCBS$^{PC}$ values vs. E(g) for the 96 genes in the Karlin *et al.* dataset . Correlation coefficients: $r$=0.730[***], $\rho$=0.820[***] for RCBS; $r$=0.844[***], $\rho$=0.894[***] for RCBS$^{PC}$.
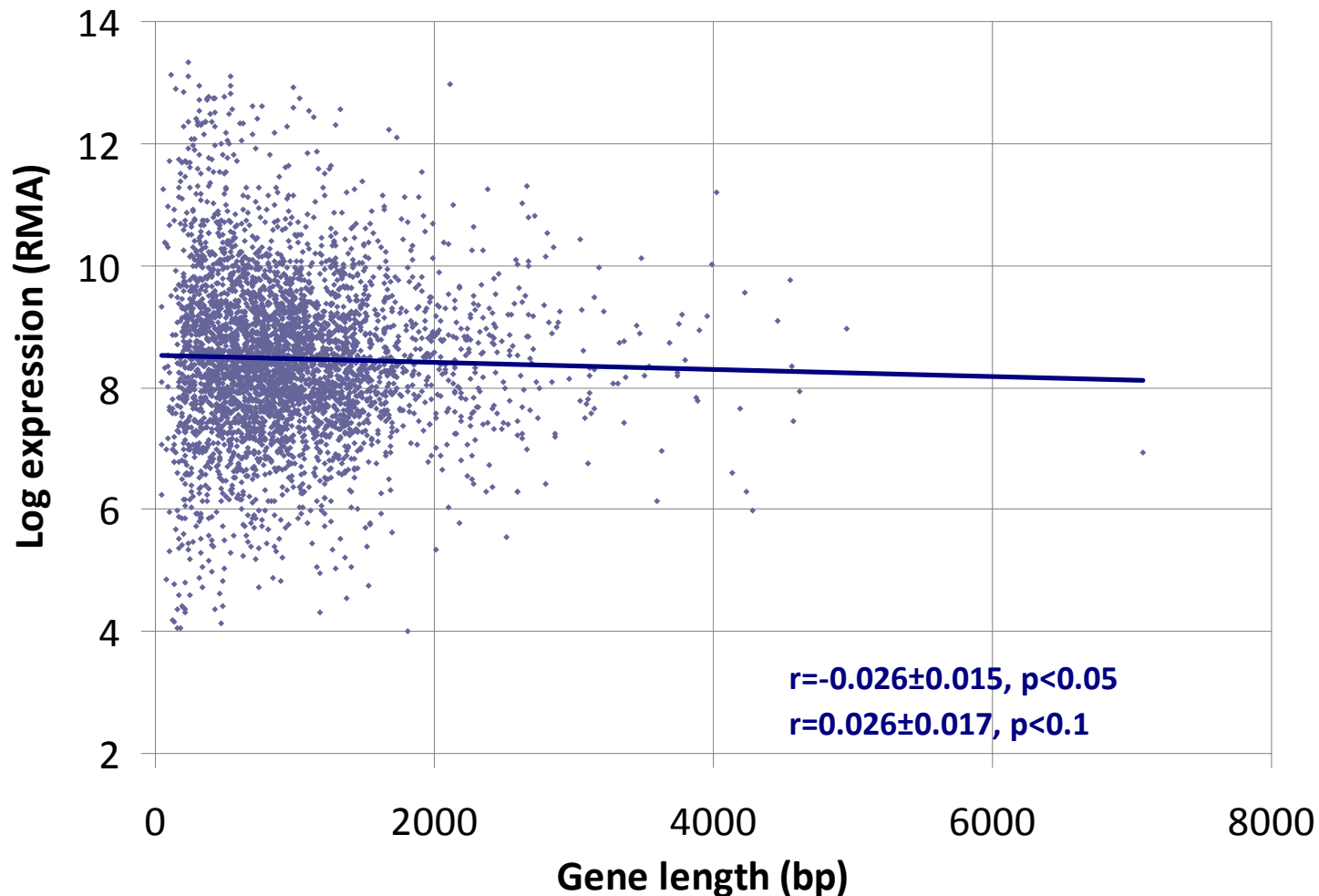
**Figure S5 –** Plot of log expression, normalized RMA values vs. gene length for 4,031 *E. coli* genes. The plot corresponds to the average log expression value for 31 experiments. Correlation coefficients were computed independently for each experiment and their mean and standard deviation are reported.

# Gene length histogram
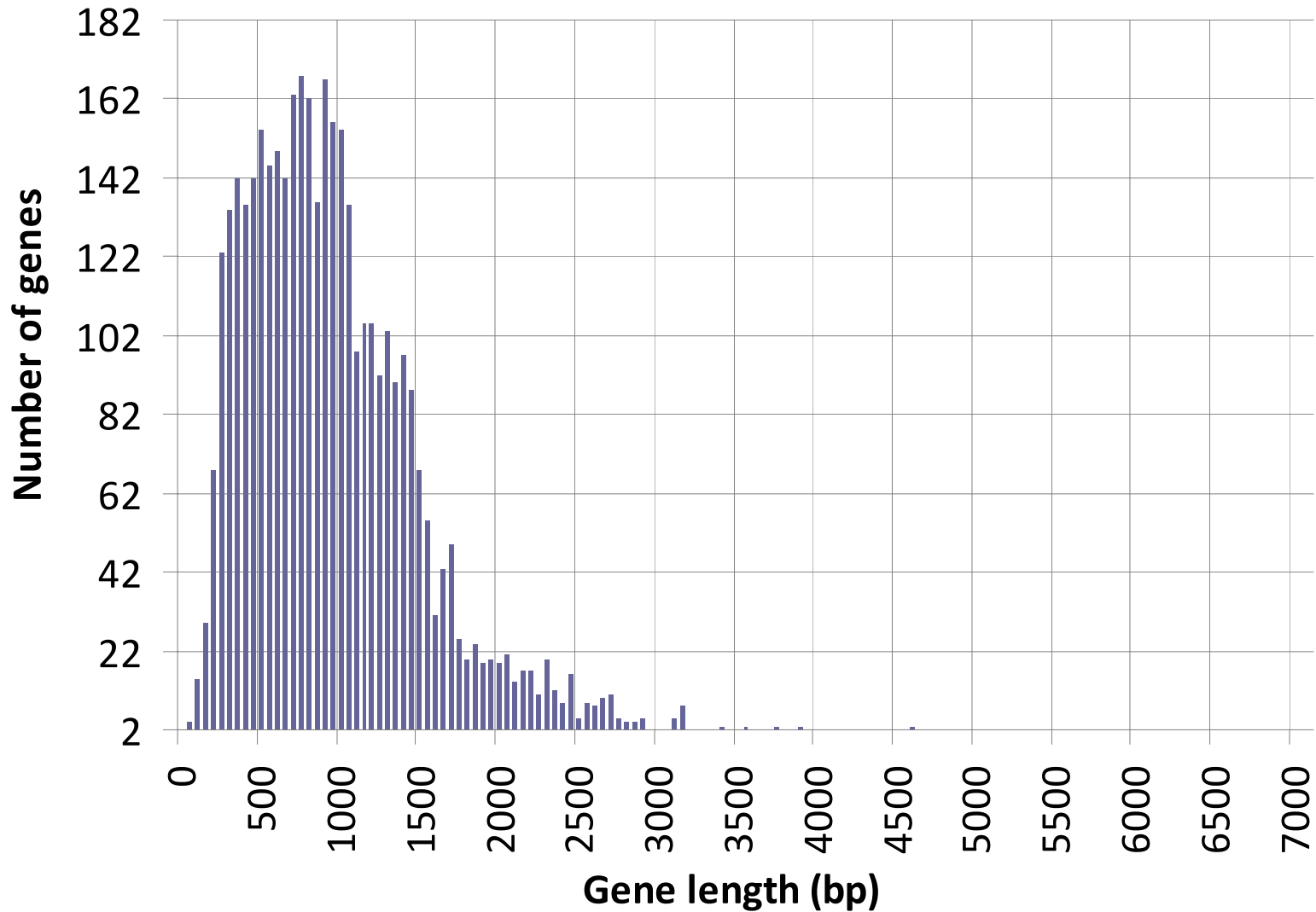
*E. coli* (4,031 genes)



**Figure S6** – Gene length histogram for 4,031 *E. coli* genes. Genes shorter than 1,000 bp account for more than a half (2,334 genes) of this dataset.