Methods S1

Justification of using log normal distribution.

The distribution from De Massy could be simplified by a gamma-type distribution. From preliminary data fitting of Tract-Seq simulated tracts, a log-gamma distribution was found to be the best fitting gamma-type distribution. Thus, we have compared the means obtained by fitting a log-gamma vs. a log-normal distribution for WT COs, WT NCOs, HI COs, HI NCOs, LO COs, LO NCOs, NCO 2 Chtd and NCO w/CO tracts. They varied on average by 1.25 bp (range=0-5 bp). We next compared the % of prototrophs at ARG4 as well as the CO association for Arg prototrophs. Both measures did not vary significantly and fit the data about equally well.

To find the best fitting distribution for our data between a log-normal and a log-gamma distribution in a more quantitative way, we computed a measure of the goodness-of-fit for non-nested models, the Deviance Information Criterion (DIC), for each model of the 8 types of tracts mentioned above. The DIC is a quality of fit statistic, similar to the residual sum of squares (RSS), that favors models with a better fit after penalizing for complexity (similar complexity for log-normal vs. log-gamma distributions); thus a small DIC is preferred. In all 8 cases, the DICs for the log-normal distributions were equal or smaller than those for the log-gamma distributions, providing empirical evidence for our choice. However, those DIC differences were minimal and both distributions provide good fits.

Supplemental References


