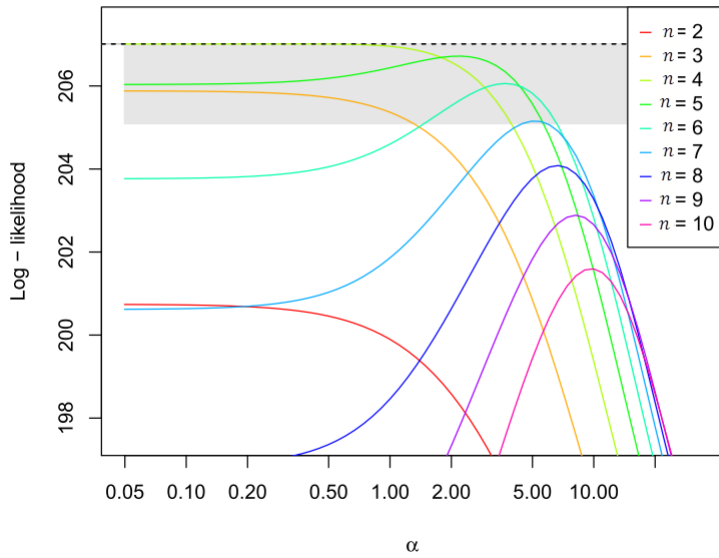


Log – likelihood profile: salmonella rplA gene



Log – likelihood profile: salmonella rpsT gene

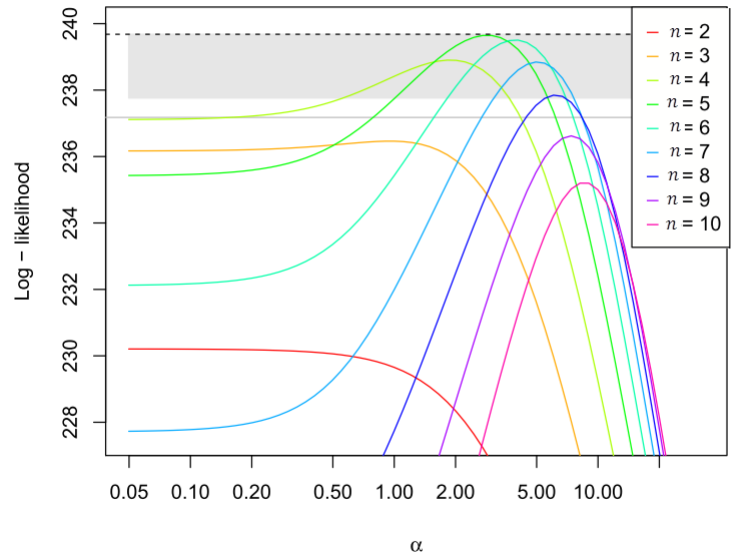


Figure S3 Log – likelihood profile for the parameter  $\alpha$  from empirical DFEs

For each gene, the log-likelihood of the fitted model (Figure 6 and Table 2) is given as a function of the parameter  $\alpha$ , for various integer values of  $n$  (see legend) and with the scale  $\lambda$  set to match the mean of the distribution ( $\lambda = 2\bar{s}/(n - 1 + \alpha)$ ). The dashed black line gives the log – likelihood of the best fitting gamma sum model (eq. (11) with  $n, \alpha, \lambda$  jointly fitted) and the gray plain line gives the log – likelihood of the best fitting pure gamma model (i.e. setting  $\alpha = 0$ ). The shaded area corresponds to log – likelihood lying within 1.92 points of the maximum log – likelihood for the best fitting model (with  $\alpha$  fitted), implying no significant difference in the fit to the data. In the *rplA* gene, the pure gamma models with  $n = 3, 4, 5$  or  $6$  are not statistically different from the best fitting model (Table 3): their maxima lie within the shaded area. For the *rpsT* gene, the ‘gamma sum’ model (eq. (11), with  $\alpha > 0$ ) provides a significantly better fit than a pure gamma model with  $\alpha = 0$  (the gray line lies below the gray shaded area). It further shows that the best fitting model (Table 3) and models with  $n = 4, 5, 6$  or  $7$ , with corresponding (increasing) values of  $\alpha > 0$ , cannot be distinguished in terms of goodness of fit (all their maxima lie within the shaded area).