bz-rates: a web-tool to estimate mutation rates from fluctuation analysis

Alexandre Gillet-Markowska, Guillaume Louvel, Gilles Fischer

G3: Genes|Genomes|Genetics Early Online, published on September 2, 2015 as doi:10.1534/g3.115.019836

bz-rates: a web-tool to estimate mutation rates from fluctuation analysis

Alexandre Gillet-Markowska*, Guillaume Louvel* and Gilles Fischer*,[1]

*Sorbonne Universités, UPMC Univ. Paris 06, Institut de Biologie Paris-Seine UMR 7238, Biologie Computationnelle et Quantitative, F-75005, Paris, France , CNRS, Institut de Biologie Paris-Seine UMR7238, Biologie Computationnelle et Quantitative, F-75005, Paris, France

ABSTRACT Fluctuation analyses is the standard experimental method for measuring mutation rates in microorganisms. The appearance of mutants is classically described by a Luria-Delbrück distribution composed of two parameters: the number of mutations per culture (m) and the differential growth rate between mutant and wild-type cells (b). A precise estimation of these 2 parameters is a prerequisite to the calculation of the mutation rate. Here, we developed *bz-rates*, a web-tool to calculate mutation rates that provides three useful advances over existing web-tools. First, it allows taking into account b, the differential growth rate between mutant and wild-type cells, in the estimation of m with the Generating Function (GF). Secondly, *bz-rates* allows the user to take into account a deviation from the Luria-Delbrück distribution called z, the plating efficiency, in the estimation of m. Finally, the web-site provides a graphical visualization of the goodness-of-fit between the experimental data and the model. *bz-rates* is accessible at http://www.lcqb.upmc.fr/bzrates.

KEYWORDS
Mutation-rate; Fluctuation-Assay; Luria–Delbrück

A classical approach to calculate mutation rates (μ) in microorganisms consists in performing fluctuation analyses through multiple cultures grown in parallel under identical conditions (Luria and Delbrück 1943). Each individual culture is started with a small inoculum such that the mutational events that occur during the culture are independent. Cultures are then plated on selective media to determine the number of mutants present in each culture. Estimating the mutation rate from these experimental data is of great interest for biologists and has been the object of many mathematical developments (for review see (Foster 2006)).

Calculating mutation rates requires to first estimate the the mean number of mutations per culture (m) under the assumptions of a Luria-Delbrück distribution model (Lea and Coulson 1949). Once a value of m has been calculated, the mutation rate μ can be easily inferred by dividing m by the total number of cells in the culture (although this can lead to an underestimation of the mutation rate (Ycart and Veziris 2014)). Most of the available estimators rely on the Maximum Likelihood (ML) method which was shown to be accurate for recovering m values (Zheng 2002; Stewart 1994; Jaeger and Sarkar 1995; Sarkar *et al.* 1992; Jones 1994; Gerrish 2008). However, ML estimators can become unstable for fluctuation assays involving cultures with large numbers of mutants. In such cases, the empirical probability GF remains robust and must be preferred over ML (Hamon and Ycart 2012).

One major parameter affecting the estimation of m is b, the differential fitness between mutant and wild type cells (i.e. the ratio between the mutant and the WT growth rates). In the case of differential growth rate, several estimators that jointly calculate m and b have long been made available (Koch 1982; Jones 1994; Jaeger and Sarkar 1995; Zheng 2002, 2005; Hamon and Ycart 2012). The code of these estimators is easily accessible but requires running command lines or installing third party tools.

In addition, the estimation of m can also be affected by another parameter: the plating efficiency, z. This criteria is defined as the fraction of the cultures that is plated on selective media. This parameter accounts for the fact that not all mutants are experimentally detected when only a fraction of the cultures is plated.

Here we propose a new integrated web-tool called *bz-rates* which provides three useful advances over the only web-tool available to estimate m (Hall *et al.* 2009). First, it allows taking into account b, the differential growth rate between mutant and wild-type cells, in the estimation of m with the GF. Note that *bz-rates* does not propose new mathematical developments but fully relies on the available GF estimator. Secondly, *bz-rates* allows the user to take into account the z deviation in the estimation of m by using the formulation suggested in (Foster 2006) and initially proposed by Stewart and collaborators (Stewart *et al.* 1990). Note that more

Manuscript compiled: Wednesday 2nd September, 2015%
[1]15 rue de l'Ecole de Médecine, UMR7238 Biologie Computationnelle et Quantitative, F-75005, Paris, France, gilles.fischer@upmc.fr

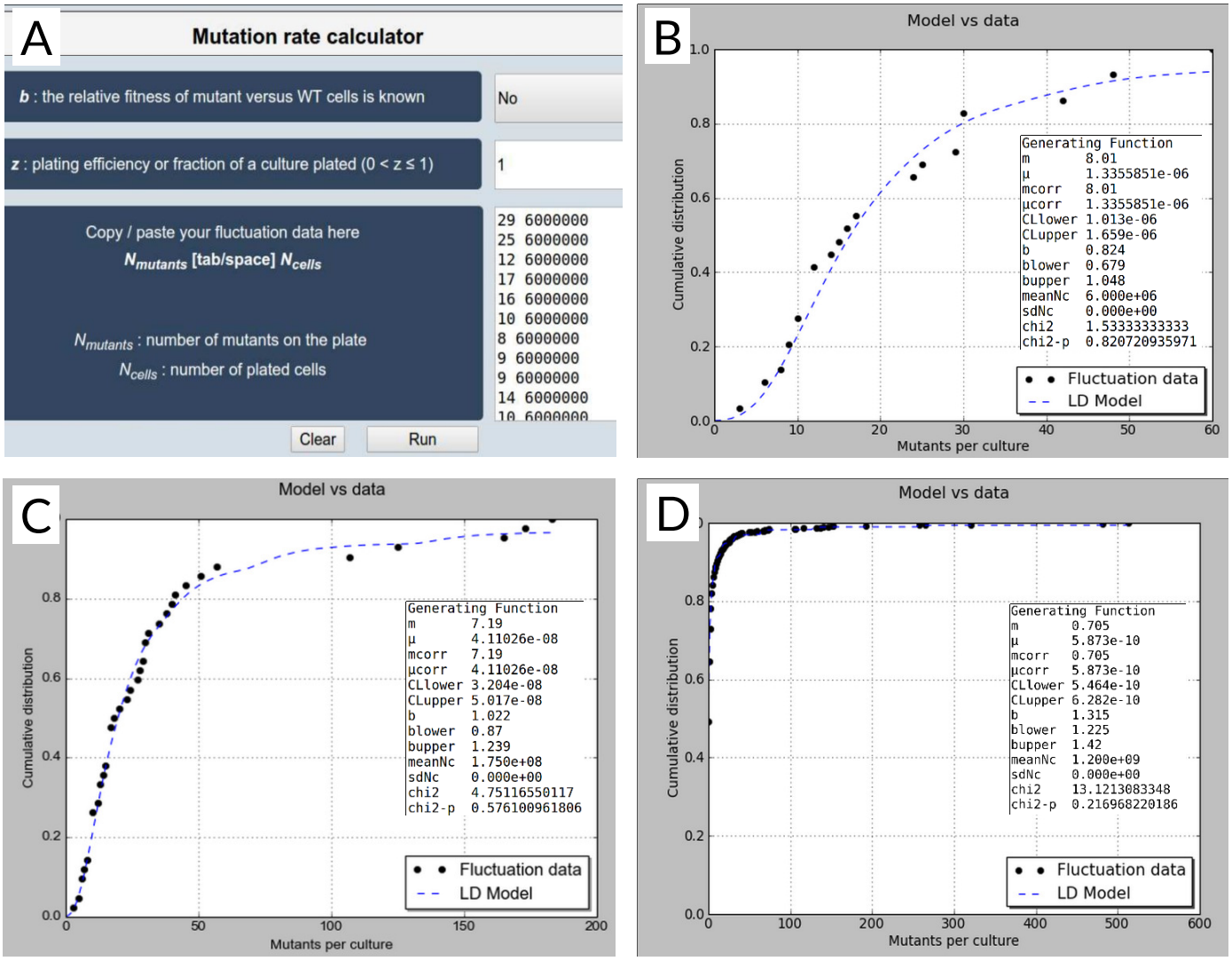

Figure 1 Screen-shots of the *bz-rates* web-site. (A) The input form is composed of 1 choice-field (for the b parameter) and 2 boxes (for the z parameter and a 2-columned data box (N_{mutant} N_{cells})). If the user chooses to manually specify a value for b, a supplementary box appears below the choice field. The z parameter is the plating efficiency which represents the fraction of a culture plated. The $N_{mutants}$ and N_{cells} box is intended to enter the number of plated mutants and plated cells in each culture, respectively. $N_{mutants}$ and N_{cells} must be spaced by a single white-space or a tabulation. Here, the $N_{mutants}$ and N_{cells} box is filled with the values from our experimental fluctuation assay described in the result section. (B to D) Each result section is composed of a numerical box (inside the plot) and a plot showing the cumulative distribution function fitted to the experimental data (B: results from our experimental fluctuation assay, (C) Results from a Luria and Delbrück fluctuation analysis of mutations conferring virus resistance in bacteria (corresponding to the pool of experiments number 1, 10, 11, 15 and 21 from table 2 in (Luria and Delbrück 1943)) (D) Results from a fluctuation experiment of mutations conferring nalidixic acid resistance in Escherichia Coli from Boe *et al.*.

recent formal mathematical treatments to this problem are also available but were not implemented here (Stewart 1991; Jones 1993; Zheng 2008a). Finally, *bz-rates* computes the goodness-of-fit – as described in (Boe *et al.* 1994) – between the experimental data and the two-parameter Luria-Delbrück model and provides the user with a graphical visualization of the fit.

METHODS

bz-rates code

bz-rates was developed in Python with the Django v1.6 framework. It is a free web-tool distributed under the terms of the GNU General Public License. *bz-rates* is accessible at http://www.lcqb.upmc.fr/bzrates. The source code, available at https://github.com/gillet/bzrates, can be easily modified to implement other estimators and clones of the tool can be set up elsewhere.

Fluctuation assay

Fluctuation assays were performed using a BY4741 yeast strain (*TRP1*Δ*5′(1-362)::natNT2*, *CYC1*Δ*::TRP1*Δ*3′*(864-958)*-hph*, *ura3*, *clb5*Δ*::KanMX4*) carrying 2 non-functional alleles of the *TRP1* gene involved in tryptophan biosynthesis on two different chromosomes. One copy is truncated in 3′ and the other copy in 5′ leaving a 400bp homology region repeated in the two alleles. A non allelic homologous recombination event between the 2 hetero-alleles generates a reciprocal translocation that restores tryptophan prototrophy. These mutant cells can therefore be easily selected by plating the cultures onto standard complete synthetic media depleted for tryptophan (CSM-TRP). Briefly, 30 parallel cultures (500μL) were started by inoculating into rich media (Broth Yeast Extract-Peptone-Dextrose) ∼100 cells per well in a 2mL deepwell plate. Cells were grown without agitation at 30°C until they reached an optical density of 0.85 (6.10^6 cells/mL) and plates were incubated for 4 days at 30°C before counting the number of mutants per plate.

Growth rates

The growth rate of 3 independent mutants and wild type cells was measured by doing growth curve experiments in 100 μL of rich media with a Tecan Sunrise robot in triplicates.

RESULTS

Implementation

bz-rates uses the empirical probability GF estimator from (Hamon and Ycart 2012; Ycart 2013). This method allows a precise estimation of m across a larger range of parameter values than the ML method. The cellular division time model chosen in *bz-rates* is not the classical exponential model but a constant division time model ('Dirac'). Although there is no universal cellular division time model as it depends on experimental conditions like the strain or the media, the 'Dirac' model is usually the most accurate for the estimation of b and as accurate as the exponential model for the estimation of m. Note that this division time model induces a positive bias in the estimation of large values of m (Ycart 2013).

When b is known, the value provided by the user is used to estimate m with the GF. However, when the mutant relative fitness b is not known, *bz-rates* estimates both m and b with the GF function.

The m_{corr} value that takes into consideration the plating efficiency z is calculated according to formula (41) in (Stewart *et al.* 1990): $m_{corr} = m \cdot (z-1)/(z \cdot ln(z))$. μ is defined as $m/\bar{N}p$ and μ_{corr} as $m_{corr}/\bar{N}t$ where $\bar{N}p$ and $\bar{N}t$ represent the mean number of cells per plate and per culture respectively. CL_{lower} and CL_{upper} provide the lower and upper confidence limits of μ_{corr} (level of confidence = 95%). σ_{Np} provides the standard deviation of the number plated cells.

To test the goodness-of-fit of the data to the model, *bz-rates* performs a Pearson's chi-squared test. The value of χ^2 gives the Pearson's chi-squared goodness of fit and the $\chi^2 - pval$ its associated p-value. The null hypothesis is rejected in the case $\chi^2 - pval < 0.01$ meaning that the cumulative distribution function does not fit with the experimental data (empirical cumulative distribution function). In this case, the user is warned that the estimation of the mutation rate is not reliable.

Interface

bz-rates is composed of a simple form (Fig. 1A). The first choice field provides the user with the possibility to indicate that b is known. In this case, the b field appears and the experimentally determined value of b can be filled in ($0 < b < \infty$). Otherwise b will be estimated computationally by the GF.

The second box allows to fill in the plating efficiency z (i.e. the proportion of cells from each culture that was plated, default value: $z = 1$).

The main field is the 'N_{mutant} N_{cells}' box that parses the fluctuation analysis counts. $N_{mutants}$ and N_{cells} are the number of plated mutants and plated cells per culture, respectively. This field is 'excel ready' thus counts can be directly copy/pasted into this box without further formatting.

The *bz-rates* result section is composed of two parts: the numerical and the graphical boxes (Fig. 1). The numerical box on the left provides the following estimates:

- m: mean number of mutations per culture not corrected by the plating efficiency (z)
- μ: mutation rate per cell per division not corrected by the plating efficiency (z)
- m_{corr}: number of mutations per culture corrected by the plating efficiency (z)
- μ_{corr}: mutation rate per cell per division corrected by the plating efficiency (z)
- CL_{lower}: lower 95% confidence limit for m_{corr}
- CL_{upper}: upper 95% confidence limit for m_{corr}
- b: mutant cells relative fitness predicted by the Generating Function (only output if b is left empty in the input field)
- b_{lower}: lower 95% confidence limit for b
- b_{upper}: upper 95% confidence limit for b
- $\overline{Nc}$: average number of plated cells per culture
- σ_{Nc}: standard deviation of the number of plated cells
- χ^2: Pearson's chi-square value
- $\chi^2 - p$: Pearson's chi-square p-value

The graphical box plots the cumulative distribution function fitted to the experimental data. It allows the user to visually judge for the correctness of the hypothesized distribution. To quantify the quality of the fit, *bz-rates* performs a Pearson's chi-squared goodness of fit as described in (Boe *et al.* 1994). If the null hypothesis is rejected (p-value<0.01), the user is advised by a red warning that the predicted and observed distributions are not in close agreement. In this case, the user should consider using another model that takes into consideration other deviations from the Luria Delbrück model such as, for instance, the post-plating growth (Lang and Murray 2008). To do so, the user should use an advanced mutation rate calculation packages to explore different models such as Salvador (Zheng 2008b).

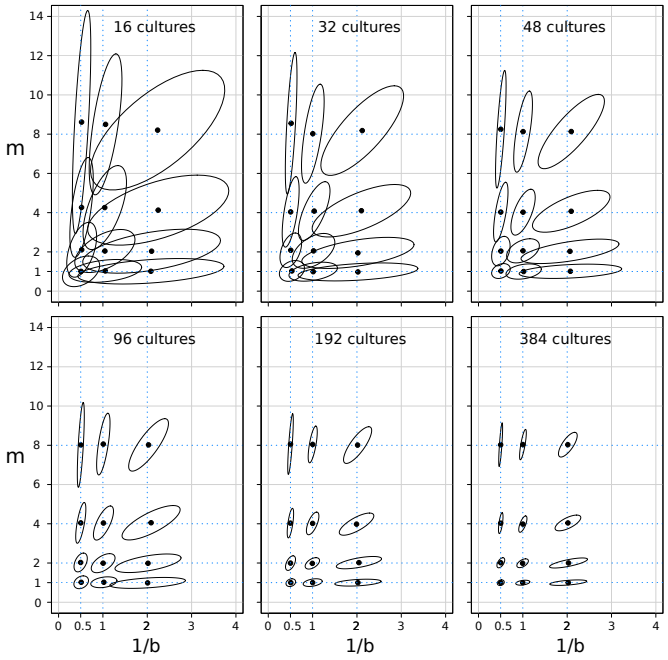

Figure 2 Performance of the *bz-rates* calculator on various simulated datasets. Each panel corresponds to simulated fluctuation datasets with either 16, 32, 48, 96, 192 or 384 independent cultures. In each panel, 200 simulations were performed for different values of m (1, 2, 4 and 8) and b (0.5, 1 and 2). The ellipses show the 95% dispersion of *bz-rates* estimations for the 200 simulations.

Experimental testing

A fluctuation assay was performed with a yeast strain carrying a genetic system that is designed to generate a functional copy of an auxotrophic gene when the cells undergo a specific chromosomal rearrangement (a reciprocal translocation, see methods). The resulting mutant cells have a strong growth defect relatively to wild type cells, probably as the result of the translocation, that was experimentally measured to 0.76 (see methods).

The form of Fig. 1 (A) is filled with the data of the 30 tubes fluctuation assay that was undertaken. We neglected to specify the mutants relative growth rate in order to compare the predicted relative growth rate of *bz-rates* to the experimental measure. Fig. 1 (B) shows *bz-rates* results. The plot indicates a good fit between the statistical distribution of the mutants and the experimental data. Pearson's chi-square goodness of fit value (1.16) and p-value (0.88) are displayed at the end of the numerical box on the left. The mutation rate (μ) is estimated to $1.33.10^{-6}$ per cell per division (95% confidence limits (CL) [$1.01.10^{-6}$ - $1.66.10^{-6}$]) and the predicted mutant relative growth rate (0.82 [0.68 - 1.05]) is in close agreement with the experimental measure (0.76).

Published datasets and simulations

In order to test our implementation and the stability of the GF estimator in *bz-rates*, we tested 2 published datasets: (i) the first dataset corresponds to a historical fluctuation assay composed of 42 parallel cultures performed by Luria and Delbrück (Fig. 1 (C)). With this dataset, *bz-rates* predicts a mutation rate ($4.11.10^{-8}$) close to the one calculated by (Luria and Delbrück 1943) ($2.48.10^{-8}$). The value of m (7.19) and b (1.022) are very close to the range of values reported in (Hamon and Ycart 2012) ([5.22-8.89] for m and [0.74-1.22] for b). We also tested one larger fluctuation dataset from (Boe *et al.* 1994) that is composed of 1102 cultures (Fig. 1 (D)). In this case, *bz-rates* reports a m value of 0.705 which is close to the one calculated in (Hamon and Ycart 2012) ([0.65-0.77]) and the one calculated in (Zheng 2005) (0.71). The mutant differential growth rate value is 1.315 which is a bit higher than the one reported by (Zheng 2005) with a maximum likelihood approach (b=1.193).

The performance of *bz-rates* was also tested on simulated datasets. We generated simulated fluctuation assays for different couples of m and b with either 16, 32, 48, 96, 192 or 384 parallel cultures (Fig. 2). As expected, the precision of the estimator increases with increasing numbers of parallel cultures. The general trend that can be inferred from these plots is that the precision on the estimation of m (and by consequence the estimation of μ) is higher for the smallest values of m. Therefore, users should not outgrow the cultures in order to limit the number of mutants that grow on selective plates.

Note that the GF estimator has also been extensively tested elsewhere (Hamon and Ycart 2012; Ycart 2013) and the reader should refer to these papers for an extensive review of the performance of this estimator.

CONCLUSION

bz-rates is a web-tool that does not require the installation of any third party tool or run any command line to estimate mutation rates. It has a minimalist design in order to provide biologists with a web-tool the most straightforward as possible. To our knowledge there was so far a single web-tool available for mutation rate calculation (Hall *et al.* 2009) but this tool does not allow to consider deviations from Luria Delbrück or to estimate the goodness of fit with the model. Therefore, *bz-rates* provides useful advances such as accounting for 2 important deviations to Luria-Delbrück distributions (b and z) as well as giving an indication of the reliability of the estimated mutation rates. We hope that *bz-rates* will reveal useful to a broad community of microbiologists and geneticists.

ACKNOWLEDGEMENT

We thank our colleagues from LCQB for fruitful discussions and particularly Nicolas Agier for invaluable tips and advice. This work was supported by a scholarship from "La Ligue contre le cancer" and by a Convergence grant (Memory) from Idex SUPER Sorbonne Université 2014.

LITERATURE CITED

Boe, L., T. Tolker-Nielsen, K. M. Eegholm, H. Spliid, and a. Vrang, 1994 Fluctuation analysis of mutations to nalidixic acid resistance in Escherichia coli. Journal of Bacteriology **176**: 2781–2787.

Foster, P. L., 2006 Methods for determining spontaneous mutation rates. Methods in enzymology **409**: 195–213.

Gerrish, P. J., 2008 A Simple Formula for Obtaining Markedly Improved Mutation Rate Estimates. Genetics **180**: 1773–1778.

Hall, B. M., C.-X. Ma, P. Liang, and K. K. Singh, 2009 Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics (Oxford, England) **25**: 1564–5.

Hamon, A. and B. Ycart, 2012 Statistics for the Luria-Delbrück distribution. Electronic Journal of Statistics **6**: 1251–1272.

Jaeger, G. and S. Sarkar, 1995 On the distribution of bacterial mutants: the effects of differential fitness of mutants and non-mutants. Genetica pp. 217–223.

Jones, M. E., 1993 Accounting for plating efficiency when estimating spontaneous mutation rates. Mutation Research/Environmental Mutagenesis and Related Subjects **292**: 187–189.

Jones, M. E., 1994 LB Fluctuation Experiements; Accounting Simultaneously for Plating Efficiency and Differential Growth Rate. J. theor. Biol **166**: 355–363.

Koch, A. L., 1982 Mutation and growth rates from Luria-Delbrtick fluctuation tests. Mutation Research **95**: 129–143.

Lang, G. I. and A. W. Murray, 2008 Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics **178**: 67–82.

Lea, D. and C. A. Coulson, 1949 The distribution of the numbers of mutants in bacterial populations. Journal of genetics **49**: 264–285.

Luria, E. and M. Delbrück, 1943 Mutations of bacteria from virus sensitivity to virus resistance'-'. Genetics **28**: 491–511.

Sarkar, S., Ma, and Sandri, 1992 On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. Genetica pp. 173–179.

Stewart, F. M., 1991 Fluctuation analysis: the effect of plating efficiency. Genetica **84**: 51–55.

Stewart, F. M., 1994 Tests: How Reliable Are the Estimates of Mutation Rates ? Genetics **1146**: 1139–1146.

Stewart, F. M., D. M. Gordon, and B. R. Levin, 1990 Fluctuation Analysis: The Probability Distribution of the Number Mutants Under Different Conditions. Genetics **124**: 175–185.

Ycart, B., 2013 Fluctuation analysis: can estimates be trusted? PloS one **8**: e80958.

Ycart, B. and N. Veziris, 2014 Unbiased estimation of mutation rates under fluctuating final counts. PLoS ONE **9**.

Zheng, Q., 2002 Statistical and algorithmic methods for fluctuation analysis with SALVADOR as an implementation. Mathematical Biosciences **176**: 237–252.

Zheng, Q., 2005 New algorithms for Luria-Delbrück fluctuation analysis. Mathematical biosciences **196**: 198–214.

Zheng, Q., 2008a A note on plating efficiency in fluctuation experiments. Mathematical Biosciences **216**: 150–153.

Zheng, Q., 2008b SALVADOR 2.3: A tool for studying mutation rates.