The rate and molecular spectrum of spontaneous mutations in the GC-rich multi-chromosome genome of *Burkholderia cenocepacia*

Marcus M. Dillon, Way Sung, Michael Lynch, and Vaughn S. Cooper

Supporting Information

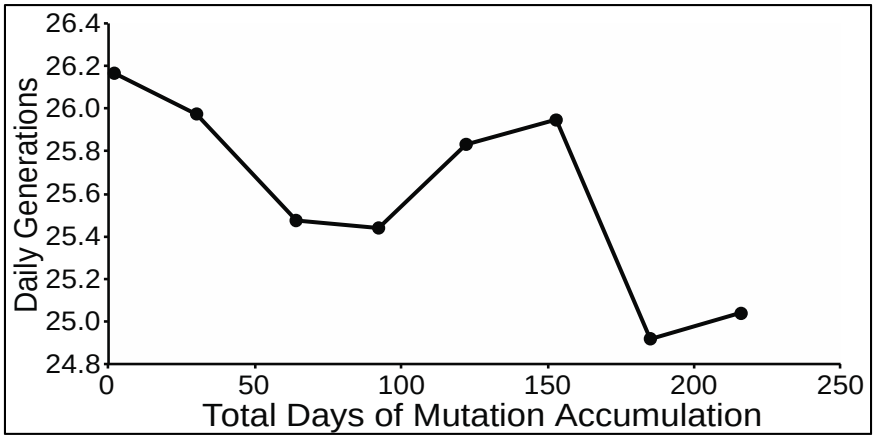

Figure S1 Estimates of the number of generations each line experienced per day, measured monthly over the course of the MA experiment.

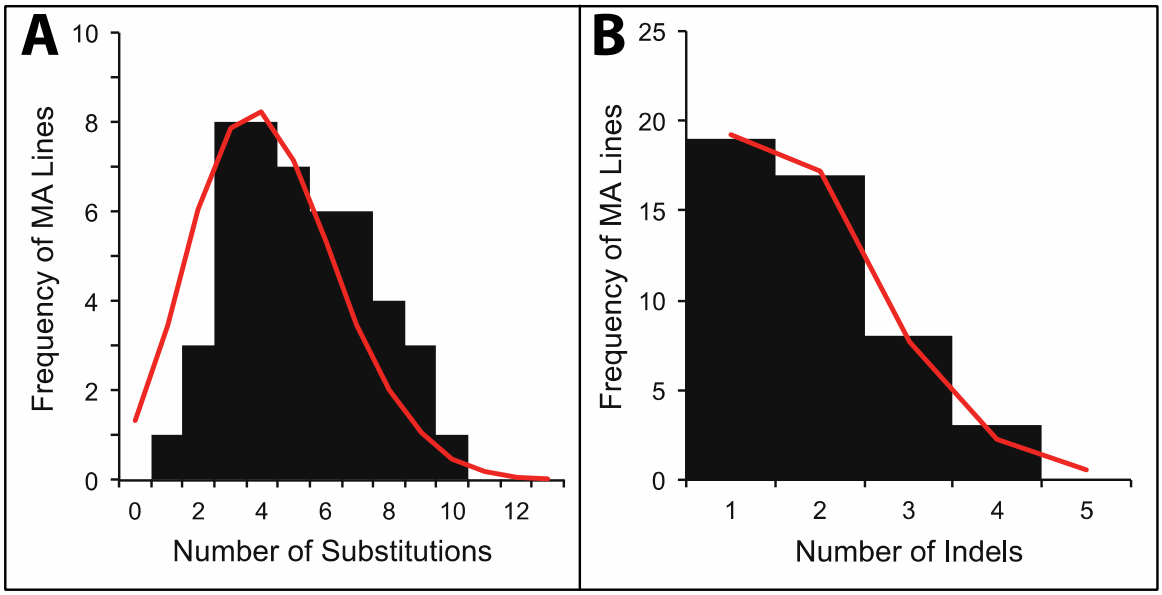

Figure S2 Frequency distributions of the number of base substitutions (A) and indels (B) identified per line. Neither distribution differs significantly from a Poisson distribution (bps: $\chi^2 = 1.81$, $p = 0.99$; indels: $\chi^2 = 0.48$, $p = 0.92$).

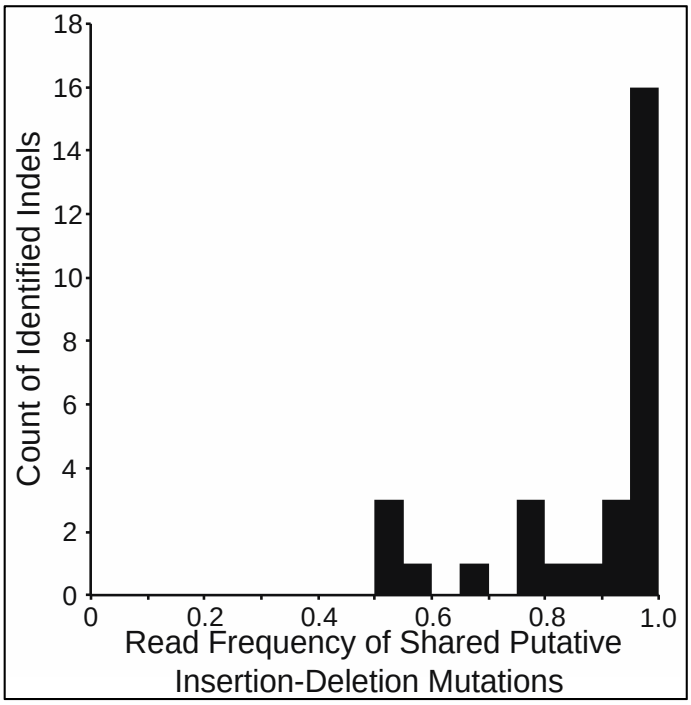

Figure S3 Distribution of the full-coverage read frequencies leading to the identification of each insertion-deletion mutation (indel) in this study. Full-coverage read frequency is the number of reads with >25 bases on each side of the indel mutation that supported the call, divided by the total reads with >25 bases on each side of the indel.

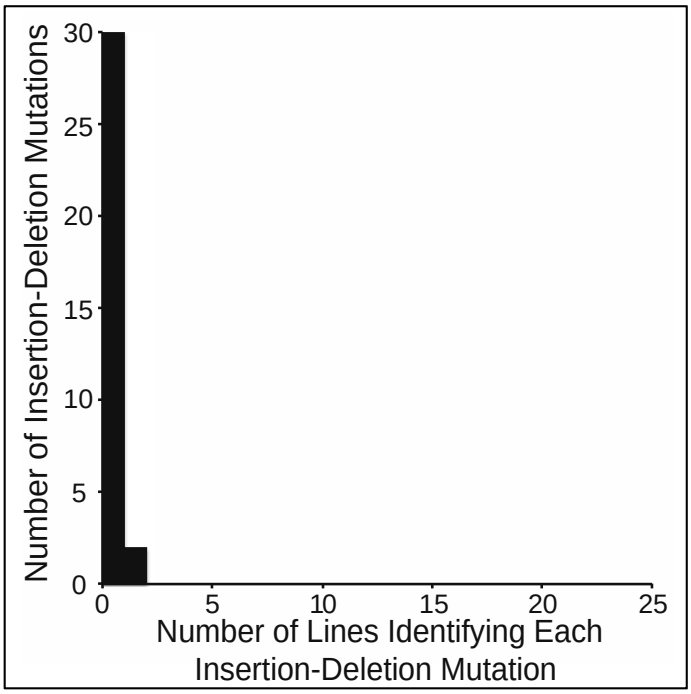

Figure S4 Distribution of the number of MA lines that identified each insertion-deletion mutation (location, size, and motif) called in this study.

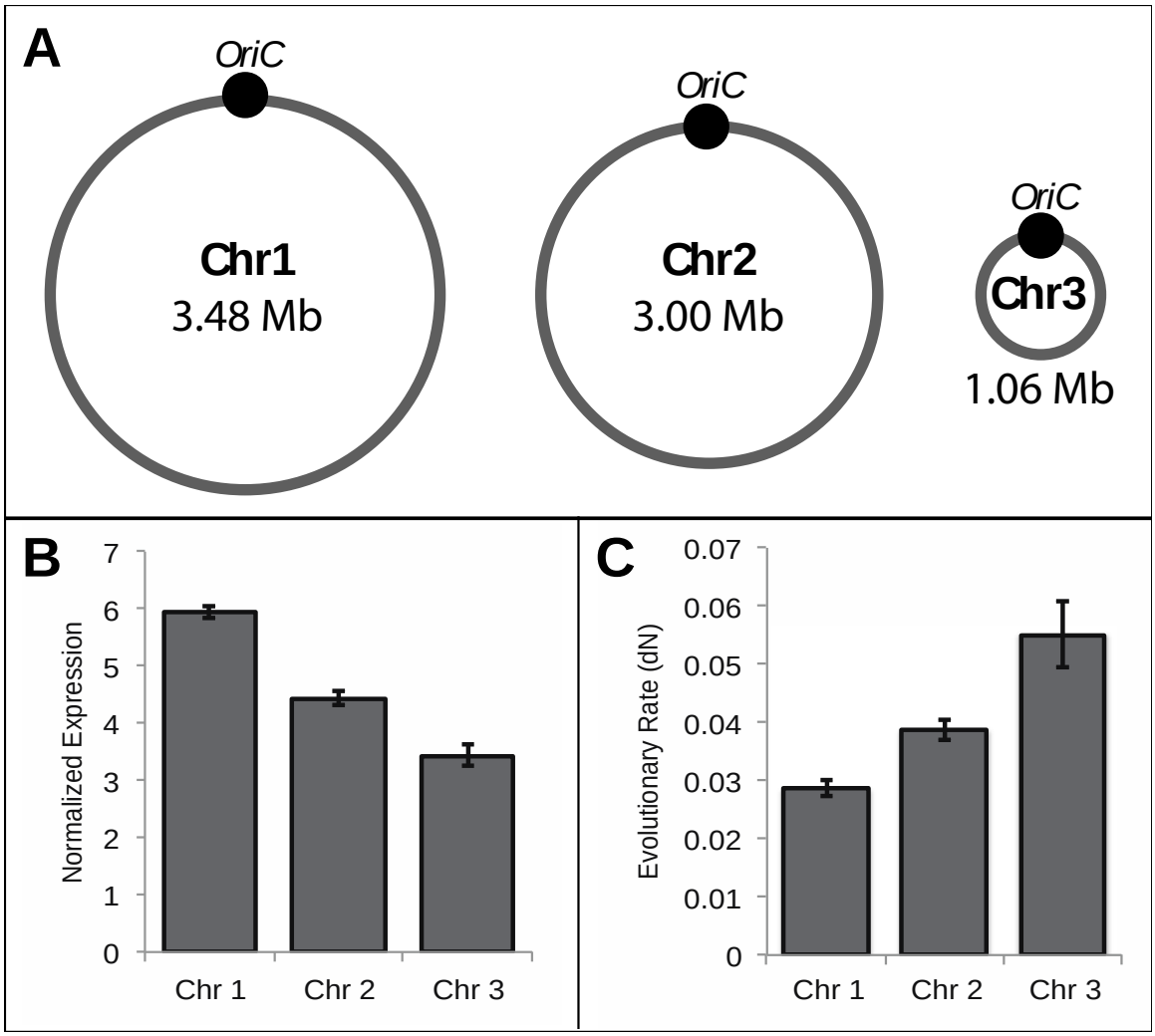

Figure S5 Chromosome size and expression (derived from RNAseq as described in (GOUT *et al.* 2013)) decline from chromosome 1 (chr1) to chromosome 3 (chr3) (A,B), while evolutionary rate is lowest on chr1 and highest on chr3 (C) (MORROW and COOPER 2012).

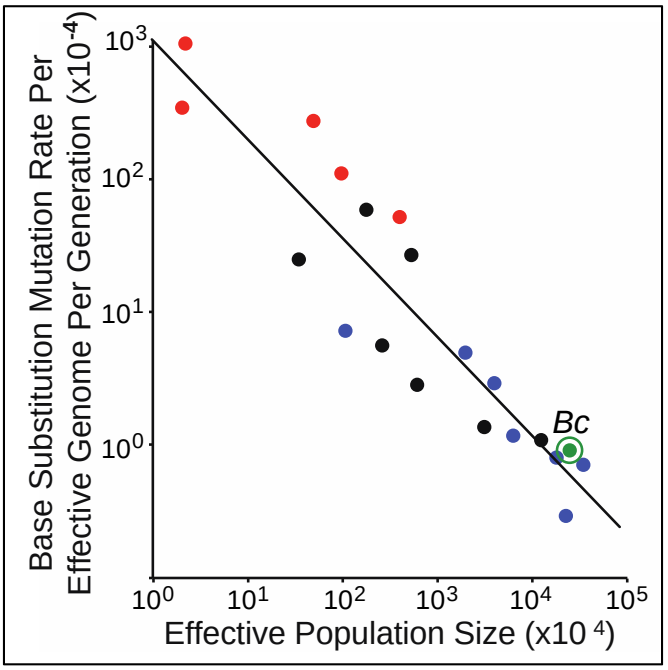

Figure S6 Relationship between base substitution mutation rate per effective genome size per generation and effective population size (N_E) in five multicellular eukaryotes (red), seven unicellular eukaryotes (black), and eight prokaryotes (blue; *B. cenocepacia* – green) (SUNG *et al.* 2012). The log-linear regression is highly significant (r^2=0.85, p<0.0001, df=19).

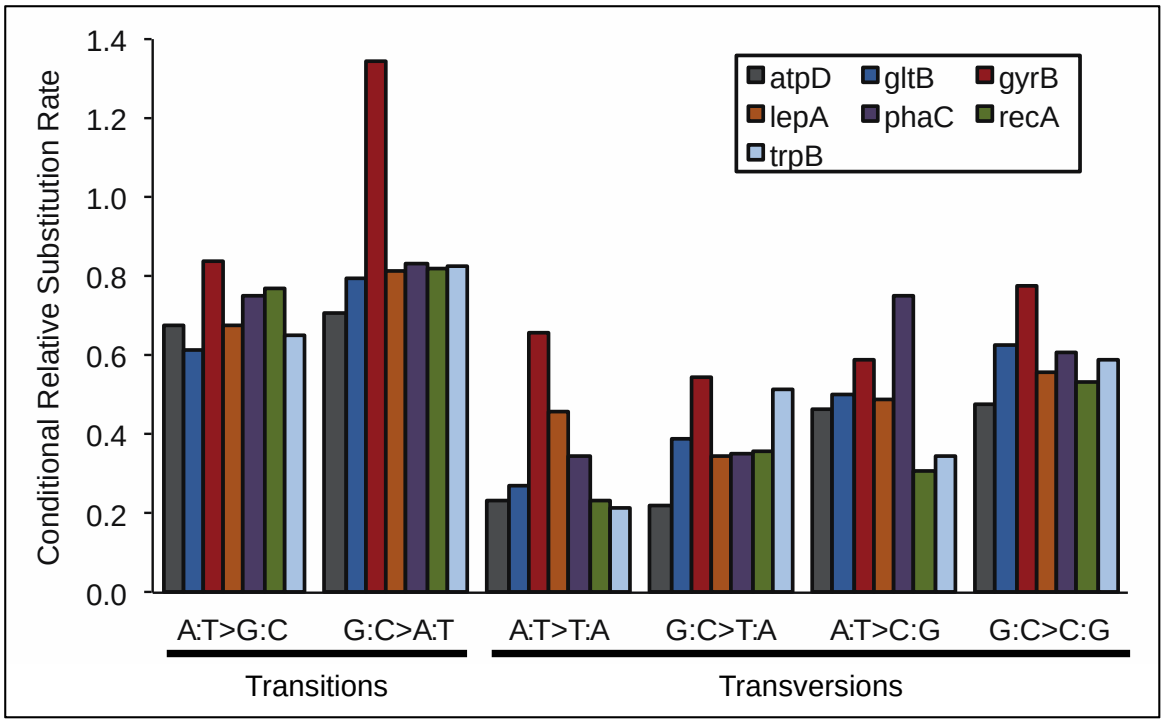

Figure S7 Relative conditional substitution rates at seven *B. cenocepacia* loci (*atpD*, *gltB*, *gyrB*, *lepA*, *phaC*, *recA*, and *trpB*). Relative conditional substitution rates are estimated by assuming that the most common nucleotide at each site is ancestral and any deviation from that nucleotide is caused by a single mutation. Substitution rates were calibrated to the nucleotide content at polymorphic sites for each gene, whereby only covered sites capable of producing a given substitution are used in the denominator of each calculation.

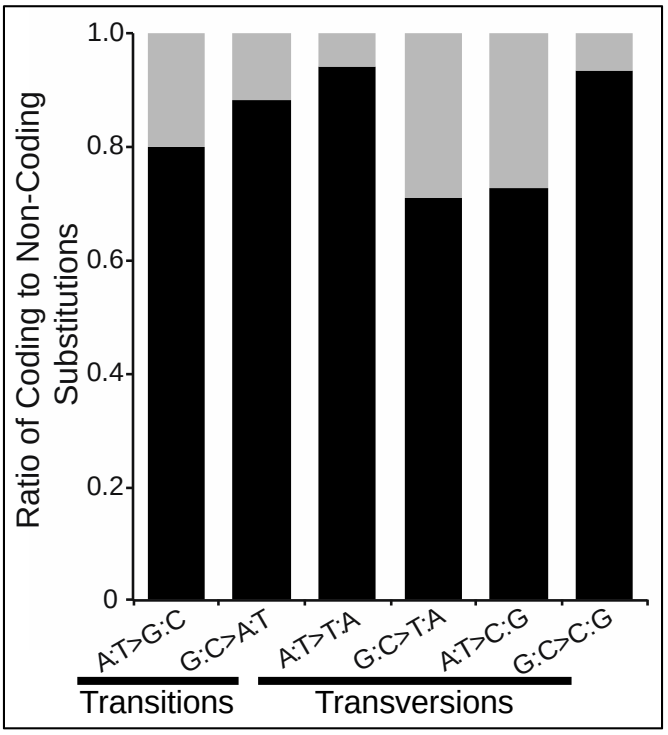

Figure S8 The ratio of coding (Black) to non-coding (Grey) substitutions for each *B. cenocepacia* substitution type.

File S1

Supplementary Tables S1 and S2

Available for download as an Excel file at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.176834/-/DC1

References

GOUT J.-F., KELLEY THOMAS W., SMITH Z., OKAMOTO K., LYNCH M., 2013 Large-scale detection of in vivo transcription errors. Proc. Natl. Acad. Sci. U. S. A. **110**: 18584–9.

MORROW J. D., COOPER V. S., 2012 Evolutionary effects of translocations in bacterial genomes. Genome Biol. Evol. **4**: 1256–1262.

RASMUSSEN T., JENSEN R. B., SKOVGAARD O., 2007 The two chromosomes of *Vibrio cholerae* are initiated at different time points in the cell cycle. Embo J. **26**: 3124–3131.

SUNG W., ACKERMAN M. S., MILLER S. F., DOAK T. G., LYNCH M., 2012 Drift-barrier hypothesis and mutation-rate evolution. Proc. Natl. Acad. Sci. U. S. A. **109**: 18488–18492.