EXTENDED PDF FORMAT
SPONSORED BY

http://discover.bio-rad.com

Adaptive Mutations in Bacteria: High Rate and Small Effects

Lília Perfeito, *et al.*
Science **317**, 813 (2007);
DOI: 10.1126/science.1142284

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 10, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/317/5839/813

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/317/5839/813/DC1

This article **cites 24 articles**, 11 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/317/5839/813#otherarticles

This article appears in the following **subject collections**:
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

Downloaded from www.sciencemag.org on August 10, 2007

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright c 2007 by the American Association for the Advancement of Science; all rights reserved. The title SCIENCE is a registered trademark of AAAS.

6. F. Watzold, K. Schwerdtner, *Biol. Conserv.* **123**, 327 (2005).
7. D. J. Rohlf, *Conserv. Biol.* **5**, 273 (1991).
8. E. Stokstad, *Science* **309**, 2150 (2005).
9. T. D. Male, M. J. Bean, *Ecol. Lett.* **8**, 986 (2005).
10. D. J. Pain *et al.*, *Anim. Conserv.* **9**, 322 (2006).
11. A. J. Cahill, J. S. Walker, S. J. Marsden, *Oryx* **40**, 161 (2006).
12. Further information on the Birds Directive, the suite of conservation measures it introduced, and a list of species classified in the analyses as Annex I species are given in the Supporting Online Material (SOM) text.
13. G. M. Tucker, M. F. Heath, *Birds in Europe: Their Conservation Status* (BirdLife Conservation Series No. 3, BirdLife International, Cambridge, 1994). The raw data underlying this assessment, and used in the present paper, were published in (*25*).
14. BirdLife International, *Birds in Europe: Population Estimates, Trends and Conservation Status* (BirdLife Conservation Series No. 12, BirdLife International, Cambridge, 2004). These data are freely available at www.birdlife.org/action/science/species/birds_in_europe/species_search.html.
15. A. Stewart-Oaten, *The Before-After/Control-Impact-Pairs Design for Environmental Impact Assessment* (Marine Review Committee, San Francisco, 1986).
16. Materials and methods are provided on *Science* Online.
17. Austria, Finland, and Sweden joined on 1 January 1995 and so were member states for 6 of the 11 years of the 1990–2000 census period.
18. P. F. Donald, F. J. Sanderson, I. J. Burfield, F. P. J. van Bommel, *Agric. Ecosyst. Environ.* **116** 189 (2006).
19. F. J. Sanderson, P. F. Donald, D. J. Pain, I. J. Burfield, F. P. J. van Bommel, *Biol. Conserv.* **131**, 93 (2006).
20. A. Balmford, P. Crane, A. Dobson, R. E. Green, G. M. Mace, *Philos. Trans. R. Soc. London Ser. B* **360** 221 (2005).
21. P. F. Donald, R. E. Green, M. F. Heath, *Proc. R. Soc. London Ser. B* **268**, 25 (2001).
22. R. D. Gregory *et al.*, *Philos. Trans. R. Soc. London Ser. B* **360**, 269 (2005).
23. H. M. Pereira, H. D. Cooper, *Trends Ecol. Evol.* **21**, 123 (2006).
24. A. Balmford *et al.*, *Science* **307**, 212 (2005).
25. BirdLife International/European Bird Census Council, *European Bird Populations: Estimates and Trends* (BirdLife Conservation Series No. 10, BirdLife International, Cambridge, 2000).
26. We thank R. B. Bradbury, D. J. Cartwright, D. Elston, A. Gammell, D. W. Gibbons, R. E. Green, D. J. Pain, W. J. Sutherland, and two anonymous referees for help and comments and the many observers across Europe who collected the data. The 1970–1990 data set was compiled by G. Tucker and M. Heath, in association with the European Bird Census Council. S.M.B. gratefully acknowledges funding from the EU, contract no. GOCECT-2003-506675.

Supporting Online Material
www.sciencemag.org/cgi/content/full/317/5839/810/DC1
Materials and Methods
SOM Text
Fig. S1
References
Appendix S1

4 June 2007; accepted 18 June 2007
10.1126/science.1146002

Adaptive Mutations in Bacteria: High Rate and Small Effects

Lília Perfeito,[1] Lisete Fernandes,[1,2] Catarina Mota,[1] Isabel Gordo[1]*

Evolution by natural selection is driven by the continuous generation of adaptive mutations. We measured the genomic mutation rate that generates beneficial mutations and their effects on fitness in *Escherichia coli* under conditions in which the effect of competition between lineages carrying different beneficial mutations is minimized. We found a rate on the order of 10^{-5} per genome per generation, which is 1000 times as high as previous estimates, and a mean selective advantage of 1%. Such a high rate of adaptive evolution has implications for the evolution of antibiotic resistance and pathogenicity.

The rate at which new mutations arise in natural populations and their fitness effects are of key importance in evolutionary genetics. Classical mutation accumulation experiments have indisputably shown that among the spontaneous mutations that affect fitness, those that cause deleterious effects are far more common than those that cause increases in fitness. Whereas there are currently several direct and indirect estimates of the deleterious mutation rate in different organisms, data are lacking for beneficial mutations (*1*). The latter are of particular interest because they constitute the driving force of adaptation and survival of populations in new environments.

Several theoretical studies have made some general predictions about the long-term process of adaptation toward an optimum (*2*, *3*). One prediction suggests that the effects of beneficial (advantageous) mutations (s_a) are exponentially distributed, in that many have very small effects and those that cause strong increments in fitness are rare (*3*). These are plausible predictions given that organisms are in general well adapted to their environments, so only small and rare changes lead to fitness increases (*4*–*11*).

The true distribution of newly arising beneficial mutations in an organism in a given environment is difficult to estimate because the probability of fixation of a beneficial mutation that increases fitness by s_a is only $2s_a$, which means that mutations of small effect are not likely to increase in frequency. This implies that the distribution of mutations that escape stochastic loss (become fixed or reach high enough frequencies to be observed) is truncated for small values (*12*, *13*). In addition, clonal interference occurs in large populations with a high beneficial mutation rate (U_a) and no recombination and will slow adaptation (compared to sexual populations of the same size) (*14*). Namely, if multiple beneficial mutations appear in different lineages, they compete with each other for fixation. This translates into an adaptation rate less than that predicted by the mutation rate and population size, and into the fixation only of mutations of large effect (*15*). Recently, there has been a considerable effort to predict the rate and distribution of beneficial mutations and the effect of clonal interference on the adaptation rate (*16*, *17*).

Current estimates for U_a fall around 10^{-9} to 10^{-8} for RNA viruses and *Escherichia coli* (*4*, *5*, *16*). A similar beneficial mutation rate was estimated for *Pseudomonas fluorescens* under adaptation to stressful conditions (*9*). A caveat for all of these estimates is that they were obtained from populations with very large effective population size (N_e) and followed adaptation to a new environment under conditions in which clonal interference had a strong effect. This led to downward biased estimates of U_a. Here, we provide estimates for the genomic mutation rate for beneficial mutations in *E. coli* that are less biased by clonal interference.

In this work, we used populations with an intermediate effective population size—big enough that genetic drift is unlikely to drive slightly deleterious mutations to a high frequency but small enough to minimize the effects of clonal interference between beneficial mutations. To estimate the beneficial mutation rate and the distribution of fitness effects of single mutations, we used a microsatellite marker system pioneered by Imhof and Schlotterer (*4*). Mutations at a microsatellite locus coded by a nonconjugative plasmid can generate neutral allelic diversity in a very short time (*4*, *18*), and selective sweeps, occurring in the bacterial genome, can be identified by following the rapid increase in the frequency of the linked microsatellite allele (*4*).

We allowed populations of *E. coli* to adapt to a given laboratory environment for 1000 generations and followed the allelic distribution of the microsatellite at periodic intervals. From this distribution, the number of mutations that escaped stochastic loss during this period was inferred for populations with a small effective size ($N_e = 2 \times 10^4$) and for populations with a very large effective size ($N_e = 10^7$). The latter allowed us to compare our estimates with those previously published (*4*, *16*, *19*).

The beneficial mutations that escape stochastic loss are expected to follow a gamma distribution with shape parameter 2 and with a mean

[1]Instituto Gulbenkian de Ciência, Rua da Quinta Grande, number 6, 2780-156 Oeiras, Portugal. [2]Escola Superior de Tecnologia da Saúde de Lisboa, Lisboa, Portugal.

*To whom correspondence should be addressed. E-mail: igordo@igc.gulbenkian.pt

Downloaded from www.sciencemag.org on August 10, 2007

equal to twice that of the distribution of the spontaneously arising mutations (*16*). Figure 1 shows the observed distributions of effects of favorable mutations segregating in the populations. In the populations with the smaller effective size ($N_e = 2 \times 10^4$), the mean value of the selection coefficient [$E(s_a)$] measured was 0.013, which is slightly smaller, although close to previous estimates (*4*, *16*). In these populations (Fig. 1A) we find that a gamma distribution with such parameterization provides a good fit to the data (Kolmogorov-Smirnov: not significant, $P = 0.6$). In Fig. 1B, we show the distribution of selective effects measured in the populations with larger effective size ($N_e = 10^7$). As expected, in these populations, the effect of clonal interference was clearly observed in the distributions of microsatellite allelic variation [for an example, see fig. S2 (*19*)]. As predicted theoretically (*15*), the effect of interference between clones carrying different beneficial mutations is reflected in an increased value of the mean selective effect of mutations segregating in the population [$E(s_a) = 0.023$, as shown in Fig. 1B]. This is because many newly arising beneficial mutations of small effect are lost in competition with mutations of larger effect.

To measure the rate of spontaneous beneficial mutations, we quantified the total number of mutations that escaped stochastic loss in all the populations with the same effective size during the course of the experiment. We observed 75 such events in the populations with $N_e = 2 \times 10^4$ and 87 in the populations with $N_e = 10^7$. Assuming that the effect of clonal interference is negligible, in the populations with larger N_e we would infer a mutation rate of 2×10^{-8} beneficial mutations per genome, per generation (*20*, *21*). This value is close to those previously measured for this species with the use of populations with similar effective sizes (*4*, *16*). However, with such a large N_e, the effect of clonal interference is very important and leads to an extreme underestimation of the true value of U_a. Indeed, in the populations of smaller effective size, our estimate of the mutation rate was 1000 times as high: $U_a = 2 \times 10^{-5}$ beneficial mutations per genome, per generation (*20*, *21*). Given that clonal interference is much weaker in these populations, we take this value to be a much more accurate measure of the real U_a.

To complement these results, we measured the mean fitness of each evolved population relative to the ancestral one. Mean fitness of an evolved population was assessed by its competitive ability against a reference strain (*19*). As expected in view of the results obtained above, there was an overall increase in fitness in all populations after 1000 generations of adaptation. In the populations with the smaller N_e, this increase was about 17%, which, as expected, was smaller than the one observed for the populations with a larger N_e (overall mean increase in fitness of 40%). We then asked if our estimates of U_a could explain such increments in fitness (*22*). To do this, we compared the results of Monte-Carlo simulations of adaptive evolution, assuming several different values of U_a and $E(s_a)$ with those obtained in the experiments (*19*). In all the simulations, we assumed that the distribution of incoming beneficial mutations is exponential and that mutations interact in a multiplicative way (*2*, *23*). Different combinations of U_a and $E(s_a)$ were consistent with the fitness increase in populations of a given effective size, but the set of parameters that more closely matched the combined data of both population sizes was U_a between 10^{-5} and 10^{-4} and $E(s_a)$ between 1 to 2% (Fig. 2). These parameters agree with the estimates obtained from the microsatellite allelic distribution (small effective size populations in Fig. 1, in which the measured mutation rate was 2×10^{-5}). It is also clear that a mutation rate of about 10^{-9} or 10^{-8} [as inferred in other experiments (*4*, *16*)] cannot explain the fitness increases observed.

Our results show that the mutation rate to new beneficial alleles is 1000 times as high as previously inferred in the same bacterial species (*4*, *16*). The difference in results can be explained by the differences in the effective population size analyzed. If only very large effective sizes are analyzed, and the effect of clonal interference is not accounted for, then our estimates for U_a and $E(s_a)$ for the populations with $N_e = 10^7$ are similar to those previously obtained (Fig. 1B). However, if these estimates were close to the true values, then we would not expect to see the sweeps of beneficial mutations in the populations with lower N_e that we observed (Fig. 1A). Hence, neglecting the effect of clonal interference underestimates the value of U_a. In addition, we showed that clonal interference changes the distribution of segregating mutations: When comparing the distribution of beneficial mutations for the populations with high N_e (strong effect of clonal interference) with that for populations with low N_e (Fig. 1, A and B), a significant difference was observed (Kolmogorov-Smirnov: $P = 0.001$). As predicted theoretically (*15*, *24*), we observed a distribution with a higher mean selection coefficient when the effect of clonal interference was stronger. In the limiting case where the supply of new beneficial mutations per generation (N_eU_a) is very high, the speed of adaptation will no longer depend on N_eU_a but on the mutations of largest effect available, be-

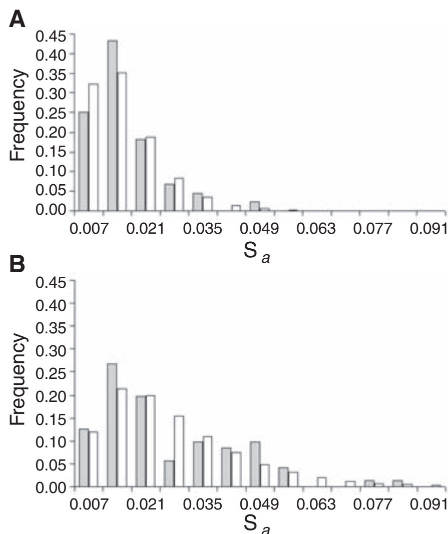

Fig. 1. Distribution of fitness effects of beneficial mutations that escaped stochastic loss, measured in populations of $N_e = 2 \times 10^4$ (**A**) and $N_e = 10^7$ (**B**). The gray bars show the distribution of effects of beneficial mutations inferred in the experimental populations and the white bars correspond to a gamma distribution with shape 2 and scale parameters 158 (A) and 85 (B). Both distributions are supported by the data [Kolmogorov-Smirnov: $P = 0.6$ in (A) and $P = 0.5$ in (B); not significant].

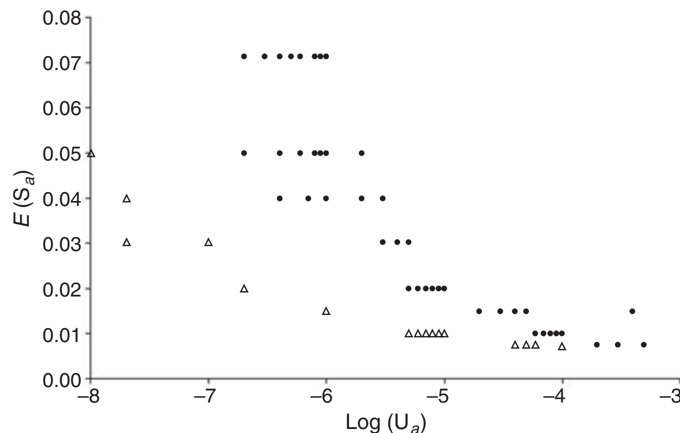

Fig. 2. Mutation rates (U_a) and mean effect of beneficial mutations [$E(s_a)$], used as parameters in Monte Carlo simulations (*19*), which produced mean fitness increases consistent with those observed in the evolved populations (difference was not significant; Student's *t* test $P > 0.05$). The circles show the parameter values consistent with the mean fitness observed in the populations of $N_e = 2 \times 10^4$ and the triangles in the populations of $N_e = 10^7$.

Downloaded from www.sciencemag.org on August 10, 2007

cause these are the only mutations that will fix. This might help explain why similar beneficial mutation rates are estimated in very diverse organisms under very diverse environments. These estimates are obtained in populations with very large effective sizes (*4*, *5*, *9*, *16*), which are likely to produce strong underestimations of U_a.

It is plausible that, in natural habitats, population sizes will be large. If the effective size of a bacterial species is much higher than 10^4 (*25*), then our results imply that clonal interference plays a major role in limiting the adaptation of these asexual organisms. As such, if there is a chance for recombination, clonal interference will be much lower and organisms will adapt faster. This has been predicted theoretically (*14*), although the empirical evidence is still very preliminary (*26*, *27*). Given our results, we anticipate that clonal interference is important in maintaining sexual reproduction in eukaryotes. Notably, mutation accumulation experiments in *Saccharomyces cerevisiae* and *Arabidopsis thaliana* have detected a significantly large number of mutants with increased fitness (*28*, *29*).

Given the estimates for the overall mutation rate in *E. coli* (*30*) and its genomic deleterious mutation rate (*1*), our estimate of U_a implies that 1 in 150 newly arising mutations is beneficial and that 1 in 10 fitness-affecting mutations increases the fitness of the individual carrying it. Hence, an enterobacterium has an enormous potential for adaptation and may help explain how antibiotic resistance and virulence evolve so quickly.

References and Notes

1. T. Bataillon, *Heredity* **84**, 497 (2000).
2. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, UK, 1930).
3. H. A. Orr, *Nat. Rev. Genet.* **6**, 119 (2005).
4. M. Imhof, C. Schlotterer, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1113 (2001).
5. R. Miralles, P. J. Gerrish, A. Moya, S. F. Elena, *Science* **285**, 1745 (1999).
6. C. L. Burch, L. Chao, *Genetics* **151**, 921 (1999).
7. C. Zeyl, *Genetics* **169**, 1825 (2005).
8. S. Estes, M. Lynch, *Evol. Int. J. Org. Evol.* **57**, 1022 (2003).
9. R. D. Barrett, R. C. MacLean, G. Bell, *Biol. Lett.* **2**, 236 (2006).
10. D. R. Rokyta, P. Joyce, S. Caudle, H. Wichman, *Nat. Genet.* **37**, 441 (2005).
11. R. Kassen, T. Bataillon, *Nat. Genet.* **38**, 484 (2006).
12. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
13. A. J. Betancourt, J. P. Bollback, *Curr. Opin. Genet. Dev.* **16**, 618 (2006).
14. W. G. Hill, A. Robertson, *Genet. Res.* **8**, 269 (1966).
15. P. Gerrish, R. Lenski, *Genetica* **102–103**, 127 (1998).
16. D. E. Rozen, J. A. de Visser, P. Gerrish, *Curr. Biol.* **12**, 1040 (2002).
17. M. Hegreness, N. Shoresh, D. Hartl, R. Kishony, *Science* **311**, 1615 (2006).
18. C. Schlotterer, M. Imhof, H. Wang, V. Nolte, B. Harr, *J. Evol. Biol.* **19**, 1671 (2006).
19. Materials and methods are available as supporting material on *Science* Online.
20. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).
21. $U_a = k/(N_e \cdot 2E(s_a) \cdot T)$, where k is the number of observed mutations, T is the number of generations and $E(s_a)$ is the mean selection coefficient. This assumes that there is no clonal interference. If its effect is major, the value of U_a will be greatly underestimated. Also, small effect mutations are likely to be missed because the time it takes for a mutation to increase in frequency is $\propto 1/s_a$.
22. The expected fitness increase over 1000 generations without clonal interference is $N_e \cdot U_a \cdot 2\ E(s_a) \cdot E(s_a) \cdot 1000 = 0.14$ in the small size populations.
23. H. A. Orr, *Genetics* **163**, 1519 (2003).
24. R. D. H. Barrett, S. P. Otto, L. K. M′gonigle, *Genetics* **174**, 2071 (2006).
25. J. Charlesworth, A. Eyre-Walker, *Mol. Biol. Evol.* **23**, 1348 (2006).
26. B. Grimberg, C. Zeyl, *Evol. Int. J. Org. Evol.* **59**, 431 (2005).
27. N. Colegrave, *Nature* **420**, 664 (2002).
28. S. B. Joseph, D. W. Hall, *Genetics* **168**, 1817 (2004).
29. R. G. Shaw, D. Byers, E. Darmo, *Genetics* **155**, 369 (2000).
30. J. W. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7160 (1991).

31. This work was supported by POCTI/BSE/46856/ 2002 and SFRH/BD/18161/2004 through Fundação para a Ciência e Tecnologia. We thank M. Imhof for help in the initial setup of this work and R. Azevedo, D. Bachtrog, A. Coutinho, F. Dionisio, K. Xavier, and anonymous reviewers for comments on the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5839/813/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
Table S1
References

8 March 2007; accepted 2 July 2007
10.1126/science.1142284

Downloaded from www.sciencemag.org on August 10, 2007

Divergence of Transcription Factor Binding Sites Across Related Yeast Species

Anthony R. Borneman,[1]* Tara A. Gianoulis,[2] Zhengdong D. Zhang,[3] Haiyuan Yu,[3] Joel Rozowsky,[3] Michael R. Seringhaus,[3] Lu Yong Wang,[4] Mark Gerstein,[2,3,5] Michael Snyder[1,2,3]†

Characterization of interspecies differences in gene regulation is crucial for understanding the molecular basis of both phenotypic diversity and evolution. By means of chromatin immunoprecipitation and DNA microarray analysis, the divergence in the binding sites of the pseudohyphal regulators Ste12 and Tec1 was determined in the yeasts *Saccharomyces cerevisiae*, *S. mikatae*, and *S. bayanus* under pseudohyphal conditions. We have shown that most of these sites have diverged across these species, far exceeding the interspecies variation in orthologous genes. A group of Ste12 targets was shown to be bound only in *S. mikatae* and *S. bayanus* under pseudohyphal conditions. Many of these genes are targets of Ste12 during mating in *S. cerevisiae*, indicating that specialization between the two pathways has occurred in this species. Transcription factor binding sites have therefore diverged substantially faster than ortholog content. Thus, gene regulation resulting from transcription factor binding is likely to be a major cause of divergence between related species.

Differences in related individuals are generally attributed to changes in gene composition and/or alterations in their regulation. Previous efforts to examine divergence of regulatory information have relied on the analysis of conserved sequences in putative promoter regions (*1*, *2*). However, these approaches are limited because transcription factor (TF) binding sites are often short and degenerate, making their computational detection difficult (*3*). In addition, requiring the conservation of motifs across species precludes the detection of sequences that are evolutionarily divergent.

The detection of binding sites with chromatin immunoprecipitation and microarray (ChIP-chip) analysis (*4*, *5*) offers the ability to globally map TF binding locations experimentally rather than computationally. For species such as yeasts, where genome sequences of numerous related species are available (*6*), this approach can allow for the evolutionary comparison of binding sites of conserved TFs across species.

We have used this approach to investigate evolutionary divergence in the targets of two developmental regulators in the *Saccharomyces* sensu stricto yeasts *S. cerevisae*, *S. mikatae*, and *S. bayanus*. In *S. cerevisiae* diploids, Ste12 and Tec1 act cooperatively to regulate genes during pseudohyphal development (*7–9*), whereas in haploid cells, Ste12 regulates mating genes (*10*). The binding sites of Ste12 and Tec1 were mapped in all three species under low-nitrogen

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA. [2]Program in Computational Biology, Yale University, New Haven, CT 06511, USA. [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA. [4]Integrated Data Systems Department, Siemens Corporate Research, Princeton, NJ 08540, USA. [5]Department of Computer Science, Yale University, New Haven, CT 06511, USA.

*Present address: Australian Wine Research Institute, Glen Osmond, Adelaide, SA 5064, Australia.

†To whom correspondence should be addressed. E-mail: michael.snyder@yale.edu