

The following resources related to this article are available online at www.sciencemag.org (this information is current as of November 12, 2009):

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/306/5698/978a>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/306/5698/978a/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/306/5698/978a#related-content>

This article **cites 28 articles**, 15 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/306/5698/978a#otherarticles>

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/306/5698/978a#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Technical Comments

http://www.sciencemag.org/cgi/collection/tech_comment

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>

Comment on “The Origins of Genome Complexity”

Lynch and Conery (*1*) argued that genome complexity reflects a history of genetic drift caused by small genetic population size (N_e), which in turn enables the spread of mildly deleterious selfish elements and duplications. Under this argument, large organisms, which tend to have small N_e , also exhibit larger (complex) genomes, and the small size of bacterial genomes reflects their position at the extreme large end of the range for N_e . Genome size can vary among bacteria by at least an order of magnitude, but the fraction of noncoding DNA is consistently low (~15%). Consequently, larger bacterial genomes are expected to result from more selection, because they maintain more functional DNA. Indeed, large genome size in bacteria commonly has been considered an adaptation to changing environments (2–6). Furthermore, small N_e in symbiotic bacteria appears to result in reduced genomes through gene loss (7, 8). Thus, the relation between N_e and genome size in bacteria is, if anything, expected to be the opposite of that proposed by Lynch and Conery (*1*).

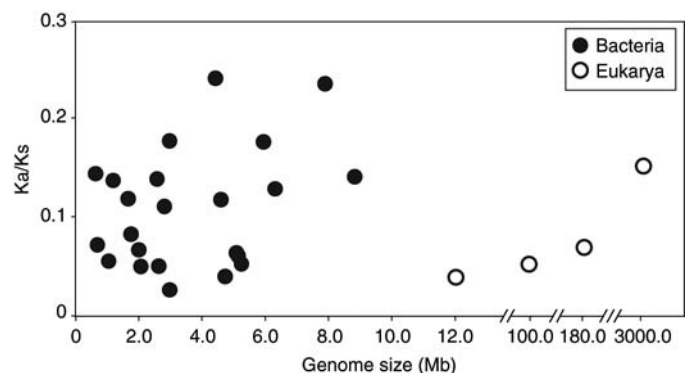
The claimed relationship between N_e and genome size is based on inferring $N_e u$ (where u is the mutation rate per nucleotide) from estimates of polymorphism using sequence divergences among isolates (*1*). This approach, however, has several drawbacks, which are especially severe in bacteria. One concern is the assumption of similar mutation rates across bacteria, which is required if estimates of $N_e u$ are interpreted as directly reflecting variation in N_e . Lynch and Conery (*1*) justify this assumption by citing a survey (*9*) that included only a single bacterial strain; other evidence indicates that bacterial lineages vary substantially in mutation rates (*10*)—as would be expected, because their genomes differ markedly in content of genes known to impact mutation. Thus, polymorphism levels reflect mutational input as well as N_e .

Even more problematic is the assumption that taxonomic names correspond to cohesive species, which is implicit in the use by Lynch and Conery (*1*) of polymorphism data to estimate N_e . Although this assumption can be considered approximately true for most sexual eukaryotic species, it is often far from valid for bacterial strains grouped under a particular species name. These usually constitute phylogenetically and ecologically distinct clusters separated by niche and genetic boundaries; thus, genetic drift does

not act within the named “species” as a whole but within smaller, divergent subunits. The position of bacteria at the extreme high end of the range of $N_e u$ values cited by Lynch and Conery (*1*) is based in large part on lumping independently evolving strains for calculations of polymorphism. For example, the highest $N_e u$ value listed by Lynch and Conery, for the free-living marine cyanobacterial genus *Prochlorococcus*, is derived from divergent species with different ecological niches (*11*), dramatically different gene inventories reflecting habitat differences, and extreme differences in genomic base composition (*12*). This use of divergence to calculate N_e is comparable to using the divergence of cat, dog, human, and mouse to calculate N_e for the species “mammal,” an exercise that would give an enormous population size.

Similarly, Lynch and Conery (*1*) treat *Salmonella enterica* as a cohesive species, yet it consists of distinct, highly clonal lineages living in different vertebrate hosts (*13, 14*). Lateral gene transfer, which is relatively common in bacteria [and reported for the *S. enterica* data set (*13*)] will also radically affect the calculated degree of polymorphism if undetected (or ignored). The complexities of relating polymorphism to N_e in bacteria are widely recognized, having been noted in early enzyme electrophoresis studies, which revealed that polymorphism levels were unexpectedly low in view of the census population sizes of bacterial species (*15*).

Fig. 1. Average K_a/K_s (*26*) versus genome size for 23 pairs of related bacteria (filled circles) and several pairs of eukaryotes (open circles). Pairs span the range of bacterial genome size and include all available genome pairs for which K_s is not saturated (values < 2). Because the value of K_a depends on gene function, we selected the single-copy orthologs (defined as reciprocal unique BLAST hits) present in at least 15 of the 23 genome pairs, to ensure that comparisons among genomes encompass similar gene sets. Similar lack of relation was observed using the complete set of orthologs for each genome pair and using different methods of calculating K_a/K_s (*26–28*). For comparison, K_a/K_s and genome size values for four pairs of eukaryotic genomes (in order of increasing genome size: yeasts, nematodes, flies, and mammals) are presented. Data and taxon names for bacteria are presented in the Supporting Online Material.



Even if issues of population subdivision, recombination, and variation in mutational input could be resolved, estimates of N_e based on polymorphism would reflect the very recent evolutionary past rather than the longer period over which genome features evolved. Major reductions in polymorphism are observed for recently derived pathogens relative to ecologically generalized parental species [see, for example, (*16*)], but genomic traits, including genome size, are little altered.

A more reliable index of genetic drift over evolutionary time is the ratio of K_a (nonsynonymous substitutions per site) to K_s (synonymous substitutions per site) for a large set of genes, based on comparisons of related species. The K_a/K_s ratio, which is almost always less than one, is widely used as an indicator of the extent of purifying selection acting to conserve coding sequences [see, for example, (*17*)]. Although K_a/K_s can be elevated in particular genes as a result of positive selection for amino acid changes or selection conserving codon choice, a pattern of genome-wide differences in K_a/K_s indicates persistent levels of differences in N_e . Therefore, the proposal of Lynch and Conery (*1*) predicts a positive correlation between K_a/K_s and genome size and, thus, higher K_a/K_s ratios in eukaryotes. Another advantage of this approach is that it is not invalidated by variation in mutation rate among lineages.

No clear relation is observed between average K_a/K_s and genome size across bacteria, whether all genes or a common set of genes are used (Fig. 1). Because K_s values can be depressed by selection favoring certain codons, we used different approaches to define a genome-wide K_a/K_s , but found no relation with genome size (Fig. 1 and Table S1). Furthermore, genomes known to have

strong codon biases, such as *E. coli*-*S. enterica* or *Bacillus* species, show low K_a/K_s , and cases known to have low codon bias, such as *Buchnera* (18), have high K_a/K_s , which indicates that purifying selection on codon use has less effect on K_a/K_s than does selection on amino acid residue.

As is evident in Fig. 1 [and as noted previously based on more limited data (19)], average K_a/K_s can be very high in certain bacteria, including not only symbiotic lineages but also others, such as several groups of soil-dwelling bacteria. At least for symbionts, high K_a/K_s corresponds to very low levels of nucleotide polymorphism (20, 21), implicating small N_e as the likely cause. The conventional view—that large N_e is typical of bacteria—was initially established on the basis of the limited early sequence data available for *E. coli* and *S. enterica*, which, as it turns out, show exceptionally low K_a/K_s (Supporting Online Material). Genomic data and polymorphism studies now suggest that this early view was simplistic and that N_e varies widely across bacterial groups.

For comparison, we have included K_a/K_s values calculated for eukaryotes [derived from genome-wide comparisons of orthologs for fungi, mammals, nematodes, and flies (Fig. 1)]. These fall within the range of values found in bacteria, which suggests that there is no consistent difference in the efficacy of purifying selection between prokaryotes and eukaryotes, despite the gap between their genome sizes.

As Lynch and Conery (1) argue, N_e probably does affect the ability to eliminate useless or deleterious DNA. However, their model supposes that lack of effective selection will lead to genome growth because of the dynamics of selfish elements. That

supposition may be true for eukaryotes but does not hold for bacteria, which display a mutational bias in nature favoring deletion and preventing major accumulation of so-called junk DNA (22–24). This deletional bias may itself reflect vulnerability to selfish DNA (23), resulting from distinctive biological features of bacteria, including lack of a nuclear envelope or meiotic sex. Although bacteria undergoing severe genetic drift may experience outbreaks of selfish elements, the principal outcome of reduced population size is the random inactivation and deletion of genes participating in basic cellular processes (7), ultimately resulting in an overall genome reduction [see, for example, (25)]. Hence, the model of Lynch and Conery does not explain the differences in genome sizes among bacteria. Their conclusions that N_e is consistently enormous in bacteria and that the resulting low levels of genetic drift underlie the lack of genomic and phenotypic complexity in these organisms are not warranted.

Vincent Daubin

*Department of Biochemistry and
Molecular Biophysics
University of Arizona
daubin@email.arizona.edu*

Nancy A. Moran

*Department of Ecology and
Evolutionary Biology
University of Arizona
Tucson, AZ 85721, USA
nmoran@email.arizona.edu*

References and Notes

1. M. Lynch, J. S. Conery, *Science* **302**, 1401 (2003).
2. K. T. Konstantinidis, J. M. Tiedje, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3160 (2004).
3. K. E. Nelson *et al.*, *Environ. Microbiol.* **4**, 799 (2002).

4. U. Dobrindt, J. Hacker, *Curr. Opin. Microbiol.* **4**, 550 (2001).
5. H. Ikeda *et al.*, *Nature Biotechnol.* **21**, 526 (2003).
6. S. D. Bentley *et al.*, *Nature* **417**, 141 (2002).
7. N. A. Moran, *Cell* **108**, 583 (2002).
8. S. G. Andersson, C. G. Kurland, *Trends Microbiol.* **6**, 263 (1998).
9. J. W. Drake, B. Charlesworth, D. Charlesworth, J. F. Crow, *Genetics* **148**, 1667 (1998).
10. H. Ochman, *Mol. Biol. Evol.* **20**, 2091 (2003).
11. E. Urbach, D. J. Scanlan, D. L. Distel, J. B. Waterbury, S. W. Chisholm, *J. Mol. Evol.* **46**, 188 (1998).
12. G. Roca *et al.*, *Nature* **424**, 1042 (2003).
13. F. S. Wang, T. S. Whittam, R. K. Selander, *J. Bacteriol.* **179**, 6551 (1997).
14. E. F. Boyd, F. S. Wang, T. S. Whittam, R. K. Selander, *Appl. Environ. Microbiol.* **62**, 804 (1996).
15. T. S. Whittam, H. Ochman, R. K. Selander, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1751 (1983).
16. M. Achtman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14043 (1999).
17. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
18. C. Rispe, F. Delmotte, R. C. van Ham, A. Moya, *Genome Res.* **14**, 44 (2004).
19. H. Ochman, S. Elwyn, N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **96**: 12638 (1999).
20. P. Abbot, N. A. Moran, *Mol. Ecol.* **11**, 2649 (2002).
21. D. J. Funk, J. J. Wernegreen, N. A. Moran, *Genetics* **157**, 477 (2001).
22. J. O. Andersson, S. G. Andersson, *Mol. Biol. Evol.* **18**, 829 (2001).
23. J. G. Lawrence, R. W. Hendrix, S. Casjens, *Trends Microbiol.* **9**, 535 (2001).
24. A. Mira, H. Ochman, N. A. Moran, *Trends Genet.* **17**, 589 (2001).
25. J. Parkhill *et al.*, *Nature Genet.* **35**, 32 (2003).
26. J. M. Comeron, *J. Mol. Evol.* **41**, 1152 (1995).
27. M. Nei, T. Gojobori, *Mol. Biol. Evol.* **3**, 418 (1986).
28. Z. Yang, *Comput. Applic. Biosci.* **13**, 555 (1997).
29. We acknowledge A. Rokas, C. Seoighe, and A. Cutter for sharing their sets of eukaryotic orthologous genes in closely related eukaryote species, and H. Ochman for comment on this manuscript. Funding was from Department of Energy award DEFG0301ER63147 to H. Ochman and NSF award EF-0313737 to N.A.M.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5698/978a/DC1
SOM Text
Table S1

29 March 2004; accepted 5 August 2004