

Reply to ‘The population genetics of pangenomes’

McInerney et al. reply — Shapiro has raised a number of interesting and important points about pangenome evolution, focusing mainly on the rate of gene acquisition¹. We agree that there are several aspects to the population genetics of pangenomes that need to be considered. It is true that prokaryotes that engage in horizontal gene transfer are difficult to treat as a ‘true’ species and in particular, a gene might find itself in organisms with different effective population sizes at different times in its history. However, there is a lot of evidence that prokaryote genomes and their constituent genes obey expectations from population genetic theory.

At least five factors correlate — organismal lifestyle (intracellular parasites at one extreme, global free-living organisms at the other) correlates with pangenome size (minimal accessory genome at one extreme and minimal core genome at the other), which correlates with rate of horizontal gene transfer (low to high), which correlates with size of effective population size (N_e) (from small to large), which correlates with efficiency of natural selection. It is arguable that N_e drives all these correlations. In other words, large N_e increases selection efficiency. This means that as N_e increases, smaller and smaller fitness effects are ‘seen’ by selection. In free-living prokaryotes, it is likely that fitness differences of one-tenmillionth or smaller are seen by selection².

Selection efficiency can be seen in synonymous codon changes in highly expressed genes in species with large N_e , which tend to be under selection for optimal codon usage³. In contrast, intracellular parasites are less likely to exhibit signatures of synonymous codon-usage optimization than free-living prokaryotes⁴ and a greater portion of their genome is influenced by drift. Effects that are clearly seen in species with large N_e are completely missing in species with small N_e ⁵.

The contrast between prokaryotic genomes and mammalian genomes is

instructive. Mammalian genomes are replete with junk (nearly neutral) DNA⁶ and this is a consequence of N_e . In an apparent paradox, mammals with small N_e tend to accumulate large genomes because selection against junk DNA is inefficient. However, in prokaryotes it is species with large N_e that tend to have larger genomes and pangenomes. Unless we propose that organisms with larger N_e tend to pick up genes with such a small effect on fitness that selection is ineffective, then a neutral model does not seem to fit the data.

Prokaryotic genomes typically consist of 3–3.5% pseudogenes⁷. These are likely to be effectively neutral or slightly deleterious. However, 96.5–97% of prokaryotic genes are fully intact, and usually functional, as evidenced by gene expression and ratios of non-synonymous to synonymous substitutions that are almost always far less than one⁸. Comparative genomics show us that pseudogenes do not remain in a genome for long after they have arisen — selection is usually efficient enough to remove them⁹. In contrast, a typical *Escherichia coli* genome, rather than the usual 3–3.5% (~150) pseudogenes, consists of more than 50% (~2,250) accessory genes.

In truth, we do not have good estimates for the fitness effects of many of the 85,000 accessory genes found so far across *E. coli*, and this is clearly a gap in our knowledge. However, for a nearly neutral explanation to work, we might expect to see far more pseudogenes or else fewer accessory genes.

We note several known unknowns concerning prokaryotic genomes. Firstly, for a wide range of organisms and a wide range of genes, we do not know what the fitness effects of alternative synonymous codons are. We also do not have precise measurements of the fitness effects of accessory genes. We now have some data on the dynamics of accessory gene accumulation, with some evidence showing that an immediate short-term reduction in fitness is rapidly overcome by compensatory mutations¹⁰. However, this is almost

certainly observed only because of long-term selective benefits to acquisition and maintenance of the plasmid.

The problem for any explanation of how pangenomes arise has always been that efficient selection should result in selective sweeps and a genome that is relatively homogeneous for a species. Here, we have the paradox that large N_e and efficient selection both correlate with large pangenome size. A slightly more complex model that allows for migration to new niches solves this problem¹¹. □

James O. McInerney^{1*}, Alan McNally² and Mary J. O’Connell³

¹Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. ²Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. ³Computational and Molecular Evolutionary Biology Group, School of Biology, Faculty of Biological Sciences, The University of Leeds, Leeds, UK.

*e-mail: james.mcinerney@manchester.ac.uk

Published online: 24 November 2017
<https://doi.org/10.1038/s41564-017-0068-4>

References

1. Charlesworth, J. & Eyre-Walker, A. *Mol. Biol. Evol.* **23**, 1348–1356 (2006).
2. Sharp, P. M. et al. *Biochem. Soc. T.* **21**, 835–841 (1993).
3. McInerney, J. O. *Microb. Comp. Genomics* **2**, 89–97 (1997).
4. Sung, W. et al. *Proc. Natl Acad. Sci. USA* **109**, 18488–18492 (2012).
5. Graur, D. et al. *Genome Biol. Evol.* **5**, 578–590 (2013).
6. Rogozin, I. B. et al. *Nucleic Acids Res.* **30**, 4264–4271 (2002).
7. Novichkov, P. S. et al. *J. Bacteriol.* **191**, 65–73 (2009).
8. Kuo, C. H. & Ochman, H. *PLoS Genet.* **6**, e1001050 (2010).
9. McNally, A. et al. *PLoS Genet.* **12**, e1006280 (2016).
10. McInerney, J. O., McNally, A. & O’Connell, M. J. *Nat Microbiol.* **2**, 17040 (2017).
11. Shapiro, B. J. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-017-0066-6> (2017).

Competing interests

The authors declare no competing financial interests.