CellPress

Opinion

Rates of Lateral Gene Transfer in Prokaryotes: High but Why?

Michiel Vos,[1,*] Matthijn C. Hesselman,[2] Tim A. te Beek,[3] Mark W.J. van Passel,[2,4] and Adam Eyre-Walker[5]

Lateral gene transfer is of fundamental importance to the evolution of prokaryote genomes and has important practical consequences, as evidenced by the rapid dissemination of antibiotic resistance and virulence determinants. Relatively little effort has so far been devoted to explicitly quantifying the rate at which accessory genes are taken up and lost, but it is possible that the combined rate of lateral gene transfer and gene loss is higher than that of point mutation. What evolutionary forces underlie the rate of lateral gene transfer are not well understood. We here use theory developed to explain the evolution of mutation rates to address this question and explore its consequences for the study of prokaryote evolution.

Trends

The rate of lateral gene transfer (LGT) in prokaryotes can be high, and is likely to have profound consequences for genome evolution and adaptation. However, the selective forces underlying LGT rates have remained relatively unexplored. Insights from theory developed to explain the evolution of mutation rates can be applied to the evolution of LGT rates.

We identify two main scenarios that could explain the high observed rates of LGT. In the first scenario, LGT rate is optimal while in the second scenario the rate of LGT is suboptimal.

The diverse mechanisms and fitness effects of LGT events make it challenging to address evolutionary costs and benefits. A combination of population genomic and experimental approaches will be needed to establish LGT rates and the distribution of fitness effects of LGT events.

The Fluidity of Prokaryote Genomes

It has long been recognized that even closely related strains of bacteria and archaea can greatly differ in gene content [1–3]. The rate and promiscuity with which genes change residence in prokaryote genomes is of such magnitude that many have rejected the concept of prokaryote species (e.g., [4]) or the possibility of reconstructing a bifurcating tree of prokaryote life (e.g., [5]). The realization of the high rates of gene content turnover has led to the paradigm of a core genome of genes present in all members of a taxonomic group and an accessory genome present in only a subset of members (with the total complement of genes in a taxon termed the pan genome [6]). The extent to which related genomes differ in gene content varies for different species, with some having a relatively 'closed' genome (i.e., a core genome that is large compared to the accessory genome) and some species an 'open' or 'flexible' genome (i.e., a relatively small core genome and a large pan genome) [7].

A diverse set of mechanisms underlies the evolution of gene content but they can be grouped into three main classes: gene loss, gene gain through duplication (parology), and gene gain though lateral gene transfer (LGT, xenology) [8,9]. Knowledge on the impact of LGT on bacterial evolution continues to rapidly expand (e.g., [10–13]). In many prokaryotes, LGT is known to be a more important process for determining gene gain than is gene duplication [14–16]. Gene loss can be due to two processes: mutational deletion and the lateral transfer of gene absence by recombination between a strain with and without the gene (likely helped by the presence of homologous flanking sequence [17]), although it is not well understood which of the two mechanisms is most important.

The rate at which genomes accumulate random variation, due to point mutation or wholesale changes in gene content, dictates the speed and mode of evolution. Both mutational changes and changes in gene content provide the raw material for selection to act on, but the rate of each process can be modified through selection too. Apart from the fitness effects of individual LGT events, it is therefore crucial to understand how second-order selection acts on the rate of change itself. In this paper we provide a short summary of the relatively few studies detailing the

[1]European Centre for Environment and Human Health, University of Exeter Medical School, University of Exeter, Penryn, UK
[2]Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands
[3]Netherlands Bioinformatics Centre, Nijmegen, The Netherlands
[4]Centre for Zoonoses and Environmental Microbiology, Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
[5]School of Life Sciences, University of Sussex, Brighton, UK

*Correspondence: m.vos@exeter.ac.uk (M. Vos).

 http://dx.doi.org/10.1016/j.tim.2015.07.006

rate and fate of lateral gene transfers. The main part of this paper then discusses how models developed to explain the evolution of the mutation rate can be applied to the evolution of LGT rate, which is likely to be high in many species. Specifically, we discuss whether selection drives optimal rates of LGT or whether selection for lower rates is constrained. Finally, we highlight some of the implications our findings have for our understanding of evolutionary microbiology and identify areas for future research.

Quantifying the Rate and Fate of Changes in Gene Content

A plethora of comparative genomics studies have demonstrated the rate of prokaryote gene content turnover to be very high. However, systematic attempts to estimate how fast gene turnover occurs over evolutionary time, expressed per genome per generation or relative to point mutation, have been rare. To our knowledge, the first study to explicitly address this question was by Hao and Golding [18] who used a maximum likelihood method to quantify both insertion and deletion rates of genes relative to mutation in a set of *Bacillus cereus* group genomes. According to their published estimates, genes were gained and lost at a rate approximately 4.4 times the rate of nucleotide substitution per site [18]. Similar estimates were found in *Streptococcus* [19] and *Corynebacterium* [20] genomes using the same methodology. A recent estimate of LGT rate in *Pseudomonas syringae* based on stochastic mapping methodology (after corrections necessary for working with genomes that are not sequenced to complete closure) was found to be four orders of magnitude higher than the estimate for *B. cereus* [21]. Individual *P. syringae* lineages could be shown to have acquired thousands of genes in the same period in which a 1% amino acid divergence accrued in the core genome. The rate also appears to be very high in *Escherichia coli*, where strains that have almost no single nucleotide changes in their genes they share can differ substantially in gene content [22].

These and other studies (e.g., [23,24]) have also demonstrated that most gene content changes are short-lived. In most studies this is apparent from a decrease in the rate of gene content change relative to the rate of nucleotide substitution in the deeper branches of the phylogeny compared to the tips. The same pattern is most elegantly demonstrated in the *P. syringae* and *E. coli* studies [21,22] in which the vast majority of individual gene gains are mapped to a single strain. There are three potential explanations for why most gene content changes are transient. First, the vast majority of gene content changes might be neutral (i.e., have no selective effects) and, like mutations, are lost from the population by random **genetic drift** (see Glossary). This seems an unlikely explanation, because under this model the relative rates of gene content change and nucleotide substitution would remain constant across different phylogenetic depths. Second, most gene content changes might be deleterious [6,8,18,22]. There are a variety of reasons to expect that LGT events are likely to have a negative effect on fitness [25]; apart from an added cost of translation, the high coding density of bacterial genomes means that inserted genes are likely to disrupt existing translation. Moreover, newly introduced gene products could be outright toxic through gene dosage effects, interfere with existing cellular interactions, or interfere with each other [26]. Third, accessory gene gains could be beneficial but only very transiently so. However, their removal would still imply that these genes ultimately have a deleterious effect.

The fact that the majority of lateral gene transfers are likely to be deleterious and quickly removed by **purifying selection** (unless this is made impossible by the presence of addiction mechanisms [27]), highlights the necessity of comparing very closely related genomes in order to reliably estimate the LGT rate. As a large number of deleterious events will have gone undetected, the ratio of gene content changes to mutation is likely to be significantly higher than estimated. Using data on the two pairs of most closely related *P. syringae* strains in the Nowell *et al.* [21] study, we divided the estimated number of gene gains by the number of estimated

Glossary

Distribution of fitness effects (DFE): mutations, as well as lateral gene transfer (LGT) events, have fitness effects that can be broadly divided into three categories. First, there are mutations that decrease fitness. Second, there are 'neutral' mutations, which have little or no effect on fitness. Third, there are advantageous mutations, which increase fitness by allowing organisms to adapt to their environment. However, in reality, there is a continuum of selective effects, stretching from those that are strongly deleterious, through weakly deleterious mutations, to neutral mutations and then on to mutations that are mildly or highly adaptive. The distribution of fitness effects refers to the relative frequencies of these types of mutation.

Genetic drift: the change in frequency of a mutation or accessory gene in a population as a result of chance, not selection.

Mutation rate modifier: a gene variant that influences the mutation rate.

Purifying selection: selection against genomic changes that lower fitness.