GENOME WATCH

The pyramid of knowledge

Georgios S. Vernikos

This month's Genome Watch discusses the benefits and limitations of bottom-up versus top-down comparative genomic analyses of microbial populations, drawing on an example from a recent application of new sequencing technologies to a long-term experiment.

Top-down methodologies are usually exploited when data are limited and methods cannot be fully analytical. Using these limited data, scientists try to envisage, predict or estimate the bigger picture. In other words, they use a handful of cases (the top) and try to extrapolate to a complete data set (the bottom). This is the case in the field of comparative microbial genomics. Although we have begun the transition to population genomics, with the increasing availability of more than one genomic sequence per lineage or species, most microbial populations are still represented in the biological databases by a single whole-genome sequence¹.

The main drawback of a top-down analysis is that it relies heavily on the parameters and hypotheses of a model that was built using a limited amount of data. As more data are fed into the model, the easier it is to generalize and the more realistic are the conclusions that can be drawn from the model. Theoretically, a model becomes uninformative once the sample of the available data approximates the totality of the data complexity. At that point, the power of expectations reaches the nadir and the power of the observations reaches the zenith. If, on the one hand, our predictions are reasonably close to our

observations, this is good news. If, on the other hand, our predictions show high levels of deviation from the observations, our initial hypotheses and the definitions that rely on them must be revisited. In addition, if this model affects other models, an on-going re-evaluation of our initial definitions, conclusions and discussions is required.

The advent of new sequencing technologies has substantially reduced the cost, in both time and money, of whole-genome sequencing, making it possible to accelerate our approximation of natural microbial diversity by maximizing the observations that we can derive from population genomics. With real data sampling from whole-genome sequences, bottom-up methodologies become feasible for the first time, even in the field of comparative genomics, in which researchers are trying to analyse a concept (the microbial species) that, by definition, has fuzzy boundaries.

A recent example of a bottom-up approach comes from the work of Barrick *et al.*², who sequenced whole genomes (at generations 2,000, 5,000, 10,000, 15,000, 20,000 and 40,000) sampled from a laboratory population of *Escherichia coli* that has been grown for 20 years with glucose as a limiting factor. The specific aim of this analysis was to quantify, by the application of new sequencing technologies, the dynamics of microbial evolution using a two-dimensional view: that of genomic and adaptive evolution. NimbleGen microarray-based comparative genome sequencing was used to screen for mutations in the clones from generations 2,000 and 20,000, and the Illumina 1G platform was then used to sequence all 6 clones to >50-fold coverage. In addition, any mutations identified were confirmed by targeted resequencing.

They found that there was a constant rate of genomic evolution but a decelerating rate of adaptive evolution up to generation 20,000. At around

generation 26,500, however, a mutator phenotype arose (caused by a 1 bp insertion resulting in a frameshift in the *mutT* gene) that led to an accelerated rate of genomic evolution. The most parsimonious explanation of these findings involves neutral mutations that spread stochastically by genetic drift, rather than beneficial mutations that spread by selection in the population. However, the key aspect of this analysis is that the observations reject the hypothesis that a neutral pattern of evolution is the dominant, underlying source of these results, suggesting that many of the mutations that are acquired are in fact beneficial. This rather counterintuitive and perhaps more complex, multifactorial pattern of evolution remains open to further analysis, although some alternative hypotheses (such as clonal interference) were put forward by the authors.

Looking at the pyramid of knowledge, the top is undoubtedly the ultimate goal, but there are two alternative routes by which it can be reached: a shortcut that involves starting at the top and envisaging the bottom (the top-down approach) and a longer route that starts at the bottom and tries to reach the top in a stepwise manner (the bottom-up approach). For scientific fields still in their infancy, the shortcut is the only solution; however, as those fields evolve and new technologies develop, the bottom-up approach will ultimately lead to a more realistic understanding of the big picture.

Georgios S. Vernikos is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk
doi:10.1038/nrmicro2300

1. Medini, D. *et al.* Microbiology in the post-genomic era. *Nature Rev. Microbiol.* **6**, 419–430 (2008).
2. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).

Competing interests statement

The author declares no competing financial interests.