Clipboard
DOI 10.1007/s12038-014-9441-0

Continuous evolution of laboratory strains of bacteria and yeast

The genomic composition of an organism is shaped by the selective pressures imposed by its niche environment. This interaction between genes and the environment creates an optimum phenotypic state of the organism. An alteration in either the genomic content or the environment perturbs this equilibrium.

The natural environment of bacteria is characterized by various physio-chemical fluctuations, and microorganisms have evolved strategies to withstand these changes rapidly and effectively. The optimum laboratory conditions that are used to grow bacteria aim to minimize external fluctuations. However, in principle, laboratory conditions constitute a new niche for bacteria whose genomes have been optimized by and for their natural niche.

The optimized laboratory growth conditions with reduced physio-chemical stress and abundant nutrients constitute a lifestyle conducive for higher growth rate. There are several observations that suggest accumulation of mutations in certain genes under laboratory conditions. One such frequent mutation encountered in several laboratory strains of *E. coli* is a mutation abrogating the function of RpoS (Atlung *et al.* 2002; Martinez-Garcia *et al.* 2003; King *et al.* 2006). RpoS is a stationary-phase sigma factor which is required for the expression of several stress-related genes. Under minimum stress conditions, the positive selection for maintenance of *rpoS* on the genome is relaxed. Other genes that frequently accumulate mutations in laboratory strains are *mutS*, a DNA mismatch repair gene, and *spoT*, which encodes the enzyme guanosine-3′,5′-bispyrophosphate (ppGpp) 3′-pyrophosphohydrolase responsible for cellular ppGpp degradation (Brown *et al.* 2001; Spira *et al.* 2008). The RpoS level is modulated by the cellular levels of ppGpp (Sarubbi *et al.* 1989). All these observations reveal that the genome undergoes changes in the new environment even when it is stress-free and stable. The availability of whole genome sequences of commonly used laboratory strains provides an opportunity to look at the genome evolution that happened in laboratory strains over the last six decades in retrospect.

Representatives of the two *E. coli* strains K12 and B have been used for most of the laboratory investigations across the globe. The classical experiments that laid the foundation of bacterial genetics and physiology were carried out with both K12 as well as B strains. Comparing the whole genome sequences of the routinely used laboratory strains MG1655 and W3110, both derivatives of *E. coli* K12 (Daegelen *et al.* 2009), revealed that although they are almost alike, there are certain critical differences. W3110 but not MG1655 has six nonfunctional genes, namely, *rpoS*, *dcuA*, *rcsC, gatA*, *dcuC* and *tnaB*. Many of these genes are disrupted by insertion elements (Hayashi *et al.* 2006). Comparative study between the whole genome sequences of the two widely used *E. coli* B strains REL606 and BL21 (DE3) is an interesting case study. The genomes of these two strains showed a large number of differences (Jeong *et al.* 2009). Moreover, the genomic changes showed strange distribution. For instance, REL606 carries a ~65 kbp region that shows an unusually large number (76% of the total SNPs) of SNPs (Studier *et al.* 2009). The differences were far too numerous to have accumulated since the time of their divergence from their ancestral strain to the present. This puzzle was solved by comparing genome sequences available for both B and K12 strains as well as tracing the practices used by laboratories working with B lineages. It appears that this ~65 kbp region was moved from *E. coli* K12 to REL606 by P1 transduction. The REL606 lineage had undergone two 1-methyl-3-nitro-1-nitrosoguanidine treatments resulting in 90 unselected SNPs and a couple of single base pair deletions as well. Another *E. coli* strain of B lineage, BL21 (DE3), had unusually large number of deletions, most of which were accounted for by the two rounds of UV treatment to which it was subjected in one of the laboratories (Studier *et al.* 2009).

Another case of misinterpretation of data due to the lack of information regarding genetic changes present elsewhere on the genome is the higher transformability of DH10B and DH5α, K-12 strains routinely used for plasmid preparation. The enhanced transformation ability of these strains was attributed to a

Keywords. DH5a; genome evolution; laboratory strains; MG1655; S288C

Published online: 1 July 2014

mutation in *deoR,* a transcriptional repressor of genes related to transport and catabolism of deoxyribonucleosides (Hanahan *et al.* 1991). It turns out that both DH10B as well as DH5α have a functional allele of *deoR* (Durfee *et al.* 2008; Xia *et al.* 2011). This warrants a re-investigation of the genetic basis of enhanced efficiency of transformation in these strains. DH10B also has a 13.5 times higher mutation rate of transposition of insertion element, especially IS*150*, as compared to the parent strain MG1655. These transposition events have led to further divergence of DH10B by driving a large duplication (~113 kb), which doubles gene dosage of 106 genes as well as an inversion (Durfee *et al.* 2008).

Therefore, in addition to the selection of spontaneous mutations, laboratory practices can further add to the forces that drive 'isogenic' laboratory strains to accumulate large-scale changes in their genomes. The uncharacterized genetic changes in laboratory strains described above are by no means exhaustive. These are intended only as selected examples to highlight the issue.

1. Single gene deletions in laboratory yeast strains drive compensatory mutations

The observations described above can be extended to yeast strains as well. Nutrient-rich laboratory conditions select for faster growth. A genome-wide comparison study carried out on *Saccharomyces cerevisiae* strain lineages that led to the laboratory strain (S288c) and a wild strain (YJM789) revealed that the laboratory strain tends to evolve faster than the wild strain (Gu *et al.* 2005). Several genetic studies have revealed that in the case of critical genes (whose functionality is maintained by strong purifying selection), loss of function rapidly results in the accumulation of compensatory secondary mutations (Barth *et al.* 1995; Stoebel *et al.* 2009). Conversely, what would happen if one tinkers with the genome while outside conditions remain more or less unaltered? What are the consequences of such genomic perturbations under normal laboratory conditions (normal in this context means the absence of deliberate selection)? Under laboratory settings, we often create targeted deletions of genes in order to understand their function. If the deletion of a gene does not result in a fitness loss compared with its wild-type counterpart, it is normally deemed to be non-essential. However, the phenotype conferred by a gene, due to its role in multiple genetic interactions, is contingent on the strain's genetic background (Dowell *et al.* 2010). Hence, the phenotypic consequences of a deletion mutation are likely to be concealed by the robustness of the genetic network of which the gene is a constituent and could vary depending on the genetic background as well as the environment.

In a similar analysis, Teng *et al.* (2013), from Johns Hopkins University School of Medicine, Baltimore, have systematically explored the consequence of genome-wide single gene knockouts available in yeast. They discovered that 'mutation of any single gene may cause a genomic imbalance with consequences sufficient to drive adaptive genetic changes'. They consider this to be a 'logical consequence of losing a functional unit originally acquired under pressure during evolution'.

The authors have screened for hidden heterogeneities in the survivability of the knockout strains by observing heat stress response as well as nutrient sensing under low amino acid conditions using replicates of the deletion strains obtained from different colonies. The presence of secondary mutations was confirmed by following their segregation in tetrads, subsequently verified by whole genomes sequencing in specific cases. Strains carrying deletions in the same gene, obtained from different sources, were evolved under non-stress conditions to determine whether they accumulate the same secondary mutation.

Genomic analysis revealed that these heterogeneities are due to secondary genomic changes and not due to stochastic changes in gene expression or other epigenetic changes, as both are often used to explain the heterogeneity in presumably isogenic populations. Moreover, the driver for these secondary changes is the original gene that is knocked out, as evident from the observation that independently constructed deletions of the same gene most often accumulated the same secondary mutations or mutations in the same complementation group (Teng *et al.* 2013). In many cases, the secondary mutations arose while growing the replicates of the deletion strains from individual colonies without selection, whereas in some cases they preexisted in the original deletion strain.

These results reinforce the fact that one must be cautious while interpreting the gene interaction studies involving deletions. Although the rest of the background is supposedly isogenic, the deletions may have unexpected consequence on the fitness of the strain resulting in the accumulation of second site suppressor mutations that are not documented. Therefore, prolonged manipulation of laboratory model organisms – whether bacteria or yeast – can result in unspecified genomic changes. Technological advancements in the last decade have made it feasible to screen for these genome-wide changes and thereby acquire accurate

information on the genotype of these organisms. The information obtained by whole genome sequencing complemented with hypothesis-driven experiments will advance our understanding of the evolvability of an organism both within the confines of the laboratory as well as in their natural environments.

Acknowledgements

I thank S Mahadevan and the two anonymous referees for their critical comments on the draft.

References

Atlung T, Nielsen HV and Hansen FG 2002 Characterization of the allelic variation in the *rpoS* gene in thirteen K12 and six other non-pathogenic *Escherichia coli* strains. *Mol. Genet. Genomics* **266** 873–881

Barth M, Marschall C, Muffler A, Fischer D and Hengge-Aronis R 1995 Role for the histone-like protein H-NS in growth phase-dependent and osmotic regulation of σS and many σS-dependent genes in *Escherichia coli*. *J. Bacteriol*. **177** 3455–3464

Brown EW, LeClerc JE, Li B, Payne WL and Cebula TA 2001 Phylogenetic evidence for horizontal transfer of *mutS* alleles among naturally occurring *Escherichia coli* strains. *J. Bacteriol*. **183** 1631–1644

Daegelen P, Studier FW, Lenski RE, Cure S and Kim JF 2009 Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394** 634–643

Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, *et al*. 2010 Genotype to phenotype: a complex problem. *Science* **328** 469–469

Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, *et al.* 2008 The Complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *J. Bacteriol*. **190** 2597–2606

Gu Z, David L, Petrov D, Jones T, Davis RW and Steinmetz LM 2005 Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **102** 1092–1097

Hanahan D, Jessee J and Bloom FR 1991 Plasmid transformation of *Escherichia coli* and other bacteria. *Method Enzymol.* **204** 63–113

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, *et al.* 2006 Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* **2** 1–5

Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, *et al*. 2009 Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol*. **394** 644–652

King T, Seeto S and Ferenci T 2006 Genotype-by-environment interactions influencing the emergence of *rpoS* mutations in *Escherichia coli* populations. *Genetics* **172** 2071–2079

Martinez-Garcia E, Tormo A and Navarro-Llorens JM 2003 Polymorphism in the *yclC-rpoS* region of enterobacteria. *Curr. Microbiol*. **46** 365–370

Sarubbi E, Rudd KE, Xiao H, Ikehara K, Kalman M and Cashel M 1989 Characterization of the spoT gene of *Escherichia coli*. *J. Biol. Chem.* **264** 15074–15082

Spira B, Hu X and Ferenci T 2008 Strain variation in ppGpp concentration and RpoS levels in laboratory strains of *Escherichia coli* K-12. *Microbiology* **154** 2887–2895

Stoebel DM, Hokamp K, Last MS and Dorman CJ 2009 Compensatory evolution of gene regulation in response to stress by *Escherichia coli* lacking RpoS. *PLoS Genet.* **5** e1000671

Studier FW, Daegelen P, Lenski RE, Maslov S and Kim JF 2009 Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.* **394** 653–680

Teng X, Dayhoff-Brannigan M, Cheng W-C, Gilbert CE, Sing C, *et al.* 2013 Genome-wide consequences of deleting any single gene. *Mol Cell.* **52** 485–494

Xia X-X, Qian Z-G and Lee SY 2011 Comparative proteomic and genetic analyses reveal unidentified mutations in *Escherichia coli* XL1-Blue and DH5a. *FEMS Microbiol. Lett.* **314** 119–124

SAVITA CHIB
National Centre for Biological Sciences
Bangalore 560 065, India
(Email, savitac@ncbs.res.in)