Streptococcal genomes provide food for thought

Hervé Tettelin

The genome sequences of two streptococci have implications for both the food industry and infectious disease.

The food industry relies heavily on bacteria in processes such as wine making, milk fermentation, cheese ripening and yogurt production. In particular, lactic acid bacteria are widely used and of major economic importance. In this issue, Bolotin *et al.*[1] present the complete genome sequences of two strains of *Streptococcus thermophilus*, a lactic acid bacterium involved in the manufacture of yogurt and cheese.

Lactic acid bacteria are Gram-positive bacteria that possess similar morphologic, metabolic and physiologic characteristics. A key metabolic distinction of this group is the rapid production of lactic acid from sugars. In the United States alone, these organisms are involved in the manufacture of food products with a yearly value of $20–30 billion[2]. The importance of the lactic acid bacteria group motivated the creation of the US Lactic Acid Bacteria Genome Consortium in the spring of 2001 (ref. 2). Its goal is to obtain a list of every protein in these organisms, including enzymes involved in the processes of interest. Related efforts are ongoing around the world, including projects initiated by the Institut National de la Recherche Agronomique (INRA) in France[3].

In 2001, Bolotin and colleagues published the complete genome sequence of *Lactococcus lactis*, the bacterium used as a starter in cheese fermentation[3]. This work led to the identification of new possibilities for fermentation pathways and for aerobic respiration. Now, Bolotin and a team of scientists from France, Belgium and the United States have struck again. They seem to have a clear passion for cheese, with which I concur, and the collaboration with Belgians was probably motivated by the perfect combination of cheese and beer. Surprisingly, however, the strains sequenced were isolated from yogurt—undoubtedly a healthier choice.

S. thermophilus is closely related to the nonpathogenic *L. lactis*, but it is even more closely related to other streptococcal species that constitute deadly human pathogens (**Fig. 1**). The latter include *Streptococcus pneumoniae* (or pneumococcus), which causes bacterial sepsis (blood infection), pneumonia, otitis media and meningitis; *Streptococcus pyogenes*, or group A *Streptococcus*, the agent of pharyngitis (strep throat), impetigo and rheumatic fever; and *Streptococcus agalactiae*, or group B *Streptococcus*, which infects neonates and results in bacterial sepsis, pneumonia or meningitis. *S. thermophilus* is also related to *Streptococcus mutans*, a cause of tooth decay. The genome sequence of *S. thermophilus* will likely allow the flavor and texture of yogurt to be improved or the lactose content to be reduced; it should also provide insight into any potential for pathogenicity of this species, which is generally recognized as safe.

From a metabolic perspective, the authors' genome analysis revealed an unusually high number of nonfunctional (decaying) genes, a third of which are related to sugar uptake, fermentation and degradation. These genes were sequenced in eight other *S. thermophilus* strains and turned out to be nonfunctional there as well. The authors conclude that these genes lost their function as *S. thermophilus* adapted to a milk environment with limited carbon sources.

The availability of more than one *S. thermophilus* genome sequence provides for comparative analyses. The two genomes reported by Bolotin *et al.* are 1.8 megabases in length and contain about 1,900 genes, more than 90% of which are identical in sequence. Most interesting is the comparison of these two genomes with the genome sequences of other streptococci. **Figure 1** shows the streptococcal genome projects that are ongoing or completed to date. This wealth of information enables thorough analysis of genes shared by all species, genes specific to *S. thermophilus* and genes specific to pathogenic versus nonpathogenic strains. The comparison of pathogenic and nonpathogenic strains is of particular interest as it provides insights into virulence mechanisms and disease processes.

Approximately 80% of the genes in *S. thermophilus* are similar to other streptococcal genes, highlighting its relatedness to pathogenic species. Nonetheless, many if not most of the genes known or predicted to be involved in virulence are either nonfunctional or absent in *S. thermophilus*. These include genes encoding cell-surface proteins known to interact with human cells and the immune system (lipoproteins, adhesins, choline-binding proteins, cell-wall anchored proteins, IgA proteases) and proteins involved in anchoring these proteins on the cell surface (sortases). On the other hand, some of the virulence-associated genes involved in synthesis of the polysaccharide capsule that surrounds pathogenic streptococci are maintained in *S. thermophilus*. Ironically, these happen to contribute to the production of external polysaccharides that confer the desirable texture of yogurt. It would be interesting to sequence a wild *S. thermophilus* isolate and compare it with industrial *S. thermophilus* strains and pathogens to trace the adaptive evolution of these organisms.

The authors also make use of comparative genomics among Gram-positive bacteria to study genome plasticity as it relates to the adaptation of pathogens to their hosts. Bacteria

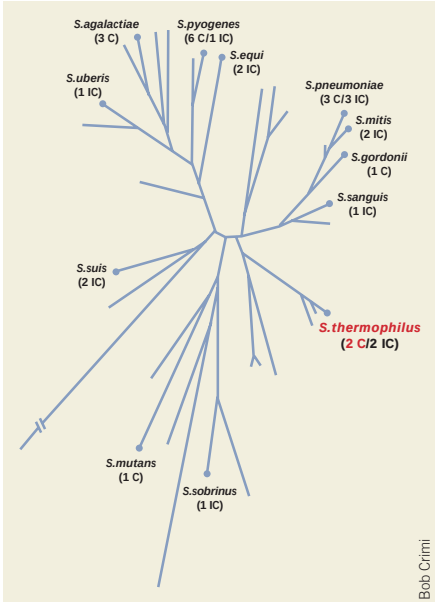

Figure 1 Schematic phylogenetic tree of streptococci (see a complete tree in ref. 7). Only streptococcal strains whose genome sequencing has been undertaken are displayed. Completed genomes (free of gaps) are indicated by 'C,' and incomplete (draft) genomes by 'IC.' The number of strains sequenced for each species is indicated next to C or IC. The two complete genomes of *S. thermophilus* reported by Bolotin *et al.*[1] are indicated in red. Links to sequencing centers and other information can be obtained by searching the GOLD database[8] for '*Streptococcus*' (http://www.genomesonline.org/).

Hervé Tettelin is at The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: tettelin@tigr.org

such as *S. thermophilus*, which are adapted to a 'sedate' nonthreatened lifestyle, have a relatively stable genome. Yet, as the authors note, all streptococci lack the genome-stabilizing helicase RecQ. Pathogenic streptococci lack two additional stabilizing enzymes involved in DNA double-strand break repair, whereas *S. thermophilus* and other lactic acid bacteria have maintained these functions.

Thus, pathogenic species harbor more plastic genomes with more potential for host adaptation and tissue tropism. An extreme example is *S. pneumoniae*, whose natural competence and extremely efficient recombination machinery result in modification of its genome content almost 'at will.' A comparison of three strains of *S. pneumoniae*, one of which is avirulent, revealed several major regions of plasticity. The gene content of these regions is variable and often related to virulence, but flanking regions are conserved in all three genomes[4].

We have come a long way from classical methods of strain selection, characterization and manipulation. Since the advent of genomics and the publication of the first complete bacterial genome just under a decade ago[5], it is difficult to imagine studying a microorganism without knowing its genome sequence. Not only does the sequence jump-start research, it also promotes high-throughput, global experimental designs such as transcriptomics, proteomics and metabolomics.

The genomics revolution has accelerated advances and discoveries in environmental and clinical microbiology, biotechnology and virtually every other field involving bacteria. The work of Bolotin and colleagues will greatly contribute to the dairy industry and to the characterization of streptococcal biology, pathogenicity and diversity at large. The two closed genomes of *S. thermophilus* will serve as the reference for this species and are of much greater value than unfinished, draft genomes[6]. And by the way, it turns out that yogurt made with *S. thermophilus* is safe for consumption after all.

1. Bolotin, A. *et al. Nat. Biotechnol.* **22**, 1554–1558 (2004).
2. Extraordinary Dairy. Dairy Industry Technol. Rev. Technical Bull. 1–6 (July, 2002) (http://www.extraordinarydairy.com/archive/innov_015_july_02.pdf).
3. Bolotin, A. *et al. Genome Res.* **11**, 731–753 (2001).
4. Tettelin, H. & Hollingshead, S.K. in *The Pneumococcus* (eds. Tuomanen, E., Mitchell, T., Morrison, D. & Spratt, B.) 15–29 (ASM Press, Washington, DC, 2004).
5. Fleischmann, R.D. *et al. Science* **269**, 496–512 (1995).
6. Fraser, C.M. *et al. J. Bacteriol.* **184**, 6403–6405 (2002).
7. Kawamura, Y. *et al. Int. J. Syst. Bacteriol.* **45**, 406–408 (1995).
8. Bernal, A., Ear, U. & Kyrpides, N. *Nucleic Acids Res.* **29**, 126–127 (2001).

A new harvest of fluorescent proteins

George H Patterson

A suite of improved fluorescent proteins offers better tools for cell biology.

In this issue, Shaner *et al.*[1] describe a veritable cornucopia of new fluorescent proteins, including mCherry, mBanana, mOrange, mStrawberry, mTangerine and mHoneydew. Fluorescent proteins make possible the relatively straightforward study of favorite proteins within living cells, and cell biologists have become increasingly dependent on them as research tools. Why should this latest crop interest the average cell biologist? The limitations of existing fluorescent proteins have direct consequences for experimental design and interpretation. Shaner *et al.* address some of these drawbacks and present new variants that should extend our capabilities.

The most prominent fluorescent proteins have been derived from the *Aequorea victoria* green fluorescent protein (avGFP)[2,3], and these molecules have been thoroughly altered and scrutinized since the expression of avGFP in heterologous organisms about 10 years ago. The limitations of most fluorescent molecules are generally associated with molecular brightness and/or stability. However, avGFPs and its initial descendents have additional complications involving protein folding, chromophore maturation and self-association. Most limitations associated with these properties have been alleviated and the proteins vastly improved (**Fig. 1**) with a variety of mutations to produce reliable fluorescent markers (see refs. 4 and 5 for reviews).

The first of the red fluorescent proteins, *Discosoma* DsRed[6], also has several pronounced undesirable properties, but mutant proteins that overcome many of these problems were quickly reported (see refs. 4 and 5 for reviews). One of these undesirable properties is that DsRed is an obligate tetramer. Although oligomerization may seem innocuous, protein-protein associations driven by the read-out marker can disrupt the normal localization, trafficking and protein-protein interactions of the protein of interest. This is particularly problematic if one considers single molecules participating in more than one oligomeric group at a time and essentially linking several oligomeric units together into a much larger complex. The production of monomeric RFP1 from DsRed by means of 33 mutations was a major breakthrough for red fluorescent proteins[7]. Yet, as good as it is, mRFP1 has several drawbacks relative to DsRed, such as decreased brightness and reduced photostability, which limit its usefulness as an imaging tool.

Shaner *et al.* avoided the effort required to monomerize the numerous other red fluorescent molecules available[8] by using mRFP1 as the starting molecule. By doing so, they evolved monomeric fluorescent proteins having altered spectra or improved brightness with each round of mutagenesis (**Fig. 1**). Taking advantage of the similarity between the avGFP and mRFP1 structures (see ref. 4 for review), they directed their improvements with knowledge gained from many years of work performed with avGFPs. For instance, the fluorescence of mRFP1 is significantly decreased when fused to the C terminus of a protein of interest, whereas avGFP is generally unaffected when fused to either end of the protein of interest. Therefore, by simply replacing the N terminus of mRFP1 with that of avGFP and adding the avGFP C terminus to mRFP1, the authors produced a protein that develops fully the red fluorescence signal. Secondly, Shaner *et al.* targeted the residues Q66 and Y67 in mRFP1 (equivalent residues in avGFP are S65 and Y66, respectively), which are known to have key roles in determining the spectral characteristics of the avGFPs. The final result, after mutagenizing these positions along with additional random mutagenesis, is a series of monomeric fluorescent proteins with emission peaks ranging from 537 nm to 610 nm.

What is the significance of this diversity of fluorescent proteins? The new proteins essentially fill the 'gap' between the most red-shifted

George H. Patterson is at the Cell Biology and Metabolism Branch of the National Institutes of Health, Bldg. 18T Rm. 101, 18 Library Drive, Bethesda, Maryland 20892-5430, USA.
e-mail: pattersg@mail.nih.gov