Genomes full of promise

Since the first bacterial genome was sequenced in 1995, over 100 others have been completed. They include many pathogens, as well as other bacteria chosen for their special interest, whether academic or applied. The resulting knowledge is revolutionising our understanding of the bacterial world.

**Streptomyces** is a genus of soil-dwelling bacteria with two unusual attributes - a complex developmental cycle and the ability to produce many of the antibiotics applied in medicine, as well as important drugs with other applications, such as anti-cancer and anti-parasitic agents, and immunosuppressants for use in tissue transplantation. Over the past 2 years, the complete genome sequences of two representative streptomyces have been published. The inventories of genes deduced from the sequences are throwing a powerful light on the strategies used by the streptomyces to compete in the soil habitat, as well as providing a huge potential toolbox for making novel antibiotics by genetic engineering.

**Streptomyces genomes contain many genes**

The genomes of such well-known bacteria as Escherichia coli, the workhorse of molecular genetics over the past 50 years, and Bacillus subtilis, a much-studied model for bacterial spore formation, are circular DNA molecules just over 4000 base pairs long, enough for about 4000 genes. In contrast, the Streptomyces chromosome is a linear molecule of about twice the size, with double the number of genes.

The Streptomyces colony is much more complex than that of the other two bacteria. Instead of consisting of a mass of separate rod-shaped cells like those of E. coli and B. subtilis, the Streptomyces colony is a mould-like system of interconnected, branching hyphae that first colonises the substrate as a so-called vegetative mycelium and then, when the food source is exhausted, gives rise to a sporulating aerial growth (Figure 1). Such a developmental programme might be expected to require a large number of genes to implement, and many have indeed been characterised, but this does not explain why the Streptomyces chromosome has thousands more genes than a Bacillus, which also needs to programme developmental events, in its case the production of heat-resistant endospores inside the rod-shaped mother cells. Most of the ‘extra’ genes seem to play other roles in adapting Streptomyces to live in the stress-rich environment of the soil.

Soils contain a huge variety of potential food sources, ranging from simple sugars and inorganic sources of nitrogen to hard-to-digest polymeric carbohydrates like cellulose and chitin (derived from the skeletons or walls of dead plants, insects and fungi), and complex nitrogen sources such as proteins. Moreover, the food sources vary from time to time and from place to place, so the Streptomyces genome encodes many suites of enzymes that can be called into play to deal with the different food sources as they are encountered.

Soils have many other variables too, such as temperature, pressure, pH, and the availability of oxygen and water, as well as the presence of other organisms that may represent competition to be met or potential colleagues with which to establish a symbiosis. The genome is full of genes whose products would meet these opportunities and threats, ensuring that the organism can thrive under a much wider set of conditions than most other bacteria, which have instead evolved to be supremely well adapted to a limited set of habitats.

With such an arsenal of genes, many of which are needed only under specific circumstances, it is no surprise that the genome is also provided with an unprecedented number - for a bacterium - of regulatory genes to switch on different sub-sets of genes in response to specific signals: one eighth of all the genes fall into this category, twice the proportion found in genomes half the size.

**The Streptomyces chromosome is linear**

Why is the Streptomyces genome linear, as in eukaryotes, rather than being circular like those of most bacteria? The answer is not obvious, especially because linearity brings with it the need for a special replication strategy to avoid the loss of coding sequences from the ends of the chromosome in each round of replication (a consequence of the fact that all DNA synthesis can only start with an RNA primer that is removed once the synthesis gets under way; if this is at the end of a molecule, a gap is left in the daughter strand). Eukaryotes overcome this problem using their complex telomeres, which constantly renew lost...
end sequences. Instead, streptomycetes have evolved a unique system to patch the gaps, using primer proteins permanently bound to the free ends.

The question about linearity becomes even more intriguing when we find that Streptomyces chromosomes occasionally mutate to a circular form by fusion of the ends, and they continue to replicate perfectly well. Linearity almost certainly represents an earlier state found in the ancestors of modern streptomycetes, and indeed in present-day actinomycetes with smaller genomes and narrower ecological niches, such as the mycobacteria that cause tuberculosis and leprosy.

Perhaps a clue to chromosome linearity is the finding that the genomes of the two sequenced streptomycetes show a biphasic structure, with a central core containing unconditionally essential genes such as those for cell division, central metabolism, DNA replication, transcription and translation, and arms representing nearly half the genome and packed with genes that would be adaptive under various sets of conditions (Figure 2).

Comparing the two Streptomyces genomes, the arms differ more strongly than the core in gene content, telling us that the arms are probably evolving at a faster rate by acquiring genes through horizontal transfer from other microorganisms, often on transposons. The arms can also exchange their ends with those of linear, transmissible plasmids, providing a potential route to such horizontal transfer. This recombination process depends on genome linearity.

**Engineering novel antibiotics**

What are the consequences of Streptomyces genome sequencing for antibiotic discovery and development? Over the past 10 years, a new field of biotechnology has grown from a glint in the eye to one that has produced drug candidates in Phase I clinical trials. This is ‘combinatorial biosynthesis’ of ‘unnatural natural products’. It stems from genetic studies of the biosynthesis of two chemical classes of antibiotics, the polyketides and the non-ribosomally synthesised peptides. The former includes blockbuster antibacterials like the tetracyclines and erythromycin, as well as anti-tumour drugs such as Adriamycin and the important anti-parasitic agent avermectin. The peptides include the most important immunosuppressants.

Genetic studies have revealed that both classes of compounds are made on giant enzymatic assembly lines that determine the complex product structure by a linear arrangement of catalytic sites acting in succession on the molecule as it travels along the assembly line. Such programming of the chemistry – by the nature, number and arrangement of the catalytic sites – is encoded in the Streptomyces genome and is readily amenable to manipulation by genetic engineering in a combinatorial fashion to generate compounds that are ‘natural’ because they are made in microorganisms but ‘unnatural’ because they are not found in nature. Kosan Biosciences Inc, a leader in this field based in Hayward California, calls this approach “doing chemistry by genetics”.

Current examples of this technology involve gene clusters for already known...
metabolites, but one of the excitementseven of whole genome sequencing is that it hasrevealed far more clusters of biosynthetic
genes for structurally complex chemicals
than had even been suspected, never
mind proven. The genome sequence of
Streptomyces coelicolor, the most
studied laboratory model for the genus,
revealed two dozen clusters of such
genes, while that of Streptomyces
avermitilis, the industrial producer of
avermectin, showed more than 30.1,2
Most of these clusters are in the arm
regions, emphasising the conditionally
adaptive nature of their products in the
soil environment.

Nearly all the clusters are different
between the two streptomycetes, telling
us that sequencing more Streptomyces
organisms will reveal an enormous
core of the chromosome is in dark blue and the arms in light blue. Ori denotes
the origin of chromosome replication and the blue circles at the ends of the
chromosome are protein molecules responsible for priming the special DNA
synthesis that ensures complete replication of the linear chromosome. The outer two
multi-coloured circles show the predicted genes on the two DNA strands as
coloured bars; note that the gene density is just as high in the arms as in the core.
The next, incomplete, circle includes a selection of essential genes, for cell division,
DNA replication, transcription, translation and amino acid biosynthesis; note
their location in the core region of the chromosome. For a full
explanation of the figure, see reference 1 [reprinted by permission from

natural products are made by engineered
strains at very low levels. This arises from
a variety of causes, such as poor substrate
availability, siphoning off of substrates by
competing pathways, low tolerance of the
novel compound by the engineered
producer, and a plethora of regulatory
influences.

An exciting goal is to engineer a ‘super-
host’ that made more of the desired
product by addressing each of these
problems in the light of the genome
sequence. For example, a missing
pathway for a novel substrate could be
introduced from another microorganism,
potential transport proteins for product
export could be added, and genes for
competing pathways could be identified
in the host genome and deleted. Thus,
while yield optimisation will continue to
be in part empirical and involve
traditional strain improvement by random
mutagenesis and screening, because of
the sheer complexity of the regulatory
circuits involved, rational steps will
increasingly become possible as we come
to understand more and more about the
genetic endowment of streptomycetes by
the application of the new techniques of
functional genomics.

References
1. Bentley SD, Chater KF, Cerdeño-Tarraga A-M
   et al. Complete genome sequence of the model
   actinomycete Streptomyces coelicolor A3(2).
2. Ikeda H, Ishikawa J, Hanamoto A et al.
   Complete genome sequence and comparative
   analysis of the industrial microorganism
   Streptomyces avermitilis.
3. Lin YS, Kieser HM, Hopwood DA & Chen CW.
   The chromosomal DNA of
   Streptomyces lividans is linear.
   Molecular Microbiology 1993; 10:923-933.
   Deciphering the biology of Mycobacterium
   tuberculosis from the genome sequence.
   Massive gene decay in the leprosy bacillus.
   Katz J. Modular organization of the genes
   required for complex polyketide biosynthesis.