

# Microbial Genomics: All That You Can't Leave Behind

## Dispatch

Craig M. Stephens<sup>1</sup> and Michael T. Laub<sup>2</sup>

**Projects designed to scan entire microbial genomes for essential genes have revealed a remarkably compact and conserved, but not universal, set of genes whose functions are necessary for survival or reproduction.**

What gene products are so fundamental to cellular survival or reproduction that they can be deemed 'essential'? This intriguing question has been addressed on a genomic scale in several microorganisms for which genomic DNA sequence data can be combined with sophisticated tools for genetic manipulation. Genome-wide functional profiling projects represent an important trend in genomics research which should provide new insight into critical, but poorly understood, cellular processes [1]. Identifying new essential genes in microbes could also lead to new antibiotic targets [2,3], an important motivation in light of the increasing numbers of antibiotic-resistant pathogens, and the fact that all the major classes of clinically useful antibiotics target the products of essential genes. We shall focus here particularly on new work from an international consortium of researchers describing systematic analysis of the genome of the common soil bacterium *Bacillus subtilis* [4].

Several approaches have been taken to identify essential genes in bacteria. An initial attempt simply compared the first two fully sequenced bacterial genomes, those of *Mycoplasma genitalium* and *Haemophilus influenzae* (Table 1) [5]. These two parasitic bacteria are phylogenetically distant and have highly streamlined genomes, so genes conserved in both species are good candidates for carrying out important functions. Careful identification of orthologs in the two genomes delineated 256 putatively essential genes. More than half of these are involved in the basic cellular information processing machinery: protein synthesis (ribosomal proteins, translation factors and aminoacyl-tRNA synthetases), DNA replication and repair, and RNA synthesis. The remainder are distributed among functions including glycolysis, protein export to the cell envelope, chaperones and transport processes. Over 70% of the shared genes identified by Mushegian and Koonin [5] are conserved across the evolutionary spectrum from eubacteria to archaea to eukaryotes, supporting the premise that reductive genome evolution selects strongly for functions universal to all life.

Of course, a purely bioinformatic approach cannot conclusively assess the importance of every gene in these genomes, or (in this case) be expected to generate a complete list of essential genes for the organisms being compared. For example, alternative strategies for transport or processing of critical nutrients could have evolved, driven by differences in host niches. In extending these observations to other species, it must be recognized that genes that are essential for an obligate parasite may represent only a subset of those necessary for organisms that must contend with more challenging environments outside a host. To address the issue experimentally, massive screens based on transposon mutagenesis and antisense RNA inhibition have been applied (Table 1) [2,3,6–9]. The results indicate that the computationally predicted set of essential genes [5] is largely valid [10], with some exceptions: conserved genes that are not essential, and non-conserved genes that are necessary, presumably reflecting the distinct physiologies and evolutionary histories of the experimental organisms.

Random approaches such as transposon mutagenesis and antisense RNA inhibition are potentially flawed, in that it is difficult to be certain that every gene has been hit with mutations (or inhibitory RNAs). The error rates — dispensable genes categorized as essential, or essential genes that are missed — may also be significant [1–4,6–9]. In light of this, systematic targeted approaches designed to address every gene with certainty have obvious advantages. Such efforts require more substantial organizational efforts and financial commitment than the random mutagenesis approaches mentioned earlier, and indeed may be more costly now than a microbial genome sequencing project. Nevertheless, functional profiling is a logical follow-up for model organisms whose genome sequences have been completed, and the initial results of systematic gene disruption projects are now available for the yeast *Saccharomyces cerevisiae* [1] and the bacterium *Bacillus subtilis* [4].

*Bacillus subtilis* has been studied extensively as a model for the process of endospore formation. Its genome has about 4100 predicted genes (Table 1). Kobayashi *et al.* [4] report on efforts to disrupt nearly 3000 *B. subtilis* genes; for the remaining genes, either experimental or predictive data were available already or they are carried on dispensable integrated phage genomes. Only 271 *B. subtilis* genes appear to be essential in the entire genome. It should be qualified, however, that only one growth condition was tested experimentally. The manner in which cultures are grown — identity and abundance of nutrients, temperature, pH, salinity/osmolarity, aeration — profoundly impacts the relative importance of various genes. Kobayashi *et al.* [4] used optimal growth conditions, including a complex medium with high concentrations of amino acids and other organic compounds, which minimized the range of biosynthetic pathways needed. How many

<sup>1</sup>Biology Department, Santa Clara University, 500 El Camino Real, Santa Clara, California 95053, USA.

E-mail: cstevens@scu.edu <sup>2</sup>Bauer Center for Genomics Research, Harvard University, 7 Divinity Ave., Cambridge, Massachusetts 02138, USA.

Table 1. Identifying essential genes in microorganisms.

Organism	Approximate number of protein coding genes	Number of genes demonstrated or inferred to be essential	Method for defining essential genes	Reference
<i>Bacillus subtilis</i>	4100	271	Systematic knockout	[4]
<i>Haemophilus influenzae</i>	1725	478	Targeted transposon mutagenesis	[9]
<i>Saccharomyces cerevisiae</i>	6100	1105	Systematic knockout	[1]
<i>Staphylococcus aureus</i>	2600	658	Antisense RNA inhibition	[2]
<i>Mycoplasma genitalium</i>	517	265–350	Transposon mutagenesis/ statistical analysis	[6]
<i>M. genitalium/H. influenzae</i>	517/1725	256	Prediction by comparative analysis	[5]

more genes would be found to be essential were an inorganic salts medium with glucose as sole carbon source used? Answers to such questions can be ascertained relatively easily, as the mutant strains generated in this study can be examined for growth under other conditions as well.

The number of essential genes in *B. subtilis* is consistent with earlier estimates of essential genes in *H. influenzae* and *M. genitalium* (Table 1), a fascinating observation in light of the dramatic differences in genome size and complexity, and the dramatically different niches these organisms occupy. Essential genes in *B. subtilis* also tend to be conserved in other organisms [4]. Over 75% of the essential *B. subtilis* genes have homologs in at least 75% of the eubacterial genomes examined, and over 30% have apparent homologs in organisms spanning the phylogenetic spectrum from eubacteria to archaea and eukaryotes.

Not surprisingly, components of the cellular information processing machinery — 136 of the 271 essential *B. subtilis* genes (Figure 1) — comprise many of the broadly conserved components. Cell envelope functions constitute many of the remaining essential genes, including those involved in membrane lipid assembly, protein export and elements of the cell division machinery. The degree of conservation of gene products associated with the cell envelope varies, depending on factors such as the nature of the cell wall. Several essential *B. subtilis* genes are involved in

the synthesis of peptidoglycan and teichoic acids. Peptidoglycan synthesis is widely important among eubacteria, but teichoic acids are restricted to Gram-positive bacteria, so genes encoding enzymes for its synthesis are not necessary outside this group.

Cells need a mechanism for generating ATP, and the glycolytic pathway for catabolism of glucose (accompanied by substrate-level phosphorylation) is widely conserved. Interestingly, in both *B. subtilis* and *Staphylococcus aureus*, nearly all of the enzymes of glycolysis are essential, even in rich media replete with amino acids whose degradation can provide alternative sources of carbon and energy. Dual roles for some glycolytic enzymes in gluconeogenesis does not explain their requirement, as provision of glucose in the media had no effect.

When challenged by starvation, *B. subtilis* executes an elaborate developmental program to create an endospore. Dozens of transcription factors and regulatory genes are necessary for spore formation. In contrast, remarkably few genes known to have regulatory functions were found to be essential for vegetative growth in rich media [4]. Only one known transcription factor, YycF, which is thought to regulate cell division genes [11], was found to be necessary for growth in rich media. Six members of the Era/Obg family of GTP-binding proteins are also essential. These bacterial ‘G proteins’ are not well understood, but at least some of them seem to be involved in signaling

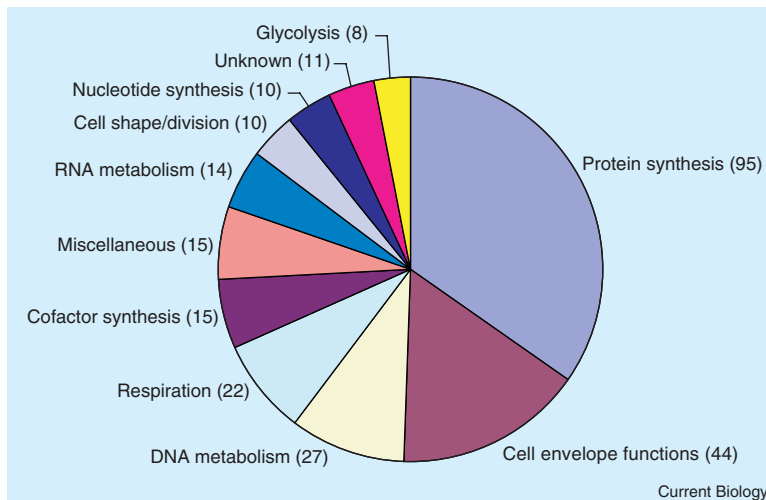


Figure 1. Functional categories of essential genes in *Bacillus subtilis*. Data are summarized from [4].

processes related to cell cycle progression [12]. Of course, the functions of several genes identified as essential in *B. subtilis* are completely unknown, and regulatory roles for these proteins are possible as well. Even so, the stark difference in essential regulators between vegetative growth and sporulation indicates that most of the regulatory systems encoded in *B. subtilis* genome mediate adaptive responses and aid survival under harsh conditions, but are dispensible under ideal growth conditions.

The potential to define a minimal gene set sufficient for life, then tinker with it to create custom-tailored microbes, garnered considerable attention in the popular press when Hutchison *et al.* [6] described their work on *Mycoplasma* in 1999. Whether this can be interpreted as defining life, or will lead to the creation of truly new life forms, is debatable, but the notion that life can be reduced to a universal, compact genetic recipe seems overly simplistic. The evidence thus far suggests a common set of gene functions — sometimes manifested in unrelated genes in different organisms — that are necessary, but not sufficient, for microbial life [10,13]. This shared set represents a genomic framework onto which evolutionary forces have bolted other genes, creating organisms suited to different habitats. The thousands of 'non-essential' genes in the microbes under study are probably there for a reason — additional effort should be put into examining on a genome-wide scale the effects of mutations on growth under varied environmental conditions, as has been done with *S. cerevisiae* [1]. Functional profiling of other simple eukaryotes in the near future should reveal more about the eukaryotic minimal gene set as well. And so, we move beyond genome sequencing to start to understand how all those genes really work!

#### References

1. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K. Andre, B., *et al.* (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391.
2. Forsyth, R.A., Haselbeck, R.J., Ohlsen, K.L., Yamamoto, R. T., Xu, H., Trawick, J.D.A., Zamudio, C., Foulkes, J.G. and Zyskind, J.W. (2002). A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* 43, 1387–1400.
3. Gerdes, S.Y., Scholle, M.D., D'Souza, M., Bernal, A., Baev, M.V., Farrell, M., Kurnasov, O.V., Daugherty, M.D., Mseeh, F., Polanuyer, B.M., *et al.* (2002). From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J. Bacteriol.* 184, 4555–4572.
4. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G. andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., *et al.* (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4678–4683.
5. Mushegian, A.R. and Koonin, E.V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10268–10273.
6. Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O. and Venter, J.C. (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286, 2165–2169.
7. Sassetti, C.M., Boyd, D.H. and Rubin, E.J. (2001). Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12712–12717.
8. Ji Y, Zhang B, Van SF, Horn, Warren P, Woodnutt G, Burnham MK, Rosenberg M. (2001). Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 293, 2266–2269.
9. Akerley, B.J., Rubin, E.J., Novick, V.L., Amaya, K., Judson, N. and Mekalanos, J.J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 966–971.
10. Koonin, E.V. (2000). How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.* 1, 99–116.
11. Fukuchi, K., Kasahara, Y., Asai, K., Kobayashi, K., Moriya, S. and Ogasawara, N. (2000). The essential two-component regulatory system encoded by *ycyF* and *ycyG* modulates expression of the *ftsAZ* operon in *Bacillus subtilis*. *Microbiology* 146, 1573–1583.
12. Morimoto, T., Loh, P.C., Hirai, T., Asai, K., Kobayashi, K., Moriya, S. and Ogasawara, N. (2002). Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. *Microbiology* 148, 3539–3552.
13. Peterson, S. and Fraser, C. (2001). The complexity of simplicity. *Genome Biol.* 2, 2002.1–2002.8.