

Minimal Genome Projects

Question: What is the minimum number of genes required for life?
(at least under laboratory conditions)

1. Bioinformatic approach

What genes are conserved in all completed genomes? This may be the minimal core set.

This approach was initially used by Mushegian and Koonin in 1996. They looked for all genes conserved among the 3 complete bacterial genomes available at the time (*H. influenzae*, *M. genitalium*, *E. coli*).

The result:

256 genes in minimal set

Some problems:

- with distantly-related species it is hard to identify orthologs
- some functions can be performed by non-orthologous genes (functional analogs). This is also known as NOD (Non-Orthologous gene Displacement). These cases can be estimated if functional data are available
- results may change with each new genome
- no evidence that minimal set alone is sufficient for life

A more recent check of the COG (Clusters of Orthologous Genes) database reveals:

- 63 genes are common to all of life (Archaea, Eubacteria, Eukaryotes)
- 91 genes are common to Archaea and Eubacteria
- 217 genes are common to Eubacteria

2. Experimental approach

An initial experiment was performed by Craig Venter and colleagues at TIGR in 1999.

They began with 2 closely-related species of bacteria:

Mycoplasma genitalium – the smallest genome known (580 kb, 480 protein-encoding genes)

M. pneumoniae (816 kb, 480 *M. genitalium* orthologs + 197 unique genes = 677 total genes)

Used transposable element (TE) insertion mutagenesis to randomly knock out genes in both species, then sequenced DNA flanking the TE to determine where it inserted.

If TE insert was within first 80% of protein-coding region and beyond nucleotide 9 of the coding sequence it was considered a “gene disruption” or “hit”

Only surviving cells could be detected, so “hit” genes must be non-essential.

total hits:

93 genes in *M. genitalium*

150 genes in *M. pneumoniae* (57 had *M. genitalium* orthologs, 93 were unique)

For *M. genitalium*, the directly observed minimal gene number is $480 - 93 = 387$.

However, this will be an overestimate, because this was not a saturation screen: some genes were not hit just by chance.

Minimal genome estimation:

Assume the 197 *M. pneumoniae*-specific genes are non-essential.

Mutations were recovered in 93 (47%) of these genes.

Assume that only 47% of the non-essential *M. genitalium* orthologs were hit, $0.47 N = 57$, $N = 121$ where N is the number of non-essential genes. $480 - 121 = 359$ essential genes (maybe fewer).

Final estimate = 265 – 350 essential genes.

111 genes of unknown function were not disrupted, many of these may be required for life.

Problems:

- Not all 197 *M. pneumoniae*- specific genes may be non-essential in *M. pneumoniae* (“new” genes can become essential)
- Genes were knocked-out individually. What about synthetic lethals?

Can a synthetic organism be engineered by combining these 359 (or fewer) genes?

3. Another experimental approach

Another experiment was performed by Kobayashi *et al.* in 2003:

They began with the common laboratory bacterium, *Bacillus subtilis* (4100 genes) and individually knocked-out genes using a targeted, homologous recombination approach. These results were combined with those of previous studies.

The results:

271 genes (6%) were essential for growth under optimal laboratory conditions.

The remaining 3,830 (94%) were non-essential!

Essential genes were classified into functional categories (about 50% were involved in DNA/RNA metabolism or protein synthesis). Over half of the protein synthesis genes encode ribosomal proteins.

The essential genes of *B. subtilis* are well-conserved in other bacterial species, and many are also found in Archaea and/or Eukaryotes. However, not all were completely conserved over all known genomes or even over all known bacterial genomes. Thus, these essential genes would not show up as part of the minimal genome by a strict bioinformatic approach. Of the essential genes, $\approx 30\%$ were conserved across 54 bacterial species and $\approx 20\%$ were conserved across 18 Archaea and Eukaryotic species.

Again, the genes were knocked-out individually. What about synthetic lethals? Can an organism with only 271 genes be engineered?

In general, the number of essential genes is not strongly correlated with the total number of genes in the genome. However, there is a negative correlation between the percentage of essential genes and the total number of genes in the genome. In organisms with few genes, such as *M. genitalium*, a high proportion of the genes are essential.

More recent work has shown that it is possible to transplant entire genomes and engineer synthetic life (see Gibson *et al.* 2008, 2010). This approach has been used to make a synthetic strain of *Mycoplasma* that has only 473 genes (see Hutchison *et al.* 2016).

References:

Mushegian AR, Koonin EV, 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* 93: 10268-10273.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC, 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286: 2165-2169.

Kobayashi K, *et al.*, 2003. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA* 100: 4678-4683.

Gibson DG, *et al.*, 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319: 1215-1220.

Gibson DG, *et al.*, 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329: 52-56.

Hutchison CA, *et al.*, 2016. Design and synthesis of a minimal bacterial genome. *Science* 351: 1414.