Shrinking the Bacterial Genome

Jennifer Tsang

Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02115, USA.
Email: jtsang@bidmc.harvard.edu

Synthetic biology is an interdisciplinary science, combining engineering and biological principles to design and construct new biological entities. Not only is synthetic biology useful for understanding life's complex circuitries, this discipline has many practical applications such as modifying biological pathways to produce nanoparticles, proteins and other molecules for specific uses. As a rapidly evolving field, particularly due to recent advances in DNA sequencing and synthesis, synthetic biology holds great potential for transforming the future of biological research.

In a recent publication in Science, the J. Craig Venter Institute and Synthetic Genomics, Inc. report the design and synthesis of a synthetic cell (named syn3.0) that contains only the minimal number of genes required for life (Hutchison *et al.*, 2016). This momentous feat was the culmination of twenty years of research, beginning with the sequencing of the *Mycoplasma genitalium* and *Haemophilus influenza* genome in 1995 (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995), transposon mutagenesis to identify nonessential genes in *M. genitalium* in 1999 (Hutchison *et al.*, 1999), and the creation of the first synthetic cell (*Mycoplasma mycoides* JCVI-syn1.0) in 2010 (Gibson *et al.*, 2010). The creation of a minimal bacterial genome is an exciting milestone that sheds light on our current understanding of life and paves the way in exploring essential

functions needed for life that are currently undiscovered.

Bacteria with large genome, such as *Escherichia coli* which contains ~4,000 − 5,000 genes, have evolved to adapt to diverse environments whereas organisms such as mycoplasmas have evolved to grow in nutrient-rich animal host environments. The smallest known mycoplasma genome (*M. genitalium*) contains 525 genes (Fraser *et al.*, 1995). Thus the already small genome size of mycoplasmas made it a great starting point to create a genome with only the essential genes for life. The minimal genome bacterium created by the J. Craig Venter Institute and Synthetic Genomics, Inc. (subsequently referred to as syn3.0) is derived from the *M. genitalium* genome and has a total of 473 genes (Hutchison *et al.*, 2016).

The minimal genome cell was initially designed based upon information from the literature and systematic, repeated transposon mutagenesis experiments. To create this minimal genome, the group embarked on several design-build-test (DBT) cycles on a modified syn1.0 genome. The original syn1.0 genome was split into eight segments and one segment was modified at a time (design), combined with the seven-eighths syn1.0 genome (build) and tested for viability in an *M. capricolum* recipient (test). Each segment was modified by removing genes deemed nonessential based upon transposon mutagenesis experiments (Hutchison *et al.*,

1999). Transposon mutagenesis is a technique where a transposon is inserted into a host's genome (Figure 1). If a transposon inserts into a gene, the gene's function is disrupted. Because transposons insert randomly into the genome, a transposon mutant library can be created such that each viable cell represents a unique transposon insertion and the collection of transposed cells have transposons inserted in all non-essential genes. By sequencing across the transposon-chromosome boundary, the non-essential genes could be identified.

However, a viable cell could not be created using information from the 1995 transposon mutagenesis experiment due to synthetic lethal pairs. Synthetic lethal pairs are pairs of genes that perform the same essential function and deletion of either gene is viable, but deletion of both genes is lethal. The challenge in identifying these genes occurs when synthetic lethal pairs are spread across different segments of the split genome. To identify these "quasi-essential"

genes, the group performed further transposon mutagenesis experiments on a viable cell containing a hybrid of syn1.0 genome fragments and reduced genome fragments. Genes deemed non-essential after this subsequent transposon mutagenesis step were deleted and resulting genome fragments were tested.

After many DBT cycles, the syn3.0 synthetic genome, comprising of 473 genes sufficient to support life, was created. As is often the case, the researchers made engineering decisions in creating syn3.0 that could affect the absolute minimum number of required genes. syn3.0 was designed for growth in rich media that supplies nearly all of the essential small molecules and thus, genes involved in transport, catabolism, and other metabolic processes have been eliminated. In contrast, transporter systems could not be eliminated from the minimal genome as the cell was devised to transport small molecules from the rich growth media. Therefore, it is likely that the minimal genome
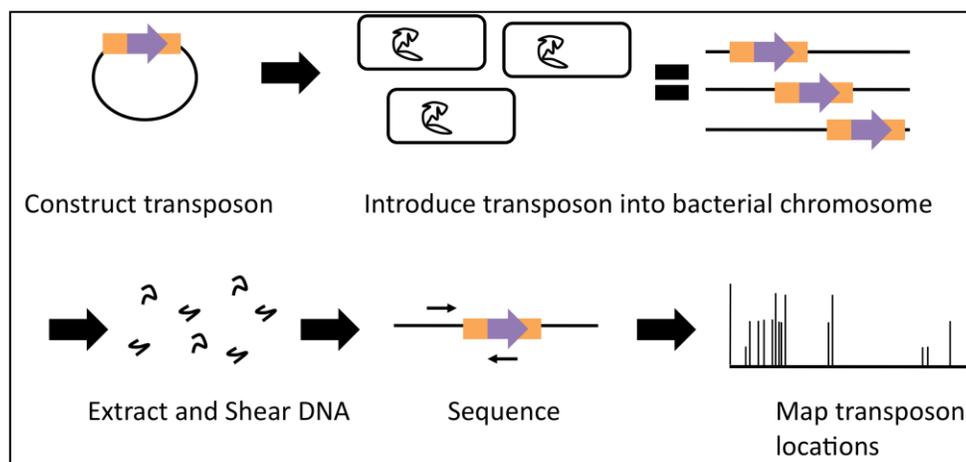
Figure 1. Outline of transposon mutagenesis. (1) Construction of transposon fragment for insertion into bacterial chromosome. (2) Transposon fragments are introduced into the bacterial chromosome at random locations. (3) DNA from viable transposants is extracted and sheared. (4) Transposon:chromosome boundaries are sequenced to determine location of transposon insertion. (5) Transposon insertion sites are mapped onto the chromosome. Sites that are not represented likely indicate transposon insertions into essential genes and sites that are represented likely indicate transposon insertions into non-essential genes.

identified here is not universal to every environment and that the genes required for a minimal genome differ depending on the growth conditions. The minimal genome for a cell grown in a more hostile or austere environment is likely larger than one grown in rich media. However, the minimal genome of syn3.0 likely represents genes that are fundamental to all minimal genomes and thus, exemplify all essential processes for life.

One of the greatest revelations from this study is that there are still so many essential genes with unknown functions. Surprisingly, out of the 473 genes in the minimal genome, 149 (31%) genes could not be assigned a specific biological function. Some unassigned genes have been found in a wide variety of organisms (prokaryotes and even eukaryotes) and may represent novel fundamental functions essential for life. Of the genes with known functions in the minimal genome, almost all of the genes involved in (1) reading and expressing genetic information (195 genes or 41% of genes in syn3.0) and (2) preserving genetic information across generations (34 or 7% of total genes in syn3.0) have been retained. Membrane-related genes account for 84 (18%) of the total genes in syn3.0. and 71 genes (17%) were involved in cytosolic metabolism.

The design of a minimal genome has many positive and negative outcomes for synthetic biology. syn3.0 has a replication time of approximately 3 hours which is much slower than the 20-minute replication time of *E*. *coli* and the 90-minutes of *Saccharomyces cerevisiae* (yeast). However, the 3-hour doubling time of syn3.0 is much quicker than the 16-hour doubling time of wild-type *M*. *genitalium*. *E*. *coli* and S. *cerevisiae* are commonly used in microbial

engineering because of their ease in manipulations and quick growth rates. The design of a minimal genome in organisms such as *E*. *coli* and *S*. *cerevisiae* may create microorganisms with an even shorter doubling time. Several research groups have created *E*. *coli* reduced genomes by deleting from 6.8% to 29.7% of genes resulting in a wide range of growth effects from impaired growth to increased cell biomass in batch cultures (reviewed in (Choe *et al.*, 2016)). Reduced genomes of *Bacillis subtillis* (bacteria) and *Schizosaccharomyces pombe* (yeast) have also been created (reviewed in (Choe *et al.*, 2016)). Future work to reduce the genome size in these organisms in a way that increases total biomass or increases growth rate will be advantageous in many industrial applications that require efficient and robust production of chemicals, drugs, biofuels, and others.

A minimal genome may also reduce the regulatory and metabolic complexities while engineering microbes. This could create microbial engineering platforms that are more predictable and the impact of any added pathways can more easily be modeled computationally. However, minimal genomes may sacrifice metabolic pathways that are necessary for the goals of industrial use. For example, genes responsible for synthesizing metabolic intermediates or co-factors necessary for the production of the desired molecule may have been deleted in the reduced genome and need to be reintroduced during the engineering process.

Taken together, the successful synthesis of a minimal genome lays the foundations for exciting work in identifying novel functions necessary for life, increasing efficiency in industrial processes,

and whole-genome design. By removing non-essential genes, the minimal genome demonstrates how much still remains unknown about the foundations of life and provides a robust platform for studying such functions.

Acknowledgements

References

Choe, D., Cho, S., Kim, S.C., and Cho, B.-K. (2016) Minimal genome: Worthwhile or worthless efforts toward being smaller? Biotechnol. J. 11: 199–211.
http://dx.doi.org/10.1002/biot.201400838

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:   496–512.
http://dx.doi.org/10.1126/science.7542800
PMid:7542800

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270: 397–403.
http://dx.doi.org/10.1126/science.270.5235.397
PMid:7569993

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52–6.
http://dx.doi.org/10.1126/science.1190719
PMid:20488990

Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., et al. (2016) Design and synthesis of a minimal bacterial genome. Science (80-.). 351: aad6253–aad6253.
http://dx.doi.org/10.1126/science.aad6253

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., et al. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286: 2165–9.
http://dx.doi.org/10.1126/science.286.5447.2165 PMid:10591650