

## J. Craig Venter Institute

# 42

### Estimation of the Minimal Mycoplasma Gene Set Using Global Transposon Mutagenesis and Comparative Genomics

John I. Glass\* (JGlass@venterinstitute.org), Nina Alperovich, Nacyra Assad-Garcia, Shibu Yooseph, Mahir Maruf, Carole Lartigue, Cynthia Pfannkoch, Clyde A. Hutchison III, Hamilton O. Smith, and J. Craig Venter

J. Craig Venter Institute, Rockville, MD

The Venter Institute aspires to make bacteria with specific metabolic capabilities encoded by artificial genomes. To achieve this we must develop technologies and strategies for creating bacterial cells from constituent parts of either biological or synthetic origin. Determining the minimal gene set needed for a functioning bacterial genome in a defined laboratory environment is a necessary step towards our goal. For our initial rationally designed cell we plan to synthesize a genome based on a mycoplasma blueprint (mycoplasma being the common name for the class *Mollicutes*). We chose this bacterial taxon because its members already have small, near minimal genomes that encode limited metabolic capacity and complexity. We took two approaches to determine what genes would need to be included in a truly minimal synthetic chromosome of a planned *Mycoplasma laboratorium*: determination of all the non-essential genes through random transposon mutagenesis of model mycoplasma species, *Mycoplasma genitalium*, and comparative genomics of a set of 15 mycoplasma genomes in order to identify genes common to all members of the taxon.

Global transposon mutagenesis has been used to predict the essential gene set for a number of bacteria. In *Bacillus subtilis* all but 271 of bacterium's ~4100 genes could be knocked out. Because of the redundancy of many genes and systems in *B. subtilis* and other conventional bacteria, in these bacteria disruptions of genes involved in essential functions are often not lethal. *M. genitalium* is a slow-growing human urogenital pathogen that has the smallest known genome of any free-living cell at 580 kb. There is almost no redundancy in this genome, and as such *M. genitalium* is often used as a model of a minimal cell. In 1999 we published preliminary estimate of the *M. genitalium* minimal gene set. Global transposon mutagenesis identified 130 of the 484 *M. genitalium* protein-coding genes not essential for cell growth under laboratory conditions, and that work predicted that in a complete study still additional genes would likely be disrupted. In our current effort we have improved the technique to allow isolation and characterization of disruption mutants. Surprisingly, after attaining saturation mutagenesis of the *M. genitalium* genome we could only identify 98 disrupted genes, suggesting that for growth under our laboratory conditions the minimal mycoplasma essential gene set is ~386 protein coding genes. Some genes were disrupted that are involved in presumably critical metabolic processes, such as lactate, pyruvate and glycerol-3-phosphate dehydrogenases. This suggests that as has already been shown for some *M. genitalium* kinases, these dehydrogenases may be somewhat redundant due to less than stringent substrate specificity.

The 15 mycoplasma genomes comprise an excellent comparative genomics virtual laboratory. Previous similar computational comparisons of genomes across diverse phyla of the eubacteria are of limited value. Because of non-orthologous gene displacement, pan-bacterial comparisons identified

less than 100 genes common to all bacteria; however determination of conserved genes within the narrow mycoplasma taxon is much more instructive. We identified 169 protein coding genes present in all of the complete mycoplasma genome sequences, and an expanded core set of 310 genes that are encoded in almost every member of our set of genomes. The additional genes in the expanded core gene set take into account that non-glycolytic mycoplasmas do not encode some glycolysis genes for instance, and that the obligate intracellular plant parasite, *Phytoplasma asteris*, has dispensed with many genes because the functions are provided by its host. At least 36 elements of the expanded core gene set are non-essential based on the gene disruption study. The combination of comparative genomics with the gene disruption data, and reports of specific enzymatic activities in different mycoplasma species enabled us to predict what elements are critical for this bacterial taxon. In addition to determining the consensus set of genes involved in different cellular functions, we identified 10 hypothetical genes conserved in almost all the genomes, and paralogous gene families likely involved in antigenic variation that comprise significant fractions of each genome and are presumably unnecessary for cell viability under laboratory conditions.

## 43

### Progress toward a Synthetic Cellular Genome

Hamilton O. Smith\* (hsmith@venterinstitute.org), Cynthia Pfannkoch, Holly A. Baden-Tillson, Clyde A. Hutchison III, and J. Craig Venter

J. Craig Venter Institute, Rockville, MD

To test our understanding of the genetic requirements for cellular life, we proposed to construct a minimal cellular genome by chemical synthesis (Hutchison, et al., 1999). A number of technical hurdles remain before this can be accomplished and we report progress on these here.

We improved upon the methodology and shortened the time required for accurate assembly of 5- to 6-kb segments of DNA from synthetic oligonucleotides. We first tested our methodology by assembly of infectious phiX174 genomes (Smith, et al., 2003). The methods have since been tested by assembling three segments (4.6, 5.3, and 6.5 kb) of the mouse mitochondrial genome. In each case assembly was straightforward, so we feel the methods are quite robust. We have made improvements in these assembly methods aimed at improving sequence accuracy of the assembled DNA.

About one lethal error per 500 bp occurred in our phiX174 genome assembly, assuming a random distribution of errors among assembled genomes. We have approached the elimination of such errors at four levels: 1) Purifying oligonucleotides prior to assembly by high-throughput capillary electrophoresis, to remove molecules containing errors that lead to mutations 2) Modifying assembly conditions to minimize DNA damage leading to mutations, 3) Global error correction of the assembled product by biochemical methods directed at producing the consensus sequence of the assembled product, and 4) Efficient correction of errors by oligonucleotide-directed mutagenesis following cloning and sequencing of the assembled product.

To drive development of improved methodology we are undertaking some larger synthetic projects. We are assembling the complete genome of the mouse mitochondrion. We are also synthesizing a region from the *M. genitalium* genome that encodes genes essential for translation of mRNA to produce proteins.

## 44

## Development of a *Deinococcus radiodurans* Homologous Recombination System

Sanjay Vashee\*, Ray-Yuan Chuang\* (RChuang@venterinstitute.org), Christian Barnes, Hamilton O. Smith, and J. Craig Venter

J. Craig Venter Institute, Rockville, MD

---

A major goal of our Institute is to rationally design synthetic microorganisms that are capable of carrying out the required functions. One of the requirements for this effort entails the packaging of the designed pathways into a cohesive genome. Our approach to this problem is to develop an efficient in vitro homologous recombination system based upon *Deinococcus radiodurans* (Dr). This bacterium was selected because it has the remarkable ability to survive 15,000 Gy of ionizing radiation. In contrast, doses below 10 Gy are lethal to almost all other organisms. Although hundreds of double-strand breaks are created, Dr is able to accurately restore its genome without evidence of mutation within a few hours after exposure, suggesting that the bacterium has a very efficient repair mechanism. The major repair pathway is thought to be homologous recombination, mainly because Dr strains containing mutations in *recA*, the bacterial recombinase, are sensitive to ionizing radiation.

Since the mechanism of homologous recombination is not yet well understood in Dr, we have undertaken two general approaches to study this phenomenon. First, we are establishing an endogenous extract that contains homologous recombination activity. This extract can then be fractionated to isolate and purify all proteins that perform homologous recombination. We are also utilizing information from the sequenced genome. For example, homologues of *E. coli* homologous recombination proteins, such as *recD* and *ruvA*, are present in Dr. Thus, another approach is to assemble the homologous recombination activity by purifying and characterizing the analogous recombinant proteins. However, not all genes that play a major role in homologous recombination have been identified by annotation.

We have over-expressed and purified *D. radiodurans* homologues of repair proteins, including RecA, SSB, RecD, RuvB, and RuvC. We showed that the properties of DrSsb DNA binding and strand-exchange properties are very similar to that of *E. coli* SSB. In addition, using antibodies we have raised, we have determined that the amount of both DrSsb and RecA protein increases in the cell when exposed to a DNA damaging agent, whereas the level of DrRecD protein remains the same.

## 45

## Development of a Novel Recombinant Cyanobacterial System for Hydrogen Production from Water

Qing Xu, Shibu Yooseph, Hamilton O. Smith, and J. Craig Venter (jcventer@tcag.org)

J. Craig Venter Institute, Rockville, MD

---

Hydrogen is a clean alternative to gasoline and other fossil fuels as it generates only water as a by-product. The development of cost-effective and renewable approaches to produce hydrogen fuel will lead to a new and efficient energy system, which will address both the adverse environmental impacts of fossil fuels and the need for energy independence. Photobiological processes are attractive routes to renewable hydrogen production. With solar energy, photosynthetic microbes such as cyanobacteria can extract energy from water via oxygenic photosynthesis. The resulting energy can be coupled to a hydrogenase system that yields hydrogen. However, one major drawback of this process is that most hydrogen-evolving hydrogenases are irreversibly inhibited by oxygen, which is an inherent byproduct of oxygenic photosynthesis.

The overall goal of our project is to develop a novel, O<sub>2</sub>-tolerant photobiological system in cyanobacteria that can produce hydrogen continuously using water as the substrate. We have undertaken two general approaches to achieve this goal. Our first approach is to transfer known O<sub>2</sub>-tolerant NiFe-hydrogenase into cyanobacteria. It is reported that purple-sulfur photosynthetic bacterium *Thiocapsa roseopersicina* has an oxygen-tolerant evolving NiFe-hydrogenase, but this anoxygenic microbe can not split water. It is therefore logic to transfer the hydrogenase into cyanobacteria to construct a novel hybrid system that physically combines the most desirable properties of two bacteria. Thus far, we have cloned O<sub>2</sub>-tolerant hydrogenase genes *hydS* and *hydL* from *T. roseopersicina* into a cyanobacterial expression vector to create plasmid pEX-Tr, which was then transformed into *Synechococcus* sp. PCC7942. Analysis of pEX-Tr *Synechococcus* transformants are currently in the process. Our second approach is to identify putative O<sub>2</sub>-tolerant NiFe-hydrogenases from marine microbes and transfer them into cyanobacteria. Screening hydrogenase genes from the environment has been a useful approach to find novel O<sub>2</sub>-tolerant hydrogenases. However, to date, only a few O<sub>2</sub>-tolerant hydrogenases have been identified. To take advantage of the environmental genetic information generated by our ongoing Global Ocean Sampling Project, we currently use probabilistic modeling methods such as HMMs to search our sequence databases for putative NiFe-hydrogenases. We will transfer them into cyanobacteria for heterologous expression, and then screen for novel O<sub>2</sub>-tolerant hydrogenases using a chemochromic screening method developed by scientists at NREL. Thus far, we have identified a putative NiFe-hydrogenase that shows strong homology to a known O<sub>2</sub>-tolerant hydrogenase. Cloning the genes is in the process.

## 46

**Biotechnology For the Production of Ethanol and Butanol from Cellulose**

Prabha P. Iyer\* (piyer@venterinstitute.org), Hamilton O. Smith, and J. Craig Venter

J. Craig Venter Institute, Rockville, MD

---

Cellulose is the largest component of all biomass and is the most abundant organic polymer in nature. It also has the potential to be fermented into a number of useful products including ethanol and butanol. This has important implications for the production of renewable, biomass-based fuels. However, the crystalline structure of cellulose makes it a challenging substrate for most bacteria to break down. Further, most of the bacteria that do degrade cellulose are not able to produce large amounts of ethanol or butanol. The goal of this project is to construct a microorganism which can break cellulose down to glucose monomers and then ferment the glucose to appreciable quantities of alcohols. This would eliminate the current requirement for large amounts of enzymes to depolymerize cellulose before it can be fermented to alcohols.

We are in the process of cloning and heterologously expressing the entire cellulosome gene cluster of *Clostridium cellulolyticum* in *Clostridium acetobutylicum*. *C. cellulolyticum* is able to degrade crystalline cellulose; however, it is not able to produce large amounts of ethanol. On the other hand, *C. acetobutylicum* is a very well characterized organism which has traditionally been used to produce solvents in acetone-butanol-ethanol fermentations. Production of a cellulosome in *C. acetobutylicum* would constitute a major advance towards consolidated bioprocessing of cellulose to ethanol.

Expression of the cellulosome in organisms capable of producing butyrate such as *Clostridium butyricum* and *Clostridium tyrobutyricum* is also being explored. This would facilitate the degradation of cellulose to butyrate which can then be converted into butanol via a second fermentation reaction.