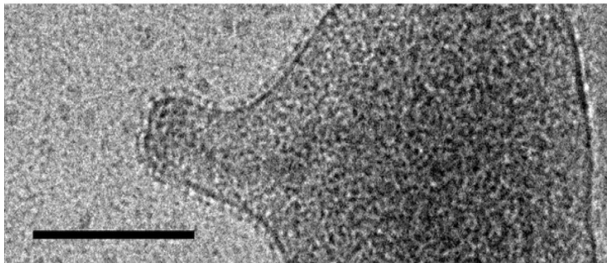
Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome

(Gibson, 2008)

Lauren Primer & Kylie Snyder
November 1, 2018
ABE 591 - Journal Club

Background

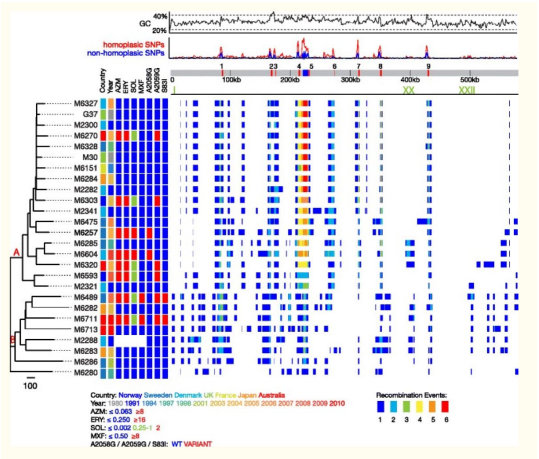
- Bacterium: *Mycoplasma genitalium*
  - Smallest genome of any independently replicating cell grown in pure culture
  - Lab use is limited to ~100/485 genes that are non-essential
- Ultimate goal:
  - Which genes are simultaneously dispensable?
  - Produce compact genomes by chemical synthesis, assembly, and cloning
  - Introduce them into cells to test viability of essential genetic functions for life



Wild type cells showing cytoskeleton supporting terminal organelle and nap layer surrounding cell membrane

Main Objectives

- Design, chemical synthesis, assembly, and cloning (in yeast) of entire *M. genitalium* JCVI-1.0 chromosome based on *M. genitalium* G37
  - Would be the largest chemically synthesized molecule of defined structure with 582,970 bp total
  - Except MG408 - disrupted by an antibiotic marker to block pathogenicity and allow for selection



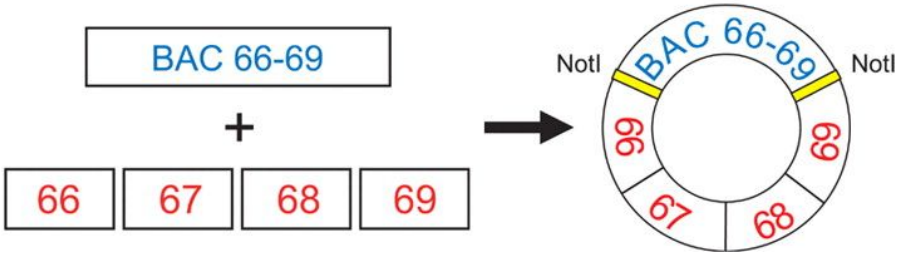
Phylogenetic tree of *M. genitalium*

Barriers

- Completed synthetic genomes were strictly viral
- **Mycoplasma = bacterial**
  - Able to reproduce without a host
- Magnitude
  - 32,000 bp (viral) vs. 582,970 bp
  - Larger sections of DNA need to be handled and may be unstable in *E. coli*

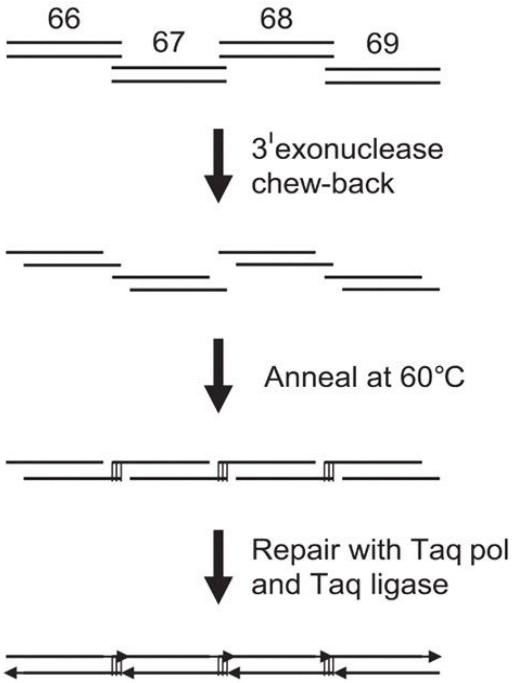
New Methods

1. In vitro recombination using *E. coli*
    - *M. genitalium* sequence was split into 101 cassettes
        - Each ~5-7 kb
    - Boundaries placed at intergenic sites
    - Contains one or more genes
    - Cassette 101 overlaps 1 -> loop

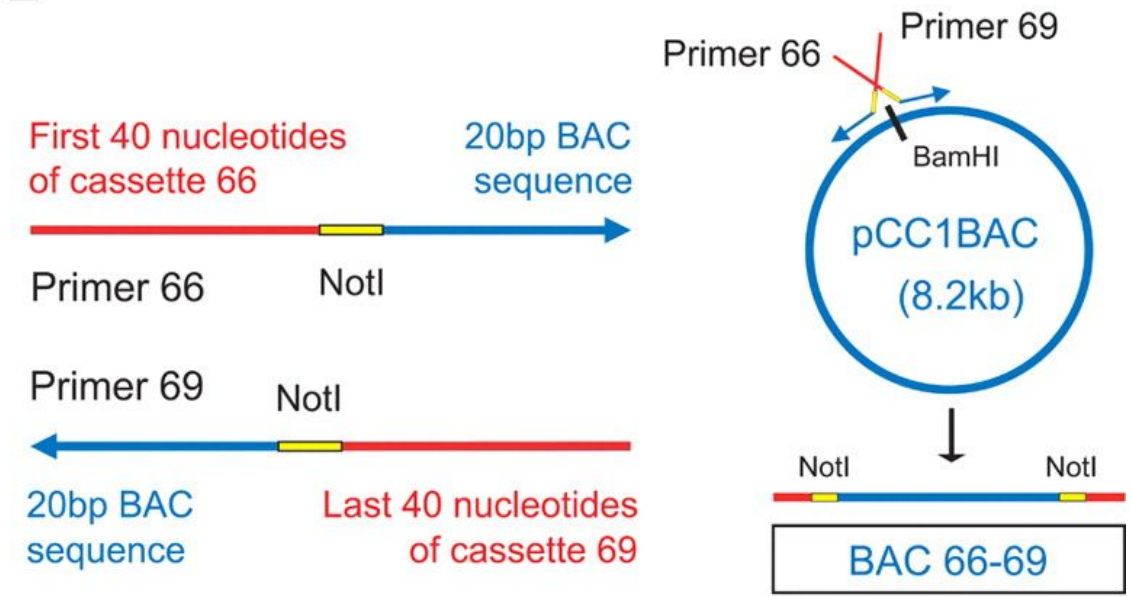


In Vitro Recombination Reaction

- Gibson method***
- Overlapping DNA molecules are digested with a 3’ exonuclease to expose overhangs
- Complementary overlaps are annealed
- Joints are repaired via ligase



Assembly of Synthetic Cassettes



- PCR amplification used to produce BAC vector for cloning each assembly
- Cassettes assembled 4 at a time in presence of appropriate BAC vector
- Not I = restriction site

Assembly Problems

- In vitro (*E. coli*) -> in vivo (yeast): difficulties in executing the planned assembly and cloning of the half and whole synthetic genomes in E. coli
  - Conclusion: larger assemblies were unstable in *E. coli*
    - Inefficient circularization of large DNA molecules
    - Breakage during the handling of DNA before transforming the E. coli
  - Solution: Turn to yeast, *S. cerevisiae*, as cloning host using TAR

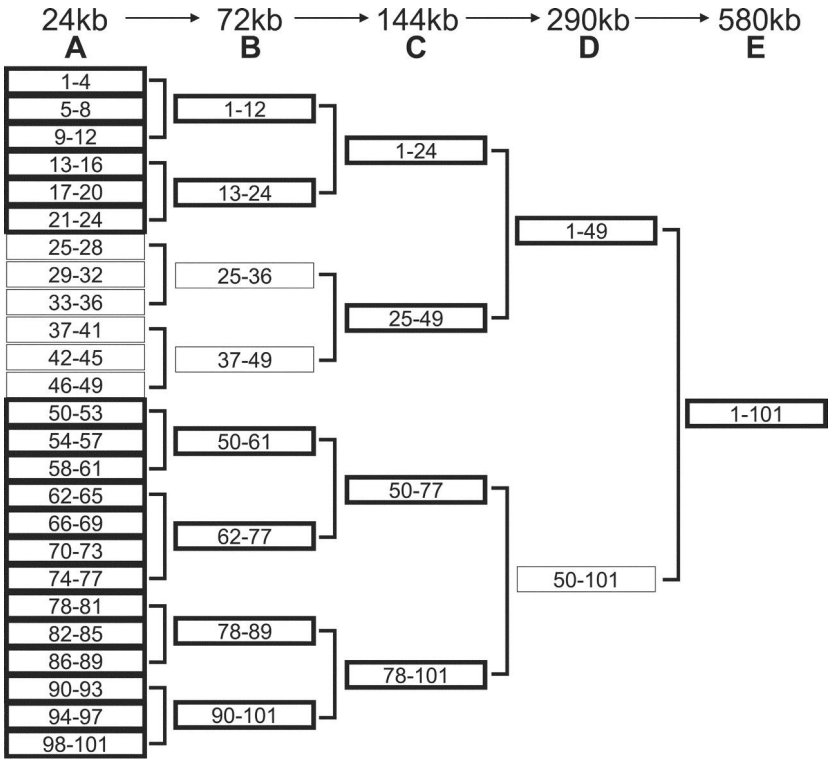
New Methods

2. In vivo, transformation-associated recombination (TAR) using *S. cerevisiae*

    - TAR cloning implemented TARBAC vector to propagate synthetic chromosome in yeast
        - Isolation for sequence verification
    - Production outsourced for large DNA synthesis
    - Blue Heron Technology, DNA 2.0, GENEART

What is required to assemble constructs in this manner?

Assembly Plan

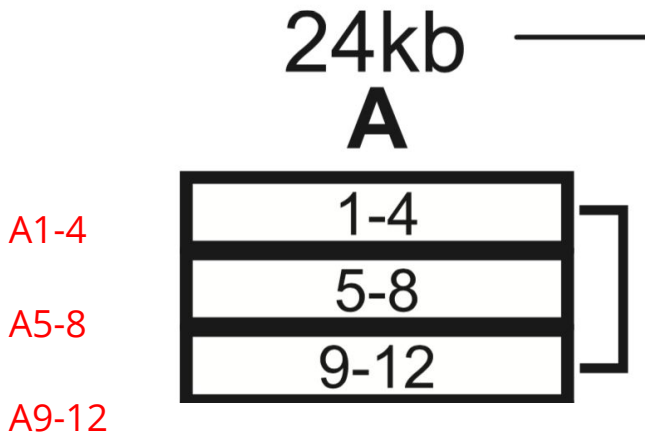


- Phases of combination
- A → C
  - In vitro recombination
- D → E
  - In vivo yeast recombination
  - Larger assemblies not stable in *E. coli*

*D and E series ultimately omitted

Stage 1 - In Vitro Recomb. , *E. coli*

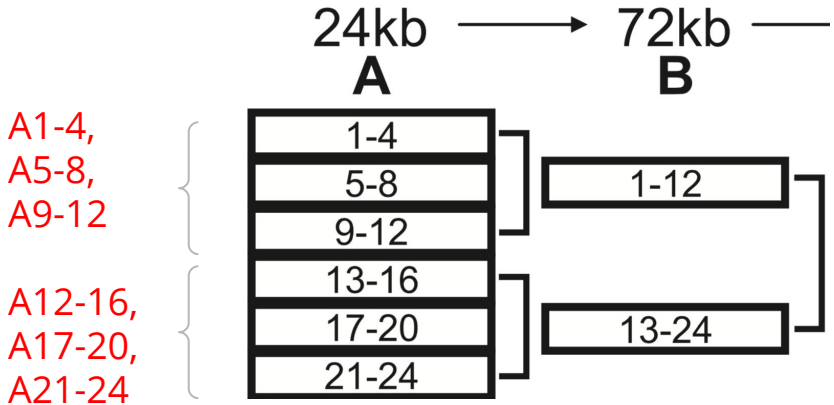
- Sets of 4 neighboring cassettes assembled by in vitro recombination
- Joined to a bacterial artificial chromosome (BAC) vector DNA to form circularized recombinant plasmids with ~24-kb inserts
  - Ex: cassettes 1-4 form A1-4
  - Ex: cassettes 5-8 form A5-8



4 neighboring cassettes form a set

Stage 2 - In Vitro Recomb. , *E. coli*

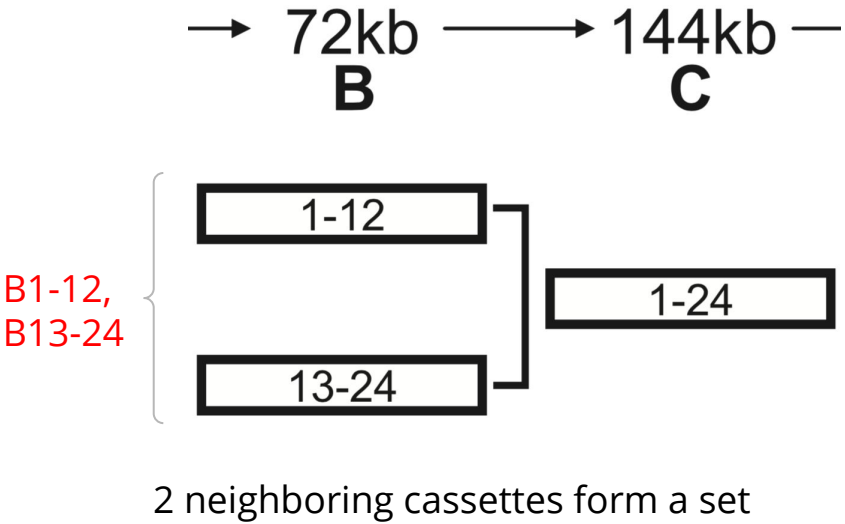
- The 25 A-series assemblies taken 3-at-a-time to form 8 B-series assemblies
- Each 1/8 the size of a genome at ~72 kb
  - Ex: A1-4, A5-8, A9-12 form B1-12



3 neighboring cassettes form a set

Stage 3 - In Vitro Recomb. , *E. coli*

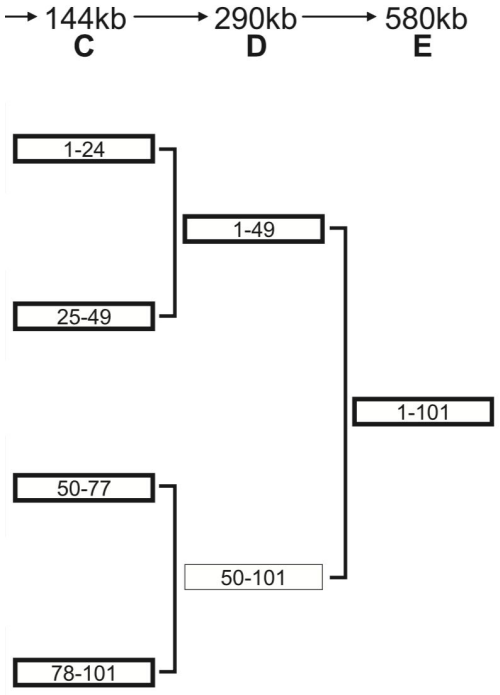
- The 1/8 genome B-assemblies were taken 2-at-a-time to form 4 C-assemblies
- Each 1/4 of genome at ~144 kb
  - Ex: B1-12 and B13-24 form C1-24



2 neighboring cassettes form a set

Stages 4 and 5 - Ideal In Vitro Recomb., *E. coli*

- Initial plan for Stages 4 and 5
  - C-assemblies would be joined 2-at-a-time to form 2 D-assemblies
  - Remaining 2 D-assemblies would form E-assembly
- For the final molecule, the D-series half molecules were not employed.
  - Rather, the whole molecule was comprised of 4 C-series quarter molecules.
- PCR Amplification → BAC vector
- FIGE Analysis/gel electrophoresis: unviable b/c too large, unstable



D and E assemblies

Stages 4 and 5 - Actual In Vivo Recomb. (TAR), *S. cerevisiae*

- Linear YAC Cloning:
  - Produced via ligation of an insert into a restriction enzyme cloning site
- TAR Cloning:
  - Cotransformation of overlapping insert and vector DNAs into yeast spheroplasts -> joined by homologous recombination -> circular clones
- Linear YAC Clones vs. TAR Clones
  - Both contain a centromere and maintained at a copy # along with a native yeast genome
  - TAR Advantage: circular TAR clones can be separated from linear yeast chromosomes
- To assemble quarter genomes into halves/wholes in yeast:
  - Use pTARBAC3 vector, containing YAC and BAC sequences
    - Construction similar to that of BAC vectors but with longer 60-bp overlaps at termini

Stages 4 and 5 - Actual In Vivo Recomb. (TAR), *S. cerevisiae*

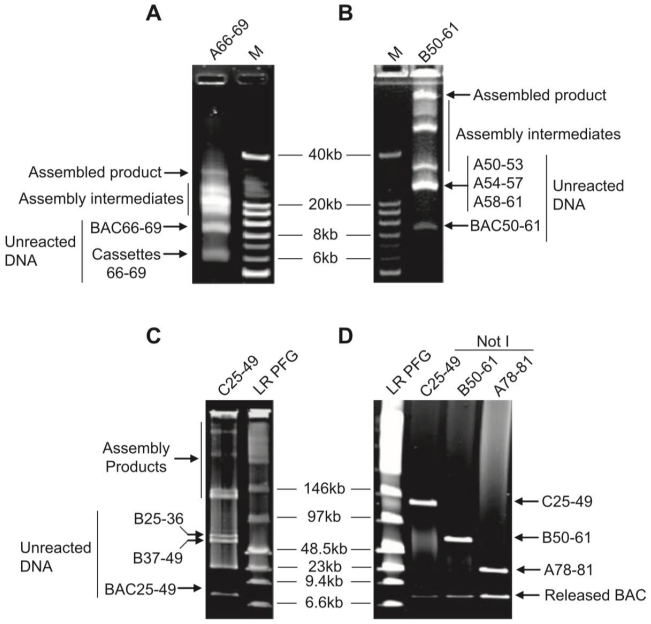
- TAR cloning:
  - Recombination stimulated by a factor of 20 at double-stranded breaks
    - Vector integrated at the cleaved intergenic BsmB I site in C50-77
    - 4 bases of the BsmB I 5' overhang were eliminated
  - DNA to be transformed had 6 pieces
    - 1 vector, 2 fragments of quarter 3, quarters 1, 2 and 4
  - To obtain a full-sized genome as an insert in pTARBAC3 vector, a single yeast cell must take up all 6 pieces and assemble via homologous recombination

Stages 4 and 5 - Actual In Vivo Recomb. (TAR), *S. cerevisiae*

- Transformation of yeast cells:
  - Equimolar amounts of vector and inserts
    1. PCR screening
    2. Southern blot with mycoplasma-specific probes
  - Positive clones tested for stability via Southern blotting of subclones
    - At least 17/94 transformants carried a complete synthetic genome
      - Clone sMgTARBAC37 was selected for sequencing
  - TAR cloning with each of the 4 sets of 2 adjacent quarter genomes
    - DNAs from transformants isolated and electroporated into E. coli
    - BAC clones of expected sizes were obtained
      - DI-49 was sequenced -> verified

Error Management

- Differences between actual and designed sequence can arise
  - Error in sequence supplied to contractors
  - Contractors could produce errors
  - During repair of assembly junctions
  - From propagations in *E. coli* or yeast at later stages
- All errors found
- Clones were sequenced for verification



Field Inversion Gel Electrophoresis (FIGE) analysis of assemblies

Overlap with Class (human practices, responsibility)

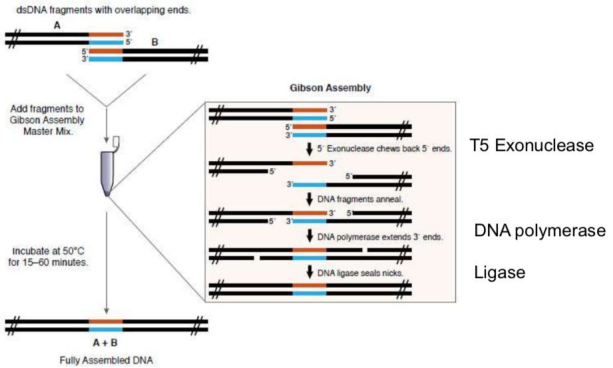
- Genome assembly: construct and clone large DNA molecules from chemically synthesized pieces and combinations of natural and synthetic DNA segments larger than those previously reported
  - Constructs may be delivered by plasmids or host genomes
  - To form circuits:
    1. Chemical addition (chem-synthesized oligonucleotides)
    2. Homologous recombination (recombineering: in vitro, in vivo)
    3. Anneal + polymerization (ability of single stranded DNA to anneal (stick) to complementary strands - in vitro, PCR amplification)
    4. Digestion + ligation (Type IIS REs defined overhang in in vivo, create 'sticky' overhangs -> annealed to join DNA pieces)
  - Recombineering: recombination cloning, isolation, and sequencing techniques used to chemically synthesize a full synthetic genome

Overlap with Class (human practices, responsibility)

- Gibson Assembly was developed in this paper (Week 9 Lecture)

Gibson Assembly

Workflow

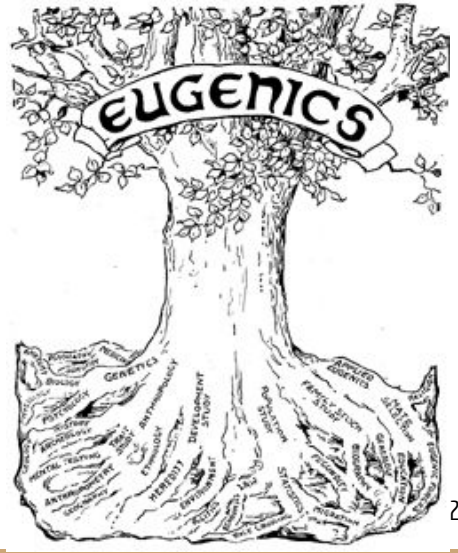


Advantages	Disadvantages
• Parts don't need to be formatted (sequence independent) • Can make combinatorial libraries (**up to 5 parts**) • One-pot reaction (fast/simple) • Scarless	• Requires long primers that may be challenging • Expensive

Foundational Principles Used

- Complete DNA synthesis
- Construction of genetic circuits
  - Parts
    - *M. genitalium*
    - *S. cerevisiae*
    - TARBAC vector
    - BAC vector
  - Assembly methods
    - In vitro recombination
    - In vivo recombination

- Responsible Design
  - Eugenics?
  - Biosafety?



Main Conclusions

- Based on *M. genitalium* G37, a 582,970-bp *Mycoplasma genitalium* JCVI-1.0 genome was designed, chemically synthesized, assembled, and cloned
  - Final product: ~10^4 synthetic oligonucleotides
- Challenge: obtain DNA molecules larger than any previously reported using a 5-stage assembly
  - Stages 1-3: in vitro recombination with *E. coli* -> assembly too large, unstable
  - Stages 4-5: in vivo transformation-associated recombination (TAR) with yeast as host
- Larger than any other chemically synthesized DNA product and largest chemically synthesized molecule of defined structure
  - Previously large non-synthetic constructs have been produced from bacterial genomic DNA and in vivo methods

Advantages/Limitations

- Advantages:
  - Previous methods for large DNA use restriction enzymes (ex: type IIS) to generate sticky ends
    - As pieces get larger, it becomes more difficult to find Type IIS REs that do not cleave within piece
    - Can use enzymes (Not I) that cleave infrequently
  - 4 quarters recombined and cloned to yield full genome
    - Ability to assemble many DNA pieces in a single reaction -> useful to generate combinatorial genome libraries
  - Methodology not restricted to chemically synthesized DNA
    - Simply need to design PCR primers with appropriate overlaps

Advantages/Limitations

- Limitations:
  - In vitro recombination: efficiency declined as assembly size increased (only 1/4 genome could be achieved) -> had to resort to in vivo TAR
  - Use of TARBAC vector to propagate synthetic chromosome in yeast
    - Unknown if...
      - Vector may interfere with production of viable cells by transplantation
      - Genomic location of the vector could affect viability
      - Vector sequences/excise vector must be altered before transplantation
  - Such large DNA synthesis -> outsourced production to contractors -> errors
  - Contains all genes of wild-type *M. genitalium* G37 *except* Gene MG408 (2514-bp insertion into cassette 89) containing aminoglycoside resistance gene
    - A strain with this defect in this virulence factor cannot adhere to mammalian cells -> eliminate pathogenicity in the best available model systems

General Lessons

- Possible to re-engineer recombineering
- Future:
  - Ability to assemble many pieces of DNA in single step
  - Methodology not restricted to chemically synthesized DNA
- Natural errors and mutations
  - *E. coli* clones had errors → Rebuild assemblies
  - Sequence verification
- Novel ways to clone genomes

Thank you