Hiram Genomics Initiative: Novel Genomics Research Where You Least Expect It

Cathy Wheeler, Prudy Hall, & Brad Goodner

Department of Biology
Hiram College, OH 44234
wheelerca@hiram.edu, hallp@hiram.edu, goodnerbw@hiram.edu

Cathy Wheeler is a Laboratory Teaching Associate responsible for the laboratory portions of genetics, cell and molecular biology, and microbiology courses. She is a co-founder of the Hiram Genomics Initiative where she incorporates research into the laboratory portions of all her courses and runs the outreach component to local high schools.
Prudy Hall is a Professor responsible for courses in botany, cell and molecular biology, introductory and advanced biochemistry, and plant physiology. She is a co-founder of the Hiram Genomics Initiative where she incorporates research into the laboratory portions of her biochemistry courses. Her research interests have recently focused on proteomics of plant-associated bacteria.
Brad Goodner is an Associate Professor responsible for courses in genetics, cell and molecular biology, microbiology, immunology, and bioinformatics. He is a co-founder of the Hiram Genomics Initiative where he administers the overall project, incorporates research into the laboratory portions of all his courses, and disseminates best practices to colleagues at other colleges and universities. His research interests include using the genomic sequence of *Agrobacterium* to better understand its biology and involving undergraduates in all aspects of microbial genome projects.

Reprinted From: Wheeler, C., P. Hall, and B. Goodner. 2004. Writing assignments in large classes with minimal support. Pages 365-371, *in* Tested studies for laboratory teaching, Volume 25 (M. A. O'Donnell, Editor). Proceedings of the 25th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 414 pages.

Overview

The Hiram Genomics Initiative is a long-term, self-perpetuating learning environment based at Hiram College. The overarching goals of the initiative are to involve students in novel research projects as part of the classroom in order to teach basic scientific principles and to excite students about research and a biology education at Hiram College. The means used to achieve these goals are the incorporation of cutting-edge genomics and microbiology research into the laboratory portion of several Hiram College courses, the establishment of research collaborations with nearby high schools, and the

establishment of an internal journal for communication of research results. During the past 1.5 years, students in five courses at Hiram College (cell & molecular biology, genetics, advanced molecular biology (bioinformatics), microbiology, & biochemistry II) and advanced biology courses at Benedictine (Cleveland, OH), Gilmour Academy (Gates Mills, OH) and Walsh Jesuit (Cuyahoga Falls, OH) high schools were involved along with eight faculty members. Here we describe three examples of our efforts in hopes of encouraging the incorporation of original research into the laboratory portions of biology courses at other institutions.

Construction of Combined Genetic/Physical Map of *Halomonas elongata* Genome in a Genetics Course

In collaboration with scientists from Purdue University, we are determining the complete genome sequence of the gamma-proteobacterium *Halomonas elongata* DSM3043 (in the same bacterial division as *Escherichia coli* and *Pseudomonas aeruginosa*). What makes *H. elongata* so interesting to study, compared to its better-known relatives, is its ability to live in the presence of a lot of salt – it requires at least 0.5 M NaCl to grow and can tolerate over 3 M NaCl (Vreeland *et al.*, 1980; Cánovas *et al.*, 1997). To better understand the halophilism and halotolerance of *H. elongata*, it would be useful to know all of the proteins made by the organism and their potential functions. Getting the entire genome sequence is just the first step, but it is not trivial (Figure 1). The genome has to be broken into manageable-size pieces, then each piece cloned and sequenced. Scientists at Purdue University will sequence several thousand random pieces of the *H. elongata* genome. Computer analysis will allow the small pieces to be joined at sequence overlaps into larger pieces (contigs). Our current part of the collaboration is to help assemble the contigs into the complete genome sequence, just like putting together a jigsaw puzzle. Having some reference points (e.g., printed design or shape of puzzle pieces) can help the assembly process tremendously, so we are generating a combined genetic and physical map of the *H. elongata* genome using a transposon mutagenesis strategy that translates well into the laboratory portion of a genetics course. This multi-week strategy is a vehicle for students learning a wide variety of skills and techniques (mutagenesis, DNA isolation, transformation, restriction mapping) while also accomplishing an original research goal that can be adapted to a wide range of bacteria.

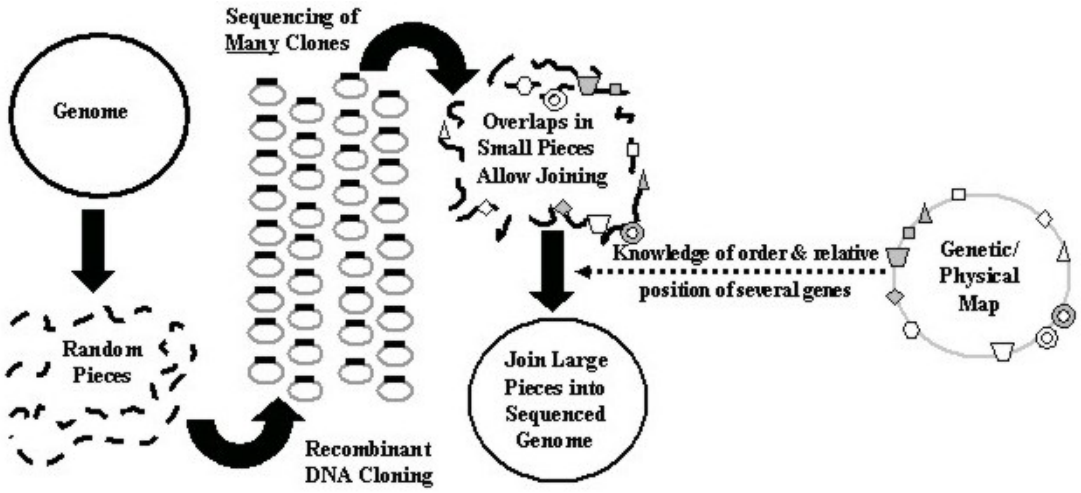

Figure 1. Basic steps of a genome sequencing project: making a shotgun genomic library, extensive sequencing coverage, computer-based assembly of sequence reads into contigs, using other information such as a genetic/physical map to orient contigs and close gaps.

In the first step of generating a genetic/physical map, students mutagenize a rifampicin-resistant derivative of *H. elongata* using the modified transposon Tn*5*RL27 (Figure 2A). This mini-transposon is a suicide derivative of Tn5 that contains a kanamycin resistance marker, the *pir*-dependent R6K origin of replication, and several rare restriction enzyme cut sites (Larsen et al., 2002). This transposon allows us to not only generate mutations, but also to physically map the location of the transposon and to easily recover the transposon with flanking genomic DNA for sequencing of the insertion site. After overnight mating of *H. elongata* with the *E. coli* strain harboring pRL27 (contains Tn5RL27 and hyperactive transposase gene) on LB plates containing 0.5M NaCl, the mating mixture is diluted and plated onto LB plates containing 0.5M NaCl, rifampicin (20 µg/ml, kills off the *E. coli* donor), and kanamycin (50 µg/ml, selects for presence of transposon) (Figure 2B).

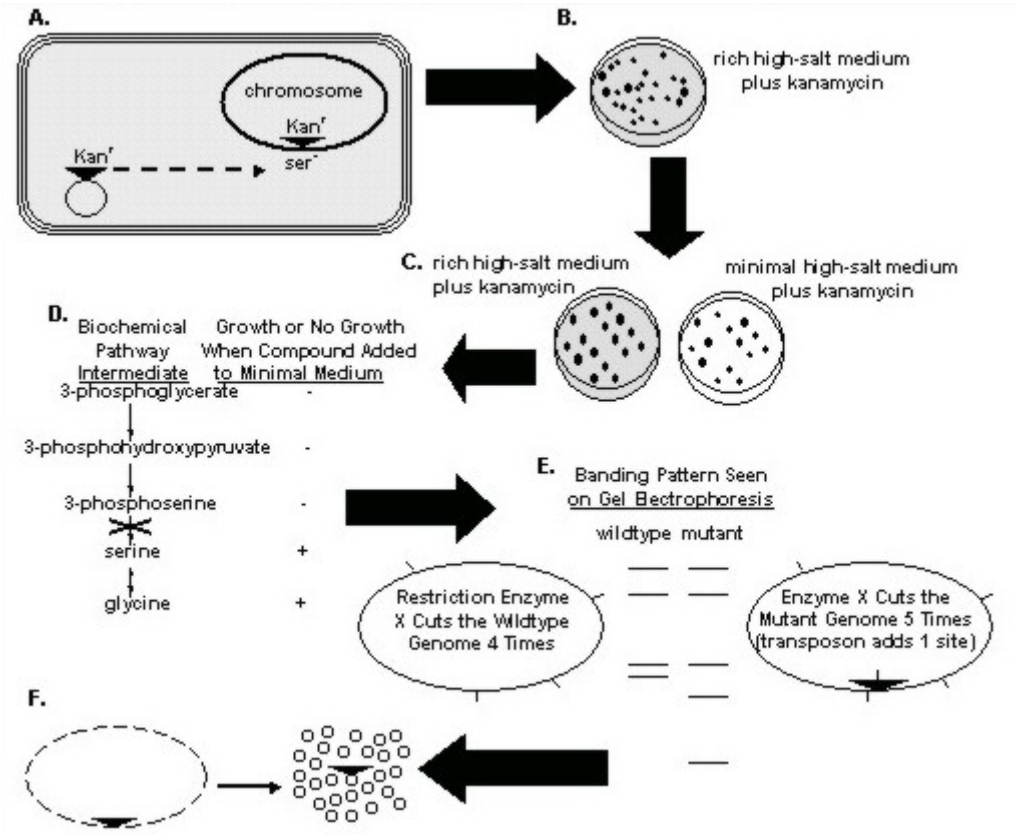

Figure 2. Steps in generation of a combined genetic/physical map of a bacterial genome. A, B: Generation and selection of transposon insertion mutants. C: Screening mutants for auxotrophy. D: Further biochemical complementation analysis of auxotrophic mutants. E: Physical mapping of transposon insertion sites using pulsed-field gel electrophoresis. F: Recovery of genomic fragment containing transposon insertion.

Second, students screen for auxotrophs by patching transposon insertion mutants onto LB and M9 minimal plates, each containing 0.5M NaCl, rifampicin (20 µg/ml), and kanamycin (50 µg/ml) (Figure 2C). While we focus on auxotrophs (grow on LB, but not on M9), please realize that the prototrophic insertion mutants can also be physically mapped and the insertions recovered for sequencing.

Third, students further characterize the auxotrophs through biochemical complementation (Figure 2D). Overlapping pools of essential building blocks can be used to determine the pathway blocked in each mutant and where possible, intermediates from the pathway can be used to determine the site of the pathway block.

Fourth, to physically map the location of the transposon in each mutant (auxotroph or prototroph), students make genomic DNA plugs in agarose (of the type needed for pulsed field gel electrophoresis), cut the DNA with a rare-cutting restriction enzyme that also cuts inside the transposon, and separate the genomic DNA fragments by pulsed-field gel electrophoresis (Figure 2E). The genomic fragment into which the transposon hopped will now be cut into two pieces. Doing this type of analysis with more than restriction enzyme and many different mutants allows for the construction of a map.

Finally, the site of transposon insertion in each mutant can be recovered for sequencing very easily by cutting genomic DNA from a mutant with a restriction enzyme that does not cut in the transposon, diluting the DNA and carrying out a self-ligation to generate circles, transforming the ligation mixture into an *E. coli* strain expressing the *pir* gene, and selecting for kanamycin-resistant transformants. The *pir*-dependent R6K origin of replication and kanamycin resistance marker on Tn*5*RL27 allow any genomic fragment to be turned into a plasmid. Sequencing off of the ends of the transposon allows for the putative identification of the mutated gene using sequence similarity to known genes in GenBank.

Over a 5-6 week period, students in a genetics lab can easily carry out this entire project, screening thousands of mutants and characterizing tens of auxotrophs. While the sequencing may have to be done off-site and later, it can be the subject of separate analysis by the next semester of genetics students. Outside of the sequencing, the costs are comparable to that for a series of separate labs on cloning, mutagenesis, and restriction analysis. Carrying out this project over 3-4 repetitions of a genetics course can generate enough data for a publishable genomic map.

Creating Knock-Out Mutants of *Agrobacterium tumefaciens* C58 in an Introductory Molecular and Cellular Biology Course

Agrobacterium tumefaciens is a common soil bacterium that causes a unique disease on many plants. If the bacterium can access a wound site, it attaches to dividing plant cells and injects a piece of its own DNA into a plant cell where the DNA becomes integrated into the plant's genome. Expression of genes on the transferred DNA causes the plant cell to make lots of its own growth-stimulating hormones, leading to the development of a tumorous gall on the plant. Also, the transformed plant cells produce compounds that *A. tumefaciens* can use as food. Much is known about *A. tumefaciens*, including the complete genomic sequence of one strain (Goodner *et al*., 2001; Wood *et al.*, 2001). However, the genomic sequence is also a starting point for functional studies such as enzyme assays. In our molecular and cellular biology course, we have students make gene disruptions for selected genes and then use these mutants to answer questions about the roles of the encoded proteins.

We usually choose the genes that will be targeted in a given semester based on the availability of associated enzyme or physiological assays and/or on the basis of their biological interest to us. For each gene, we design a pair of PCR primers that will allow the amplification of a 300-500 bp internal portion of the gene. If time allows and one wishes to address PCR primer design in the course, there is no reason why students could not design their own PCR primers using Web-accessible software such as Primer 3 (<*http://www-genome.wi.mit.edu/cgi-*

bin/primer/primer3_www.cgi>). We usually assign one gene to each group of 2-3 students. For example, in one recent semester, we targeted the *A. tumefaciens* genes encoding catalase, lysine decarboxylase, nitrate reductase (assimilatory), survival protein SurE, and urease.

Over the course of three weeks early in the semester, students are introduced to the concepts of DNA isolation, PCR, cloning, transformation, electrophoresis, restriction analysis and experimental design as a part of this project. Students first amplify their gene segment using PCR, check their PCR using gel electrophoresis, and clone their PCR product into a plasmid (we use pCR2.1-TOPO from Invitrogen). The students then isolate plasmid DNA from their transformants and carry out restriction mapping to confirm that they have cloned their PCR product.

As the final step in this part of the lab course, students introduce their plasmid containing a PCR product into wildtype *A. tumefaciens* via electroporation and select for the presence of the plasmid (carbenicillin resistance works well for PCR2.1-based plasmid constructions). Here, the choice of PCR product and plasmid come in handy. Using a plasmid that replicates in *E. coli*, but not in *A. tumefaciens* allows us to select for recombination events between the wildtype gene of interest and the cloned PCR product carried by the plasmid. The fact that the PCR product is an internal portion of the gene of interest means that a single crossover recombination event generates a gene disruption (Figure 3). Patience is crucial because it takes several days for these recombination-based transformants to grow up on the selection plates.

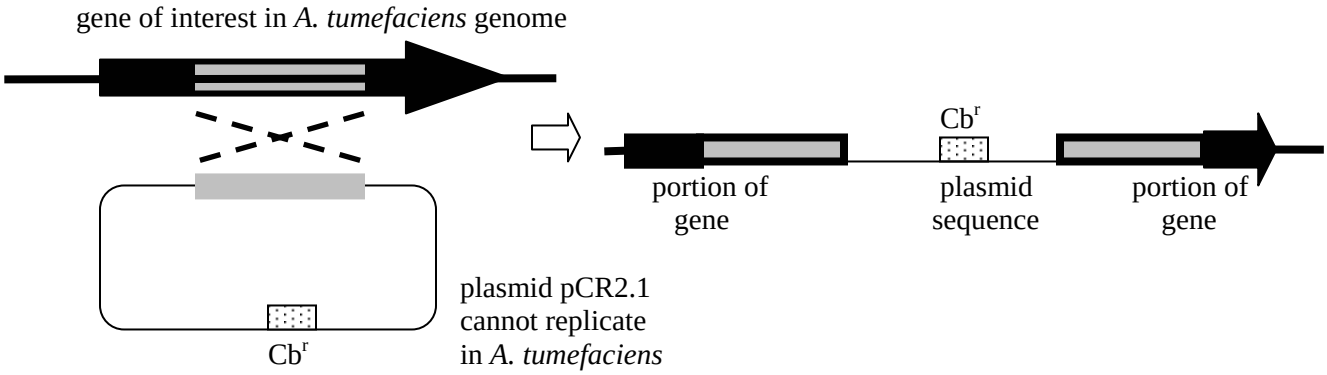

Figure 3. Generation of a gene disruption mutation using a cloned internal portion of the gene of interest.

Later in the semester, student groups brainstorm on how to assay the function of the protein encoded by their gene of interest. Such assays might involve colorimetric measures of enzyme activity, growth curves in different media, or other physiological tests. Student groups submit lists of needed materials and help prepare some of the materials if possible. We have found that having two weeks of lab open for the subsequent independent experiments gives each group time to try out an assay or test, optimize it, and/or follow up with additional experiments. Two examples of such experiments are shown in Figure 4. Finally, student groups report on their projects in a poster session. This project never gets repetitive because new genes can be disrupted in each reiteration of the course.

A. **B.**

Figure 4. Representative results from functional tests of gene disruption mutants. A: Tumorigenesis assay on carrot disks of wildtype, surE⁻ mutant, and catalase⁻ mutant derivatives of *A. tumefaciens* C58. The avirulence of the catalase⁻ mutant agrees with previous results (Xu and Pan, 2000). B: The urease⁻ mutant grows better with nitrate as a N source aerobically than does the assimilatory nitrate reductase⁻ mutant (some growth due to the remaining bifunctional NAP-type nitrate reductase), but both grow to same extent anaerobically due to remaining NAP-type nitrate reductase.

Genomic Sequence Annotation as Part of Genetics and Advanced Molecular Biology Courses

A complete genome sequence looks like someone aimlessly banged away at a typewriter with only 4 keys (A, C, G, and T) until you understand its language. The sequence can tell us what proteins an organism makes and comparison of the predicted proteins to known proteins from other organisms can allow us to hypothesize as to their biochemical functions. Ultimately, sequence analysis followed up by experiments will tell us when each protein is made, where it goes in the cell, and its roles in the biology of the organism. Knowing how to use some of the common Web-based sequence annotation tools and interpret the results is and will continue to be crucial across all of biology. Students at Hiram College first learn the basics of bioinformatics in the genetics course, where they work in pairs to annotate a 1 kbp section of the *Sphingomonas elodea* or *Halomonas elongata* unfinished genome. In our advanced molecular biology, an intensive short course on bioinformatics, students learn to use a variety of DNA and protein sequence analysis programs (as well as write some scripts of their own).

Then, each student analyzes a different 50,000 bp stretch of assembled genome sequence. They determine the location of putative genes, the identity of the encoded proteins, and propose experimental tests of function for some of the encoded proteins. For example, one gene from *S. elodea* encodes a strong bacterioferritin homolog. The student who annotated this gene determined, from a literature search, that bacterioferritin is involved in iron homeostasis. He proposed that confirmation of this gene annotation could involve generation of a mutation in this gene. He hypothesized that such a mutant would not grow in iron-deficient medium and/or medium with excess iron. All of this work is presented in a poster session format.

Literature Cited

Canovas, D., C. Vargas, F. Iglesias-Guerra, L. N. Csonka, D. Rhodes, A. Ventosa, and J. J. Nieto. 1997. Isolation and characterization of salt-sensitive mutants of the moderate halophile *Halomonas elongata* and cloning of the ectoine synthesis genes. Journal of Biological Chemistry 272:25794-25801.

Goodner, B., G. Hinkle, S. Gattung, N. Miller, M. Blanchard, B. Qurollo, B. S. Goldman, Y. Cao, M. Askenazi, C. Halling, L. Mullin, K. Houmiel, J. Gordon, M. Vaudin, O. Iartchouk, A. Epp, F. Liu, C. Wollam, M. Allinger, D. Doughty, C. Scott, C. Lappas, B. Markelz, C. Flanagan, C. Crowell, J. Gurson, C. Lomo, C. Sear, G. Strub, C. Cielo, and S. Slater. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323-2328.

Larsen, R. A., M. M. Wilson, A. M. Guss, and W. W. Metcalf. 2002. Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. Archives of Microbiology 178:193-201.

Vreeland, R. H., C. D. Lichtfield, E. L. Martin, and E. Elliot. 1980. *Halomonas elongata*, a new genus and species of extremely salt-tolerant bacteria. International Journal of Systematic Bacteriology 30:485-495.

Wood, D. W., J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. Almeida Jr, L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P.D. Karp, D. Bovee Sr, P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson, and E. W. Nester. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294:2317-2323.

Xu X. Q., and S. Q. Pan. 2000. An *Agrobacterium* catalase is a virulence factor involved in tumorigenesis. Molecular Microbiology 35:407-414.