METHODS 20, 35–46 (2000)
Article ID meth.1999.0903, available online at http://www.idealibrary.com on IDEAL®

Starting a New Genetic System: Lessons from *Bacteroides*

Abigail A. Salyers,[1] George Bonheyo, and Nadja B. Shoemaker

Department of Microbiology, 601 South Goodwin Avenue, University of Illinois, Urbana, Illinois 61801

Despite the fact that genetic manipulation is an essential adjunct of studies of microbial physiology, many important bacterial species are still not accessible to genetic manipulation. In many cases, the people who know the organism best and have studied its physiology extensively have little experience with genetic techniques and are uncertain how to set up a new system. In this article, we describe what we have learned in the process of developing a genetic system for *Bacteroides* spp. and helping to develop genetic tools for two related species of bacteria: *Porphromonas gingivalis* and *Prevotella ruminicola.* We discuss the relative usefulness of different types of genetic tools and make some suggestions about how to construct them. © 2000 Academic Press

Genetic analysis is essential for understanding bacterial physiology and ecology. Yet, despite considerable progress during the past decades, many important bacterial species are still not genetically manipulable. In many cases this situation exists because the scientists who know the organisms the best have little or no background in genetic analysis and are uncertain how to develop a genetic system. Unlike the "model organisms" of the past, such as *Escherichia coli* and *Saccharomyces cerevisiae,* there is no well-funded army of scientists dedicated to constructing a genetic system for most of the organisms for which such systems are needed. Instead, the development of genetic tools is often a lonely struggle involving one or a few laboratories with limited resources and time to spend on tool building. The purpose of this article is to describe some of the things we learned about developing a genetic system from our experience with *Bacteroides, Porphyromonas,* and *Prevotella* species, in the hope of providing encouragement and helpful suggestions to those embarking on the quest for a genetic system.

Let us first address the question of why genetic tools are so important for studies of bacterial physiology and ecology. When working out a metabolic or biosynthetic pathway for the first time, scientists usually rely initially on biochemical analyses to identify the steps in the pathway. This type of analysis can suggest hypotheses about how the pathway works, but to determine whether the pathway actually works as predicted it is necessary to disrupt genes encoding enzymes or other proteins in the pathway to demonstrate their function in the intact organism. When this is done, there can be some surprises and unexpected insights. For example, results of biochemical analyses of the starch utilization pathway of *Bacteroides thetaiotaomicron* suggested that there was a single starch-degrading enzyme that carried out the first step in the starch utilization pathway. To test the importance of this enzyme, we cloned the gene encoding it and disrupted it to determine the effect of losing the gene on the ability of *B. thetaiotaomicron* to use starch. The mutant grew more slowly on starch than wild type but it still grew (1). Clearly there was at least one other enzyme we were not detecting in our assay system. By eliminating the high-activity enzyme, we were able to detect another starch-degrading enzyme, whose activity had been obscured by the first enzyme (2). This second enzyme turned out to be essential for growth on starch. Thus, although its activity in our assay system was the much lower of the two enzymes, this enzyme proved to be the most important enzyme in the pathway. Genetics also helps you discover steps in a pathway that were not initially suspected to be involved or for which there was no good assay. For example, by using a transposon to mutagenize *B. thetaiotaomicron* and screening for mutants no longer able to grow on starch, we discovered

[1] To whom correspondence should be addressed. Fax: (217) 244–8485. E-mail: abigails@uiuc.edu.

1046-2023/00 $35.00

several outer-membrane proteins involved in the binding of starch to the bacterial surface (3, 4). It would have been very difficult to identify such proteins and establish their properties by using biochemical methods. Now that the outer-membrane proteins have been identified, it is time to return to biochemical studies to determine whether the proteins are part of the same complex and which ones are exposed on the cell surface. As our knowledge of the system increases, more sophisticated genetic studies will become possible. Ideally, studies of metabolism should be a constant interplay between biochemical and genetic studies.

ESTABLISHING GOALS AND CHOOSING APPROACHES

What Do You Want to Learn?

There are many types of genetic tools that can be constructed, but some of them will be more appropriate for a particular project than others. Thus, it is important to start by spending some time thinking about what questions are being asked and what genetic strategy is most likely to answer them. Someone who has identified the enzymes in a particular pathway and wants to learn more about the importance of these enzymes will probably find vectors for creating targeted gene disruptions or for delivering transposons most useful, because these tools will make it possible to disrupt the genes one by one and assess their function. By contrast, someone who wants to introduce a new gene into an organism to expand its performance repertoire will want a plasmid carrying a promoter whose strength can be varied. As an aid to making decisions about where to start, let us review the different types of genetic tools and what they can do.

Chemical or UV-Induced Mutagenesis

Since chemical and UV-induced mutagenesis were the earliest techniques for generating point mutants that would identify genes, and since these techniques are in all the genetics textbooks, scientists without much genetic experience often think first about using this approach. If your genetic system is in its early stages, however, your best move is to forget chemical or UV-induced mutagenesis and move straight to transposon mutagenesis. Methods for introducing DNA into cells, along with more details about the vectors and transposons available, are provided in a later section, but first it is important to understand why chemical or UV-induced mutagenesis can be problematic.

Analyzing the types of point mutants obtained with chemical or UV-induced mutagenesis is very helpful once a system has reached a fairly advanced point and interest has focused on interactions between specific proteins, but for the initial stages of genetic analysis it can pose problems that should be faced in advance. It is usually fairly easy to work out the conditions that optimize the number of mutants and to find potentially interesting mutants. The problem comes in the next step, when it is time to clone the wild-type version of the mutated gene. The traditional way to identify the affected gene is to clone it by complementing the mutant phenotype with a library of cloned wild-type DNA. Constructing a complete plasmid library from an organism and introducing this library into the organism of interest is no small undertaking, especially in early stages of plasmid vector development when frequencies of introducing DNA into the strain may be relatively low. If a complementing clone is found, it will have to be subcloned and the various fragments tested for function to narrow down the gene of interest. If this gene is part of an operon and does not have a promoter immediately upstream, narrowing down the gene may prove time consuming. Complementation itself can be tricky. There are cases in which a gene different from the one in which the mutation occurred can appear to complement the mutant gene because the protein produced from the cloned fragment, which is present in much higher abundance than usual, provides a bypass reaction. Moreover, some mutations are dominant rather than recessive, so complementation with a wild-type gene will not restore the original phenotype. Such occurrences are admittedly uncommon, but it is important to realize that cloning by complementation is not as straightforward as it may sound to the uninitiated. There is also the possibility that a mutant phenotype is due to more than one mutation. If so, complementation may be impossible if the two affected genes are too far separated.

Although we believe that chemical or UV-induced mutagenesis is generally a bad idea for scientists taking their first steps in a genetic analysis of an organism, it is important to note that there are special cases in which this strategy can be a good one. If you are interested in a particular gene, whose sequence is already known, analyzing point mutants becomes much easier. For example, for someone interested in mutations in DNA gyrase that confer resistance to fluoroquinolone antibiotics, it is easy to obtain such mutants by selection and to analyze them by polymerase chain reaction (PCR) amplification and sequencing of the mutant gyrase gene.

Transposon Mutagenesis

In general, we recommend transposon mutagenesis over chemical or UV-induced mutagenesis for several reasons. First, a transposon insertion in an open reading frame will generally inactivate the gene. By contrast, point mutations can reduce or increase the activity of a particular protein as well as inactivate it. Thus, transposons usually give good clean knockout

phenotypes. These are the easiest ones to interpret in the early stages of genetic analysis, when the main goal is to find genes encoding steps in a biochemical pathway. A second advantage of transposon mutagenesis is that the transposon carries a selectable marker, so every recipient that acquires the marker gene will be a mutant of some sort. With chemical mutagenesis, mutants are a subset of the total population. This can be a problem when you are screening for loss-of-function mutants. Third, once an interesting mutant is found, the transposon can be used as a marker for cloning the gene it interrupted. Finally, it is easy to check whether the mutant has more than one insertion in its chromosome by using hybridization to determine how many copies of the transposon are present. A problem with transposons is that it is necessary to be able to introduce them into the organism of interest at very high frequency, to get the thousands of transposon insertions needed to hit enough genes to include the ones of interest. Also, transposons can sometimes insert upstream of an open reading frame. Since most transposons have promoters firing out of both ends, the promoter on the transposon end can replace the original promoter of the gene. Conversely, transposons lacking out-firing promoters may cause polar effects by interrupting transcription of a multigene operon. In *Bacteroides,* we have found that even transposons with promoters at both ends can cancel the transcription of downstream genes in an operon (4) because the transposon-associated promoters are too weak to replace the original promoter.

Tn*4351* is a transposon we have used extensively in our work with *Bacteroides* (Fig. 1) (5). Tn*4351* carries an erythromycin resistance gene, *ermF,* which confers resistance to erythromycin on *Bacteroides* species. This gene is expressed in *Bacteroides* and *Porphyromonas* but not in *Prevotella* species. Another *erm* gene, *ermG,* has a broader host range. Tn*4351* also carries a tetracycline resistance gene, *tetX,* that does not confer resistance on *Bacteroides.* It confers resistance on aerobically grown *E. coli* because the resistance gene product, a tetracycline-inactivating enzyme, requires oxygen to oxidize tetracycline (6). Tn*4351* is introduced on the vector shown in Fig. 1B (7). Although this vector was designed with *Bacteroides* and related genera in mind, it can be used in *E. coli* because it has the R6K replication origin. Plasmids with this origin will not replicate unless the *pir* gene is provided in the strain (8). We have not determined the host range for Tn*4351,* but this is a good transposon to try if you are working with gram-negative bacteria, especially those in the same phylogenetic group as *Bacteroides* (e.g., *Porphyromonas, Prevotella, Cytophaga, Flexibacter, Flavobacterium*). Tn*4351* might also work on *E. coli* group gram-negative bacteria, although here there are a plethora of other transposons to try.

Tn*4351* has an unusual feature for a transposon. We found that about half of the time, the Tn*4351* inserted by itself, and half the time the plasmid vector cointegrated with the transposon (5). Usually, a transposon does either one or the other type of integration but not both. Because of this, if we obtain an interesting mutant and want to clone DNA adjacent to the ends of the inserted transposon, we must first determine whether a single copy of the transposon has inserted in the strain or a plasmid–transposon cointegration event has occurred. If the transposon is inserted alone, we cut chromosomal DNA with *Hin*dIII (cf. Fig. 1), shotgun clone the fragments into an *E. coli* plasmid, transform *E. coli,* and select for tetracycline resistance (*tetX*). One can use an enzyme that cuts outside the transposon in an attempt to clone DNA on both sides of the transposon, but in our experience the fragments

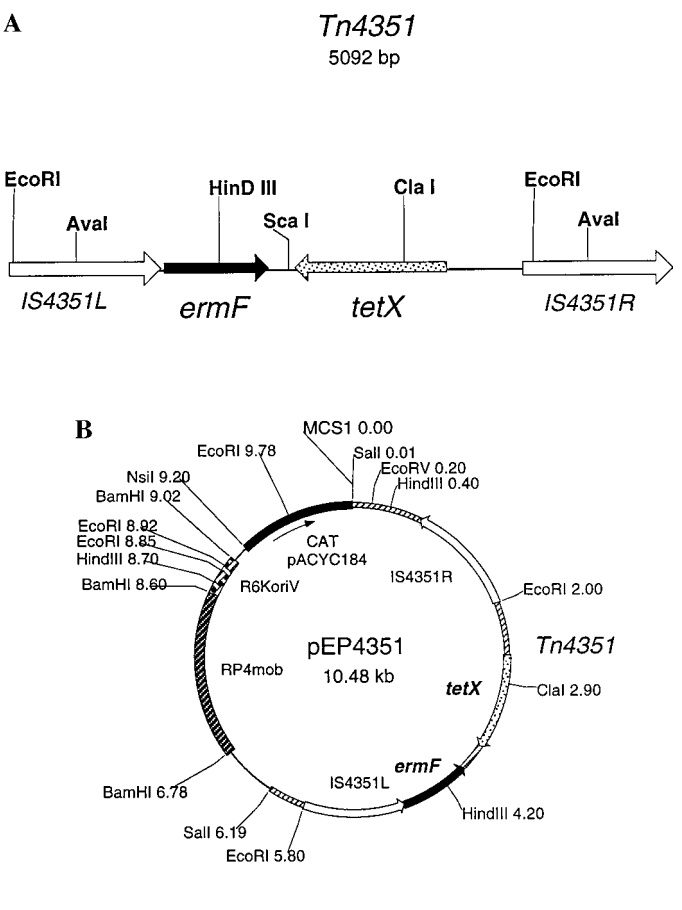


FIG. 1. (A) Diagram of Tn*4351,* a *Bacteroides* transposon that has been used for transposon mutagenesis. Both of the insertion sequences (IS) are active and can mediate transposition. Some useful restriction sites are shown. (B) The latest delivery vector for Tn*4351.* The RK2 (also called RP4) transfer origin region (RP4 mob) is recognized by transfer proteins encoded on RK2 (RP4). The *cat* gene is expressed in *E. coli* but not in *Bacteroides.*

containing the transposon can be difficult to clone because they are large and the two direct repeat IS sequences can render clones unstable unless a *recA* mutant of *E. coli* is used for cloning. If the delivery vector has integrated along with the transposon, we use the strategy illustrated in Fig. 2 to get the right or left junction DNA. In this case, we cut chromosomal DNA with a restriction enzyme that cuts in one of the two multiple cloning sites of the integrated plasmid and somewhere outside the integrated element, then ligate and transform *E. coli*, with selection for chloramphenicol resistance. Once the DNA adjacent to the transposon insertion has been cloned, this DNA can be used as a hybridization probe to clone the intact gene from a wild-type strain.

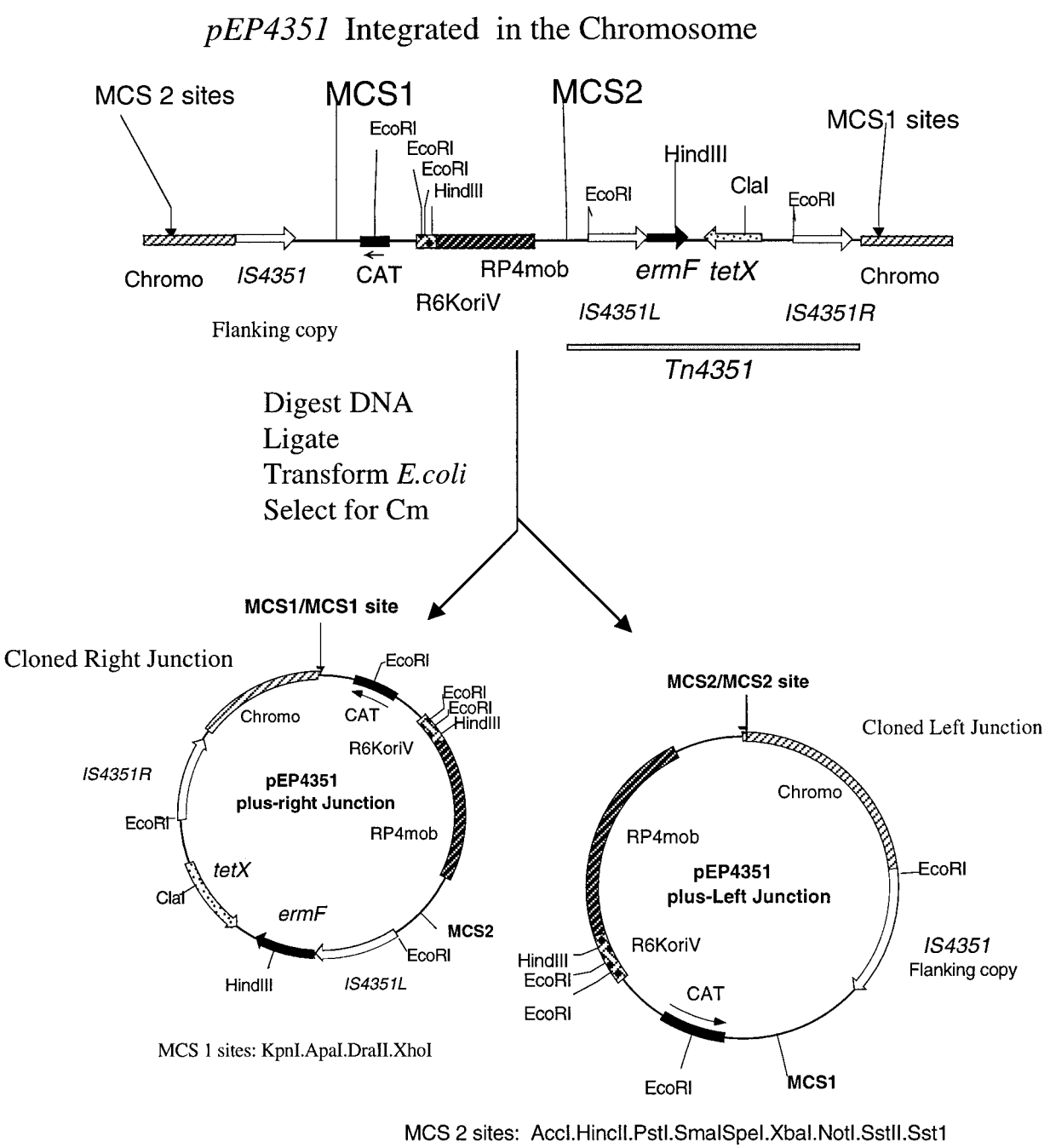
Another broad-host-range integrating element is Tn*916*. This conjugative transposon can transfer to and transpose in a variety of gram-positive and gram-negative bacteria (9). The tetracycline resistance gene it carries (*tetM*) is also expressed in a variety of bacterial genera. A problem with Tn*916* is that in some hosts it integrates site-selectively rather than randomly, but some groups have found this transposon very useful for mutagenesis. Tn*4351* also exhibits some site selectivity. Although we have obtained many different types of mutants using Tn*4351*, about one in six have proved to have the transposon insertion in an intergenic region—a high ratio, given the small size of promoter regions compared with open reading frames. Tn*916* exhibits the same slight preference for intergenic regions (9). An example of how activation of a gene by a transposon can cause initial misconceptions about the nature of the gene affected is provided by one of the starch-utilization mutants we obtained (mutant Ms-5) (1). This mutant grew so poorly on starch that it was at first scored as starch-minus. If had an α-glucosidase activity that was not only reduced compared with wild type but also constitutively expressed rather than regulated by maltose as in the wild type. This phenotype led us to think that the transposon had disrupted a regulatory gene. As we later learned, however, the transposon had replaced the promoter of the α-glucosidase gene, so that the gene expression was now directed by a promoter in the end of the transposon (1). The lower level of expression of the α-glucosidase gene and the downstream genes was responsible for the poor growth phenotype.

Targeted Disruptions

The best way to assess the function of a cloned gene is to disrupt it and determine the effect of the disruption on the phenotype of interest. This is the strategy we used to show that a starch-degrading enzyme from *B. thetaiotaomicron* that we had thought was the dominant enzyme in the starch utilization pathway, based on biochemical analyses, was not the only such enzyme in the cell (1). In another case, we found out that a *Bacteroides ovatus* α-galactosidase gene we had cloned in *E. coli* was not the gene we were after, a gene involved in breakdown of galactomannans. When we disrupted this α-galactosidase gene in *B. ovatus*, there was no change in the specific activity of the α-galactosidase we were studying, the one upregulated during growth on galactomannans, and there was no difference in the ability of the bacteria to grow on galactomannan (10). We never did find a role for this gene in *B. ovatus*, although disruption of it did adversely affect the ability of the bacteria to colonize the murine intestinal tract. This example illustrates why transposon mutagenesis is a more reliable route to finding genes than cloning genes that have a particular activity in a heterologous host and why, if the latter route is taken, it is important to check by disruption mutagenesis for the actual role of what has been cloned.

To make a disruption, an internal fragment of the gene is cloned into a suicide vector, a plasmid that replicates in *E. coli* but does not replicate in *Bacteroides*. An example of such a suicide vector, pGERM, is shown in Fig. 3 (11). pGERM has a selectable marker that works in *E. coli* (*bla*), a selectable marker that works in *Bacteroides* spp. (*ermG*), an *E. coli* replication origin (pUC19), and the transfer origin (*oriT*) of plasmid RK2. A useful donor strain is *E. coli* S17-1, which has the RK2 transfer genes integrated into the chromosome (12). Some of our other suicide plasmids are mobilized by plasmid R751, which does not recognize the RK2 *oriT*. Both RK2 and R751 have an extremely broad host range in the sense that they can mobilize DNA into a variety of different bacteria, even if they cannot replicate in the recipient themselves. That is, they provide the transfer machinery that can be used by other plasmids, such as our shuttle vectors. A surprising feature of all the *Bacteroides* plasmids we have studied to date is that they are mobilized by both R751 and RK2. Thus, *Bacteroides* plasmids can be used as a source of mobilization regions (the *oriT* where the transfer process starts and the mobilization genes that allow the plasmid to take advantage of the transfer apparatus of R751 and RK2).

A little-known feature of R751 is worth mentioning. Transfer rates for this plasmid plummet to near zero under anaerobic conditions. *Bacteroides* spp. are aerotolerant enough to allow the matings to be done under aerobic conditions, but this could not be done with *Porphyromonas* spp. or *Prevotella* spp. We found that pelleting the *E. coli* donor and recipient onto an agar slant in an anaerobic tube, removing the fluid phase, and then injecting a few milliliters of air provided enough oxygen to raise the transfer rate to the detectable level and little enough so that *E. coli* scavenged it before it could harm the anaerobe.



Junctions cloned in plasmids in *E. coli*

FIG. 2. Integrated form of Tn*4351*, in a cell where the transposon cointegrated the plasmid vector instead of transposing cleanly. Cutting chromosomal DNA from the strain, using an enzyme that cuts only once in the multiple cloning site 1 (MCS1), ligating, and transforming *E. coli* will yield a plasmid containing the cloned right junction fragment. Following the same procedure but using an enzyme that cuts in MCS2 will yield the left junction. Note that the cointegration event puts a copy of IS*4351* on either side of the integrated DNA. Trying to clone both junctions simultaneously by using an enzyme that cuts outside the integrated DNA will yield a plasmid that can be unstable due to the presence of the direct repeats. For this reason, we recommend cloning the two junctions separately.

Once an internal fragment of the gene to be disrupted has been cloned in the suicide vector, the resulting plasmid is mated into *Bacteroides* spp. with selection for erythromycin resistance. By far the most common consequence of this is a single crossover insertion into the gene via the cloned region, an event that also integrates the suicide plasmid and duplicates the cloned region. Such insertions can be unstable because they can be deleted by homologous recombination between the repeats at the end of the plasmid. This can be avoided by keeping the antibiotic selection constant. Another type of targeted disruption, which produces more stable mutants, is gene replacement. A gene that is interrupted by a transposon or a DNA segment containing a selectable marker is cloned into a suicide vector and the construct introduced into the organism of interest. Double crossovers flanking the selectable marker replace the wild-type gene with a disrupted form of the same gene. In our experience, double crossovers occur at a 1000-fold lower frequency than single crossovers. Thus, most of the recipients that integrate the plasmid DNA into their chromosomes will have single-crossover insertions and it may be necessary to screen thousands of transconjugants to find one resulting from a double-crossover insertion. In one case, we made a double crossover by first obtaining a single crossover, then screening for a second recombination event that eliminated the plasmid used in the insertion (13). Unfortunately, it was necessary to screen more than 10,000 colonies for loss of a marker on the vector to obtain this mutant.

One way to select for double crossovers preferentially, which works in *E. coli* and many related organisms, is to have a copy of the *sac* gene in the plasmid (14). Gene expression is stimulated by growth on sucrose. The product of this gene is toxic, so that if cells are plated on sucrose-containing medium, there is a strong selection against cells that acquire a single-crossover insertion. So far, this trick has not worked in *Bacteroides* spp. but it has proven very useful to scientists working on *E. coli* and other members of the proteobacteria. An analog that would work similarly in *Bacteroides* and the gram-positive bacteria is badly needed.

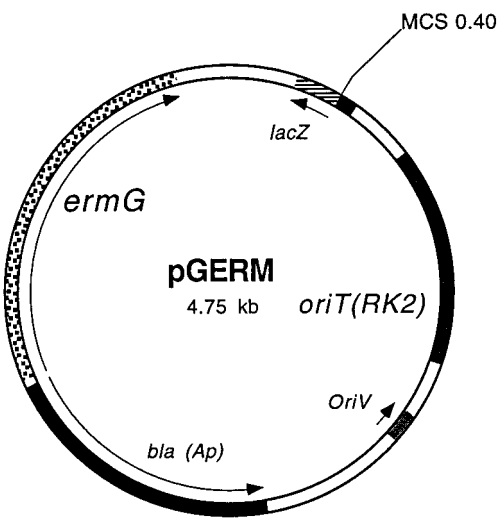


FIG. 3. Map of suicide plasmid pGERM. The sites in MCS 0.40 of pUC19 that are unique are listed below the plasmid.

It is important to keep in mind that transposon insertions or disruptions made with cloned DNA can have a polar effect on the expression of downstream genes. This is a problem because a disruption of a gene that makes the organism unable to grow on a substrate does not necessarily mean the gene is essential for growth if downstream genes are also no longer expressed. Thus, for example, an insertion in a gene encoding an α-glucosidase (*susB*) made *B. thetaiotaomicron* unable to grow on starch not because the gene was essential but because downstream genes encoding the starch-binding receptor were no longer expressed due to polarity (1). There are a variety of ways to check for polarity. If there is an assay for products of downstream genes, this assay can be used to determine whether the mutant is still making these proteins. If not, polarity is probably the reason. If there is no easy assay for the downstream genes, a disruption can be made in the next gene downstream. If an insertion in gene A abolishes growth on maltose and an insertion in gene B, which is immediately downstream of gene A, does too, then the insertion in gene A might have had a polar effect on expression of gene B and one cannot conclude that gene A is essential for maltose utilization. Finally, one can try complementation with a plasmid that contains gene A. A problem that sometimes arises in complementation experiments is that the promoter may be further upstream of gene A than you think and the cloned gene A is not expressed. Also, we have run into problems with titration of regulatory factors by the gene on the plasmid. An example of this phenomenon is described in the next section.

Shuttle Vectors

Shuttle vectors are plasmids that replicate in both *E. coli* (or the organism used for cloning) and the organism of interest. Two examples of *E. coli–Bacteroides* shuttle vectors are shown in Figs. 4 and 5 (15, 16). Note that in addition to the types of genes included in the suicide vectors, these plasmids contain a replication region from a *Bacteroides* plasmid that allows the recombinant plasmid to replicate in *Bacteroides* spp. The shuttle vector in Fig. 4, pFD160 (15), was created by Jeff Smith's laboratory and is the easiest one to use for the initial cloning step because of the number of unique restriction sites and its high copy number. The shuttle vector shown in Fig. 5, pT-COW, was originally de-

signed for use in *Prevotella ruminicola* (hence the "COW" designation), but has proved useful in *Bacteroides* as well (16). pT-COW has a broader host range than pFD160. It also has a lower copy number in *E. coli,* a desirable feature if the DNA you are trying to clone is making the plasmid unstable. Despite their different copy numbers in *E. coli,* both pFD160 and pT-COW have a similarly low copy number in *Bacteroides* spp. (10–20 copies per cell).

Shuttle vectors are used for complementation or to determine the effect of increased gene expression. As an example of the latter, we found that when a gene encoding a regulatory protein, SusR, which activated expression of starch utilization genes, was cloned into a plasmid and introduced into *B. thetaiotaomicron,* the increased level of the regulatory protein actually allowed the cells to grow more rapidly on starch (17). This result suggested that SusR was normally present in limiting amounts. A potential problem with reintroducing cloned DNA into the cell is that if the DNA is used to complement a mutant gene, the gene on the plasmid may recombine with the chromosomal copy, restoring the mutant gene to its wild-type form. For this reason, a recombination-deficient mutant is preferable for use in complementation experiments. A *recA*-deficient mutant of *B. thetaitoaomicron* has been constructed by disrupting the *Bacteroides recA* gene (7). Unfortunately, this mutant is more difficult to work with than wild-type *B. thetaiotaomicron* because it is more susceptible to oxygen. *Bacteroides* species are generally quite oxygen tolerant, although they will not grow when oxygen is present. This aerotolerance is a big advantage because it allows many of the genetic manipulations to be performed on the benchtop rather than in an anaerobic hood. We have noted that the main oxygen toxicity seems to come from superoxide produced by cysteine or sulfide (which is used to decrease the redox potential) or by resazurin (a redox indicator dye). Use of a complex medium such as brain–heart infusion supplemented with hemin overcomes most of the oxygen toxicity problem in the case of the *recA* mutant (7).

A second problem encountered when introducing genes on a multicopy plasmid is titration of regulatory proteins. For example, when we provided a cloned α-glucosidase gene *in trans* with the wild-type form of the operon, in which this was the first gene, the ability of the mutant to grow on starch decreased considerably. The explanation was that 10–20 copies of the promoter region on a plasmid were binding the activator protein that was essential for expression of the chromosomal operon, which contained genes essential for growth on starch (1). This means that we could not complement starch utilization mutants with cloned DNA that had this particular promoter on it without providing additional activator to compensate for the titration effect.

Reporter Genes

Reporter genes are used to monitor promoter activity and regulation of this activity. The best-known reporter gene is β-galactosidase (*lacZ*). This reporter gene has the advantage that the protein it encodes has high activity and is quite stable. It is easy to assay and there is a simple plate assay (X-Gal) that makes it easy to screen mutant colonies for increased or decreased

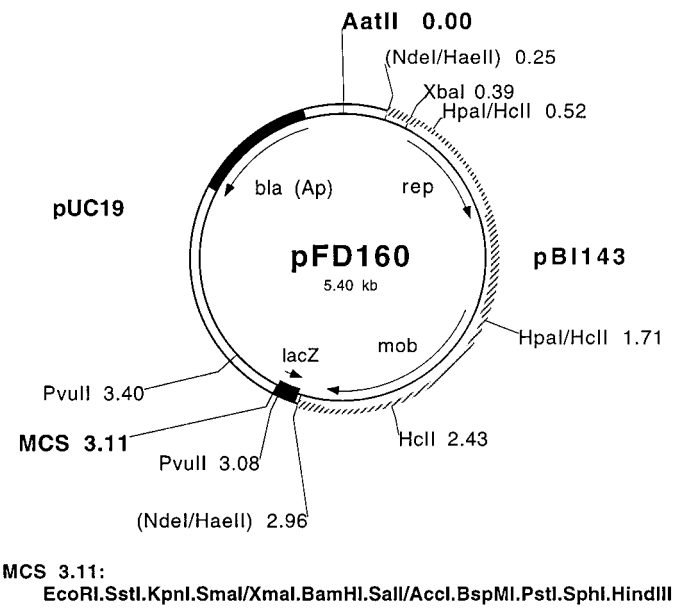


FIG. 4. Map of the shuttle plasmid pFD160 (15). A number of derivatives of this plasmid exist. Some of the marker genes that have been inserted into the *Aat*II site are listed below the plasmid.

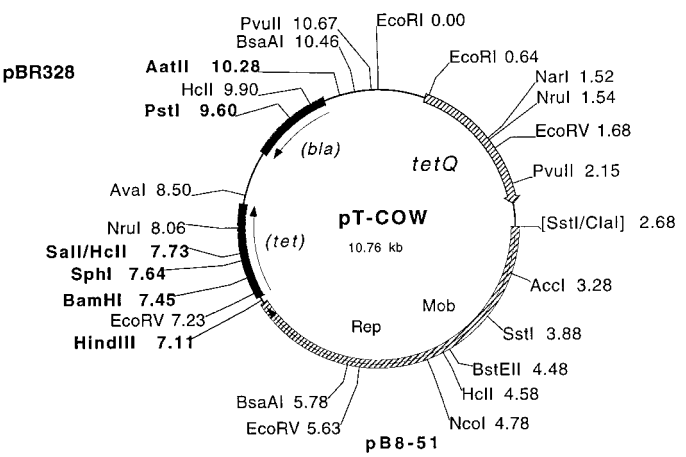


FIG. 5. Map of shuttle plasmid pT-COW (16). This plasmid has fewer usable cloning sites than pFD160 and has a lower copy number in *E. coli.* It has a broader host range than pFD160 and its lower copy number can sometimes be useful for cloning DNA segments that are toxic to *E. coli* when cloned in very high copy number plasmids.

expression of the promoter fused to *lacZ*. Some bacteria like *Bacteroides* spp. have background levels of β-galactosidase activity too high to make this an attractive reporter gene. Also, if the medium contains the reducing agents commonly used to stimulate the growth of anaerobes, X-Gal does not turn blue when it is hydrolyzed because oxidizing conditions are needed for the blue precipitate to form. There are other reporter genes, including β-glucuronidase, green fluorescent protein, and luciferase. Green fluorescent protein and luciferase do not work under anaerobic conditions, especially with reducing agents in the medium, because they require oxygen for fluorescence. Similarly, the X-Gal equivalent of the β-glucuronidase substrate does not produce a color on a reducing medium because oxygen is required for the oxidation step that produces the colored precipitate. We have used β-glucuronidase in studies to determine whether the expression of a particular gene is regulated and, if so, whether it is regulated at the transcriptional level (18). A problem with using β-glucuronidase (*uidA*) as a reporter group is that it is not a very sensitive indicator and is good only for promoters that express at relatively high levels. So far, we have encountered several cases in which we were unable to detect β-glucuronidase activity of a *uidA* fusion despite the fact that we knew the genes were expressed because disruption of them produced a mutant phenotype. No one has attempted to improve translation of the β-glucuronidase message, a step that could yield higher levels of enzyme. Smith's group has shown that the chloramphenicol transacetylase gene works as a reporter gene in *Bacteroides* spp. (19). A gene that encodes a xylosidase/arabinosidase is a very promising reporter group that has high activity (20), although some *Bacteroides* species have this activity already.

Transducing Phages

A belief held by some who practice traditional *E. coli* and *Salmonella* genetics is that a good transducing phage is the be-all and end-all of a genetic system. However, searching for a good transducing phage is a very time-consuming process. Transducing phages are very useful for moving mutations from one strain to another. This is an important feature for those using chemical mutagenesis, because the phage can be used to move the mutation of interest into a clean background, thus eliminating the possibility that the mutant phenotype is due to multiple mutations. It can also be used to check transposon-generated mutants for a second point mutant, although this can also be done by cloning and gene disruption. There is nothing inherently wrong with transducing phages. Rather, the issue is whether it is worth the time it may take to find one. Finding a transducing phage can be very difficult and time consuming. Before committing time to this undertaking, ask yourself whether it is really worthwhile for you given the unquestioned direct benefits that would accrue from developing conjugation- or transformation-based transposon mutagenesis or targeted gene disruption systems.

ATTITUDE ADJUSTMENT

Have Reasonable Expectations

One of the most important prerequisites for success in developing a genetic system is a positive attitude, but one tempered by reasonable expectations. Often researchers who want to develop a genetic system tend to use the frontline systems as their model. Thus, instead of celebrating their first breakthroughs they repine that they are not further along. Development of a genetic system proceeds one painful step at a time. For this reason, it is best to accept the idea that some genetics is better than no genetics and make the best use of what you have. So, if you gain the capability to make directed insertional mutations, even with difficulty and at very low frequency, attack those problems that can be addressed with this limited technology and adjust your strategy for genetic analysis to fit its limitations. Defining your goals realistically and vowing to be relentlessly enthusiastic help to keep you on the right side of the optimism/pessimism divide and virtually ensure your success in developing your new genetic system.

Be Flexible

For reasons that are not always clear, some strains are easier to manipulate genetically than others. We have been able to introduce DNA into many strains of *Bacteroides* spp., but there are some strains that refuse to cooperate. If you have done most of your biochemical or metabolic analyses on a particular strain, it may seem difficult to switch, even to a closely related strain of the same species. To maximize success, however, be willing to consider trying other strains in your initial attempts to develop a genetic system. Moreover, be willing to try more than one approach. Electroporation is a very attractive approach to introducing DNA into a bacterium, but conjugation has worked more often. Especially if you are working with relatively new isolates rather than strains that have been in culture for decades, restriction enzymes may be a big problem. Conjugation, which introduces single-stranded DNA, can get past some restriction systems. Be willing to try a variety of approaches. Once you succeed, it is always possible to go back to the more attractive method for introducing DNA and fix whatever problem prevented it from working in the first place.

Use Kits

One of the most painful sights in modern microbiology is a laboratory run by someone who is starting to do

molecular biology but feels there is some moral stigma associated with using the cloning, plasmid isolation, and polymerase chain reaction (PCR) kits now available. This attitude is particularly common among scientists who were accustomed to traditional biochemical or metabolic analyses and who are still somewhat uncomfortable about molecular biology techniques. The irony of this type of anti-kit mentality is that it sentences students and postdoctoral associates to spend far too much time on what is completely uninteresting (molecular biology techniques) to the detriment of time spent on what is vitally interesting (applying molecular techniques to analyze metabolism). If you are using a microscope or a pH meter or a spectrophotometer, you are already using kits, so why hesitate to take advantage of their molecular equivalents. Kits and DNA sequence facilities are the advantages that have drastically decreased the gap between the well-established genetic systems and the newest ones. Use them, and do not fuss bout the expense because in the long run kits more than pay for themselves by freeing young investigators to do interesting science. It goes without saying, of course, that people who use kits should know how they work and how to make the reagents if necessary.

INTRODUCING FOREIGN DNA INTO YOUR ORGANISM

Learning how to introduce DNA into your organism is often the most difficult step in constructing a genetic system, because in the beginning there are no positive controls and you are dealing with multiple variables. In our experience, it is best to start with conjugation using broad-host-range plasmids, such as RK2 and R751, as the mobilizing agent. The next problem is the construction of the plasmid to be mobilized. This plasmid will need to have a replication origin that works in the target organism, a mobilization region that allows it to be mobilized by R751 or RK2, and a selectable marker. The recommended approach to constructing such a plasmid is to start with a plasmid that occurs naturally in the target organism, then clone selectable markers into the plasmid and try various methods for introducing this construct into strains that do not have the plasmid. If the plasmid is found in most or all strains, you can hope that the selectable marker—if it works—will allow your construct to drive out the indigenous plasmid. When we first constructed such a plasmid for use in *Bacteroides*, the process was unpredictable because we did not know what regions of the plasmid were essential for replication. Thus, it was not clear where to clone the selectable markers so that they did not abolish the replication capability of the plasmid. Today, there is no reason for anyone to take the blind cloning approach we were forced to take. If the plasmid you are planning to use is less than 10 kb in size, have it sequenced. Then you can target specific restriction sites that do not interrupt open reading frames. Since there are so many plasmid replication genes in the databases, you will probably be able to identify potential replication genes by comparing your sequence to sequences in the databases. Second, think phylogenetically. Antibiotic resistance genes, which make very good selectable markers, have now been characterized in a variety of microorganisms. Some genes have been found in a variety of different organisms and apparently have very broad host-range promoters and function. The gene *tetM* is expressed in an amazing variety of genetic backgrounds, as are some of the *erm* genes. Genes encoding β-lactamases are a bit more problematic, possibly because they are secreted proteins, whereas the gene products of *tetM* and the *erm* genes remain in the cytoplasm. The efflux pump tetracycline resistance genes may not work very well because they are cytoplasmic membrane proteins.

A good strategy for picking possible selectable marker genes is to take advantage of the many resistance gene probes and PCR primer sets currently available. Although no one has so far compiled a list of these, most of them can be obtained from our laboratory or from the laboratory of Marilyn Roberts (School of Public Health, University of Washington, Seattle). If you can find resistant strains in the species of interest and if you can determine which gene is responsible for the resistance, you now have a gene that is very likely to work in a susceptible strain of the same species.

When you are first trying to introduce DNA into an organism, check to see whether your strain is naturally transformable. It is true that relatively few of the bacterial pathogens and human intestinal bacteria are naturally transformable but environmental isolates seem to be much more accommodating (21). There is a problem with natural transformation: it is often not plasmid-friendly. In contrast with transformation methods such as chemical transformation and electroporation, which readily introduce plasmid DNA into bacterial cells, natural transformation systems make nicks in the incoming DNA and digest away one strand of the DNA so that what is imported are single-stranded DNA segments (21). Thus, the plasmid must reform inside the cell. So far this has been true for the small number of naturally transformable bacteria studied in depth: *Bacillus subtilis, Streptococcus pneumoniae, Neisseria gonorrhoeae* and *Haemophilus influenzae.* Natural transformation can still be an excellent method for introducing linear DNA segments and doing allele replacements, in which you replace the gene with one interrupted by a marker gene. Finally, keep in mind that some natural transformation systems are regulated. Thus, it may be necessary to try bacteria at

different phases of growth or grown under a variety of conditions.

For strains that are not naturally transformable, chemically induced transformation or electroporation has worked in many cases. Nonetheless, we still recommend starting with conjugation because conjugation has worked in cases where transformation and electroporation have failed. The problem with chemical transformation and electroporation is that so many natural isolates of bacteria are loaded with restriction enzymes. DNA from *E. coli* may not have the appropriate modifications to survive in such recipients. In the case of *B. thetaiotaomicron,* we found that a plasmid isolated from a strain of *Bacteroides* can be reintroduced into that same strain at high frequencies by electroporation. Yet the same plasmid isolated from *E. coli* yielded no electroporants when introduced into *Bacteroides* (22). We tried all the standard approaches to overcoming this barrier without success and have finally given up. Nonetheless, other groups working with other species of *Bacteroides* have used chemically induced transformation (23) or electroporation (24) successfully to introduce DNA cloned in *E. coli* into a *Bacteroides* strain. Initially, conjugation can be more difficult to set up than transformation or electroporation because, in addition to selecting for the recipient, it is necessary to eliminate the donor. If the donor can grow more rapidly than the recipient, has fewer growth requirements, and is more resistant to antibiotics, finding conditions that eliminate the donor without eliminating the recipient may be difficult. The advantage of trying conjugation early in your efforts to introduce foreign DNA into your organism is that if conjugation works, at least you know that your plasmid can replicate in the strain of interest and the marker gene is expressed. Also, this gives you a positive control that can be used to develop and optimize electroporation or other methods for introducing DNA into your organism.

Fortunately, it is easy to recommend a good system for conjugal transfer of DNA from *E. coli* to other bacteria: the IncP plasmids already mentioned, R751 and RK2. These plasmids are not broad host range in the sense that they cannot multiply in bacteria that are far removed genetically from *E. coli* and *Pseudomonas,* but they are dynamite mobilizers of other plasmids to very distantly related bacteria. There are strains of *E. coli* that have RK2 transfer genes integrated into the chromosome (e.g., *E. coli* S-17) (12), which eliminates the fuss and muss of having a second plasmid in the strain. To be mobilized, a plasmid must have a transfer origin and mobilization proteins that make a single-stranded nick at the *oriT* and interact with the mating apparatus provided by the IncP plasmid to start the transfer process. A transfer origin recognized by RK2 is available on a small, easy-to-clone DNA segment. If RK2 is used to mobilize a plasmid carrying the RK2 transfer origin, RK2 itself will provide the mobilization proteins *in trans* so that it is not necessary to clone them into your vector. The RK2 transfer origin is not recognized by R751, so R751 cannot be used to transfer a plasmid with the RK2 *oriT*. In our experience, however, if R751 works so does RK2, so the inability of this origin to be mobilized by R751 should not be a problem. Initially, we relied on *Bacteroides* mobilization regions, which contain both *oriT* and mobilization genes and which are recognized by both RK2 and R751, but we have now switched to the RK2 *oriT* region. In our genetic studies of *Bacteroides* and related genera, we have seen only one case in which the RK2/R751 strategy failed, *Prevotella ruminicola* B_14. In that case, we had to first introduce the plasmid into *Bacteroides* from *E. coli,* then mobilize the plasmid from *Bacteroides* to *P. ruminicola* using a *Bacteroides* conjugative transposon (22).

If you are using R751 as the mobilizing plasmid and your recipient is an obligate anaerobe, you may be in for an unpleasant surprise. Although R751 transfers other plasmids with very high efficiency under aerobic conditions, transfer frequencies plummet under anoxic conditions. When dealing with *Bacteroides* species, which are quite aerotolerant, this is not a problem because we can do the matings under aerobic conditions. But when working with a strict anaerobe like *Porphyomonas gingivalis* or *Prevotella ruminicola,* we had to mix the donor and recipient on anaerobic agar slants, then inject a few cubic centimeters of air, hoping that the *E. coli* cells would consume the oxygen before it could damage the anaerobic recipient. This much air was sufficient to raise the transfer frequency of R751, although it was still well below the frequency attainable under aerobic conditions. Another important thing to realize about the IncP plasmids is that although they will transfer from *E. coli* to *E. coli* in liquid media, their transfer frequencies rise dramatically if the matings are done on agar. Because of this, we recommend doing matings on agar plates and not in liquid media. Still another consideration is the growth phase of donor and recipient cells. In *E. coli*-to-*E. coli* matings with plasmids like RK2, growth phase matters little, but we found that unless the *Bacteroides* cells were in very early exponential phase, no transfer was detectable. In addition, use growth conditions that minimize the production of capsular material that can interfere with formation of the mating bridge that connects donor and recipient.

When first developing conditions for introducing DNA into an organism, it is usually best to start with a shuttle vector, which replicates both in the donor and in the recipient. The number of transconjugants obtained when a plasmid that replicates in the recipient is transferred is higher than the number obtained

when introducing a transposon or a suicide vector containing DNA homologous to a segment of the recipient chromosome. A transposon delivery vector must not only get into the cell, but then the transposon must transpose into the chromosome. Similarly, a suicide vector carrying cloned DNA must first enter the recipient, then integrate by homologous recombination. In both cases, the frequency of the second step is usually 10^{-4} or lower per recipient. For this reason, once you succeed in introducing a plasmid into the organism of interest, it is necessary to optimize the procedure to get the highest frequencies possible so that transposon mutagenesis and directed insertions become feasible.

SURVIVING THE STARTUP PERIOD

Genome Sequencing

Developing a genetic system, especially getting DNA into your strain for the first time, can take a long time and can be very frustrating. While you are waiting, think big and think modern. There is something you can do today that was not feasible until very recently and will help you through the difficult genetic times: partially sequence the genome of your organism. Obtaining a complete genome sequence, with all of the sequences matched up to make a complete genome map with every gene in its place, is still a major undertaking even with recent advances in sequencing technologies. But it is not necessary to go to those lengths to obtain valuable information about your organism. If you generate a random library of clones, then have a local sequencing facility sequence in from either side of the clone, the main work for you is in keeping track of the sequence information and searching the databases. Even if you obtain enough sequences to cover only about 50% of the genome of your organism, you have probably hit virtually every gene your organism has. If you find an intriguing sequence fragment, you can easily use it as a probe to clone out that part of the genome for more complete sequence analysis. The main impediment here is cost, but the cost of sequencing is dropping rapidly. Many universities have made a big investment in DNA sequencing facilities and they want to use them. Try talking to the sequence facility personnel about a special price for your sequencing project, which could be done in such a way as to fill in gaps in their production schedule. Genome prospecting is something you can do that will fire up your laboratory, raise spirits, and give you the strength for the hard part, which is jump-starting your genetic system. Add to that the fact that if your genetic efforts succeed, you are ready with genes to disrupt and hypotheses to test. We have begun a low-level partial sequencing project on the chromosome of *B. thetatioatomicron*. Although only 200 sequences of about 600 bp apiece have been obtained, we have found a surprising number of matches in the databases and some of them were unexpected.

Thinking Phylogenetically

If you are not in the position to finance a partial genome sequencing project, consider a cheaper alternative. Genome sequences are now available for many different types of bacteria. Unfortunately, most of these are genome sequences of bacterial pathogens and they are not well distributed phylogenetically, but if you are working on one of the proteobacteria or one of the gram-positive bacteria, there are plenty of genome sequences in your immediate phylogenetic vicinity. Look through these genomes for genes of interest to you. Do an alignment and design PCR primers that might amplify the corresponding gene from your organism. This works best, of course, in cases where one or more sequences are available in closely related organisms, but it can be done even if this criterion is not fulfilled. For example, we obtained an internal region of a possible *tonB* gene from *B. thetaiotaomicron* by aligning the sequences of *tonB* genes from proteobacteria and gram-positive bacteria and identifying regions where primers might be anchored. The DNA segment that was amplified from *B. thetaiotaomicron* DNA had significant similarity to sequences of other known *tonB* genes (26). We were interested in the *tonB* gene because TonB is an energy source for receptor systems in *E. coli*, so it was possible that TonB could have a similar function in the *Bacteroides* polysaccharide utilization systems, which use receptors to bind the polysaccharide to the cell surface. Unfortunately, a disruption in the *B. thetaiotaomicron tonB* gene had no effect on polysaccharide utilization (26). This type of negative result is the down side of mining the sequencing databases, but not much effort was expended and if the result had been a positive one, it would have been a real breakthrough.

Molecular technology has been changing rapidly in recent years. This is good news for people with new genetic systems because each advance in technology has tended to narrow the gap between advanced systems and primitive ones. Scientists working on advanced genetic systems may actually be at a disadvantage if they do not adopt new technology. In seeking advice, keep in mind that people working on the most advanced genetic systems may not give you the best guidance. They have systems developed in earlier days that have been quite successful and they have no reason to abandon these older approaches. Scientists developing a new genetic system do not have this luxury, and are often better advised to consult another laboratory that has pioneered a genetic system recently. This view is, admittedly, controversial and many scientists working on advanced systems would disagree. Our experience, however, tends to support it. The important

point is to remain flexible and constantly rethink strategies and approaches.

ACKNOWLEDGMENTS

Work described in this article was supported by Grants AI17876 and AI22383 from the National Institutes of Health.

REFERENCES

1. D'Elia, J., and Salyers, A. A. (1996) *J. Bacteriol.* **178,** 7173–7179.
2. Shipman, J. A., Cho, K. H., Siegel, H. A., and Salyers, A. A. (1999) *J. Bacteriol.* **181**(23), in press.
3. Reeves, A., D'Elia, J., Frias, J., and Salyers, A. A. (1996) *J. Bacteriol.* **178,** 823–830.
4. Reeves, A., Wang, G. R., and Salyers, A. A. (1997) *J. Bacteriol.* **179,** 643–649.
5. Shoemaker, N., Getty, C., Gardner, J. F., and Salyers, A. A. (1986) *J. Bacteriol.* **165,** 929–936.
6. Speer, B. S., Bedzyk, L., and Salyers, A. A. (1991) *J. Bacteriol.* **173,** 176–183.
7. Cooper, A. J., Kalinowski, A. P., Shoemaker, N. B., and Salyers, A. A. (1997) *J. Bacteriol.* **179,** 6221–6227.
8. Metcalf, W. W., Jiang, W., Daniels, L. L., Kin, S.-K., and Wanner, B. L. (1996) *Plasmid* **35,** 1–13.
9. Salyers, A. A., Shoemaker, N. B., Li, L. Y., and Stevens, A. M. (1995) *Microbiol. Rev.* **59,** 579–590.
10. Valentine, P. J., Gherardini, F. C., and Salyers, A. A. (1991) *Appl. Environ. Microbiol.* **57,** 1615–1623.
11. Shoemaker, N. B., and Salyers, A. A., unpublished results.
12. Simon, R., Priefer, U., and Puhler, A. (1983) *Bio/technology* **1,** 784–791.
13. Salyers, A. A., and Guthrie, E. P. (1988) *Appl. Environ. Microbiol.* **54,** 1964–1969.
14. Reyrat, J.-M., Pelicic, V., Gicquel, B., and Rappuloi, R. (1998) *Infect. Immun.* **66,** 4011–4017.
15. Smith, C. J. (1985) *J. Bacteriol.* **164,** 294–301.
16. Gardner, R. G., Russell, J. B., Wilson, D. B., Wang, G.-R., and Shoemaker, N. B. (1996) *Appl. Environ. Microbiol.* **62,** 196–202.
17. D'Elia, J., and Salyers, A. A. (1996) *J. Bacteriol.* **178,** 7180–7186.
18. Feldhaus, M. J., Cheng, Q., Hwa, V., and Salyers, A. A. (1991) *J. Bacteriol.* **173,** 4540–4543.
19. Smith, C. J., Rogers, M. B., and McKee, M. L. (1992) *Plasmid* **27,** 141–154.
20. Whitehead, T. (1997) *Curr. Microbiol.* **33,** 1–6.
21. Lorenz, M. G., and Wackernagel, W. (1994) *Microbiol. Rev.* **58,** 563–602.
22. Shoemaker, N. B., Anderson, K. L., Smithson, S. L., Wang, G.-R., and Salyers, A. A. (1991) *Appl. Environ. Microbiol.* **57,** 2114–2120.
23. Smith, C. J. (1985) *J. Bacteriol.* **164,** 466–469.
24. Thompson, A. M., and Flint, H. J. (1989) *FEMS Microbiol. Lett.* **61,** 101–104.
25. Shoemaker, N. B., Getty, C., Guthrie, E. P., and Salyers, A. A. (1986) *J. Bacteriol.* **166,** 959–965.
26. Reeves, A., and Salyers, A. A., unpublished results.