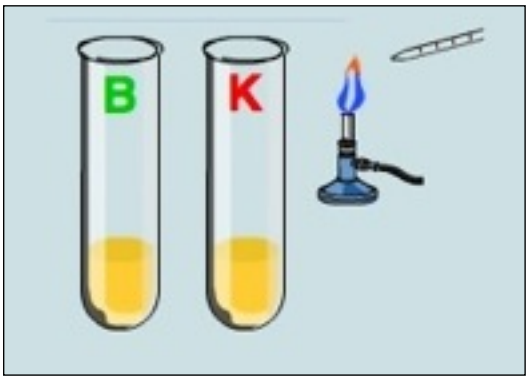
Cricket

Discovering the path of one of biology's most elegant experiments

Bruce Patterson - Wednesday, July 16, 2014



Teaching Goals

The method uses what we call 'Logic Diagrams', and the tool we currently use is LucidChart (free for a small number of projects; they were also eager to work with UA and linked free accounts to our Google-based CatMail network).

Overview

I strongly disfavor introducing students to the paper at the outset of this series of experiments. Doing so allows this to degenerate into yet another example of solving through research and *following* an example rather than discovering ways to solve novel problems. Further, as is inevitable in the modern era, the paper as well as solutions on 'cheat sites' on the web are out there. So I make the following a requirement on *each* submission, worth 1 letter grade:

> "I did not use internet resources other than those assigned, nor did I consult folks who have taken the course before, tutors, or other inappropriate resources"

When the time comes, the paper is
Crick et al., "General Nature of the Genetic Code for Proteins" Nature 192(#4809): 1227-1232.

Note also reference 9; the *foundational assumption* that they were dealing with frameshifts arose from a logical argument based on mutagen structure combined with phenotypic and reversion characterization of mutants. Event this keystone was based on logic, not 'fact'.

I've found a presentation at this link by Jeff Elhai; I think it's an interesting guided tour of the paper and thinking.

1: Which phage is wild type?

This task has the sole purpose of introducing students to their tools. These are a combination of strains (mutant and wild type phage; the latter will virtually always contain one or more single nucleotide frameshift mutations as well as two strains of E. coli, designated **B** and **K**) and lab equipment. The latter is simple stuff—a storage area, a loop tool, a pipet, and a petri plate onto which a mixture of bacteria and phage can be plated.

	E. coli **B**	*E. coli* **K**
Phage w wild-type rIIb gene	small plaques	small plaques
Phage lacking rIIb gene function	large plaques	no plaques

#2: Identify compensating mutations

Students often come at this thinking "I plated out mutant phage therefore I have mutant phage." The oversight here is that the program is simulating "the real thing." Their phage stock consists of tens or hundreds of millions of individual phage, the result of several generations of replication. Each replication event carries with it the usual hazards of mutation. In these simulations, only +1 and -1 nucleotide deletions occur, and only in the rIIb gene. But they occur. So every stock will 'throw off' a variety of mutations, and these will be revealed if the plating conditions are right.

Often all you need to do is walk the students through the underlying biology—get them to realize that little phages are not brought by storks, and that DNA replication is as error-prone for bacteria as it is for us (actually, much more so).

The only exception to the "only +1 or -1" rule is the Deletions scenario, where the students are provided with 'chunk' deletions as starting material.

#3: Do both phage contain identical mutations?

There are a number of scenarios in these exercises where a change of perspective is called for. This is a major tool of thinking, and for this reason, some of the scenario requests are NOT phrased in the way that most directly leads to the answer. Here, for example, a more powerful re-statement of "Do two phage contain different mutations?" is "Do the two phage contain complementary* regions of wild type that could be joined to reconstitute a wild-type sequence?"

Here, a better approach is actually "Do the two phage contain different mutations". Recombination is key. *Throughout* Cricket, I often find myself forcing students to draw the recombinations that they are discussing. It appears to be a natural tendency to focus on ONE of the two products of a reciprocal recombination event. Even when forced to think of the correct 'cross' (mutant 1 x mutant 2), some only 'see' the double mutant, which offers no help. If you force them to draw the event, you can help them see that true wild-type also results.

Another way of approaching this is the 'name all the ways you can describe the two phage as different'. While it is true that in one model "they have mutations in different spots", another thinking tool (different perspective; phrase differences in novel ways) is "each has wild type sequence where the other has mutant". In other words, *combining what is GOOD about each* can result in a product with no errors.

Caveat: In #2 they learned that errors in DNA replication and our capability of working with huge numbers means that a mutant phage plated out WILL yield *apparent* wild type—due to spontaneous new mutations that restore reading frame (only a subset of which do so by TRUELY reversing the initial mutation)

We need to *lead* them to this idea rather than telling them. Sometimes "What did you learn in #2" is sufficient; sometimes having them *describe* the events *and underlying biology* is needful.

*Clarification: be careful of the idea of *genetic* complementation. In the case of Cricket, the only gene that is ever mutant is the *rIIb* gene, and phage are always plated at sufficiently low density that the petri dish exhibits the effects of SINGLE infection. In other words, we are observing the effectiveness of single genes in all instances.

#4: Recover each single mutation from a double mutant

The main idea here is that recombination creates every possible jumbling of input DNA sequences. However, all the mutant phage will 'look' alike (i.e. if starting with +1 and -1 frameshifts, both will yield large plaques on B). But the stated task requires specifically identifying one of each type. What to do?s

First, I lead them through the discussion that we *cannot*, with the tools provided, tell which one is the +1 and which the -1. But we can tell whether any two are the same. Most will get this if you just stare at them. Because the solution is what they just did in task #3 (with the same caveat and solution!). If two phage are the same, mixing them (thereby providing an opportunity for recombination) will *not* result in an enhancement in the sum of small plaques achieved by plating each individually (the rate of spontaneous reversion mutations for each). On the other hand, if the two have mutations at different locations, then recombination events occurring *between* them will reconstitute one true wild-type phage and one double mutant. The former will *definitely* have wild-type phenotype; the latter *may*, depending on whether the two frameshifts have the same 'sign' (both insertion or both deletion cannot restore frame) and whether there is a 'discovered' stop codon in the intervening sequence.

This constitutes a big teaching moment. That something learned at one stage of an investigation can become a TOOL for future stages. And that's a very apt description of Cricket itself—each stage requires skills and insights from preceding ones.

#5 Which of 3 mutations is between the other two?

The underlying goal of this module is to enhance students problem-solving skills. Specifically, to confront novel problems and to learn to recognize tools they have and find approaches and insights that lead to solutions. One of the keys that I push is trying to *discover* a useful perspective on the problem. Oftentimes, this requires discarding an initial, constraining viewpoint. In order to try to 'force' this lesson, the problem title here is somewhat misdirecting. While the goal is to determine which of 3 mutation is 'in between', the *solution* is actually to determine the opposite: which *two* mutations are on thou 'outsides'. Thus, the insight students are seeking is that the one in the middle is the one NOT on the two ends.

More concretely, physical distance between mutations correlates to frequency of recombination events occurring between them. In this case, all mutations are of the same 'sign' (+1), so the primary determinant of frequency of wild-type appearing plaques is the number of recombination events. If all pairwise crosses are attempted under conditions that yield a significant number of wild-type plaques (thus, plating on strain **K**), we should observe one value that is approximately the sum of the other two, such as:

1 x 2: 26 plaques
2 x 3: 37 plaques
1 x 3: 13 plaques

From this data, we can see that phage 2 and 3 contain the mutations that are *most distant* from each other, and by inference, phage #1 must harbor the mutation that is between the other two.

Several key challenges to this scenario. Many students have already forgotten the lesson of Scen #2 (stocks contain compensating mutations) despite reinforcing it in Scen 3. Once it is raised, they tend to be flummoxed. I usually end up trying several routes, with varying success likely specific to the pair that I am talking with. Basically, we are trying to *lead* the student to propose a 'no recombination control'. I prefer to start more broadly, though. I begin by asking whether they can stop spontaneous mutations from happening. They agree they cannot. Then I ask if they can stop recombination from happening. This is an indirect question; of course it cannot be stopped either… BUT they can stop recombination from occurring btw phage #1 and #2 simply by omitting the partner from the experiment… which is exactly the correct control.

Having a prepared scenario to jog their thinking about how to measure background when you cannot prevent it is likely useful. For example, I recently succeeded by asking students how they could tell if I was a mass murderer; they would compare deaths in my neighborhood before/after my arrival.

#6: Make mine a double

This is a classic in the "How many ways can I find to think about this problem" pathway to an insight. Recombination and mutation have made several appearances at this point, but always in the context of creating/restoring wild-type function. Now the challenge will be to find the even-more-defective version. And there's the key.

The most common prompting question I end up asking is "What is special about the phage you are seeking." This is often followed by some dead ends where the student tries to imagine how it is 'more mutant' phenotypically, which it's not. Some in frustration will indicate that it has both mutations in it; I try to grab that moment and get them to amplify on the CONSEQUENCES of this. Often, a graphic compare-and-contrast can be very useful. Getting them to focus on the 'distracting' pieces can also help: if they look long and hard at the 3 kinds of mutant plaques they'll get by plating on B (each of the single mutations and the double), they may well realize that this looks a lot like the earlier case of trying to determine whether 2 phage have the same or different mutations. In some cases, getting them to *explicitly* recognize that these are the starting mutations can be helpful. All this because they *already know* how to tell if 2 phage have the same mutation—there will be a dearth of new wild types when the two are mixed.

The key insight here is that the double mutant has 'the same mutation' as EACH of the starting phage, whereas the 'contaminating' non-recombinants have the same mutation as only ONE of the parents. So they are seeking the mutant phage which recombine with NEITHER parent 1 nor parent 2 to yield significant numbers of wild type.

It is important in this scenario to call attention to the 'multiplexor', a made-up tool that conveniently allows students to take phage straight off the petri dish and use them in 2 x ## crosses against any two stored strains (most conspicuously, the 2 starting mutations) in either strain of bacteria (looking at you, *E. coli* K :-)

#7 Mapping deletions

This one is a bit of an offshoot that reflects some critical science of the period as well as an aspect that played into the actual work, but is not on the direct path followed by the rest of Cricket. For that reason, I include or leave it out based on 'events on the ground', primarily how long it has taken to progress through other aspects of the course. It is important to note that this scenario involves a different concept, in that the sequences being studied here should be considered essential; you can therefore introduce them either as some gene OTHER than *rIIb*, or as occurring in the 'critical' region of *rIIb* that must be present and translated in-frame to produce function.

The key ideas is straightforward, and actually harkens back to Scenario #3 (same or different)—while focus is generally drawn to the deleted material, the Big Idea has to do with what *remains* in each phage. If the two deletions overlap, then the *union* of their genetic information lacks that critical region and there is no hope of reconstituting function.

Some other rules should occur to the thoughtful

—larger deletions will fail to give WT offspring with MORE partners than smaller ones will

—the larger the intervening sequence of intact nucleotides separating two deletions, the greater the frequency of recovered wild type will be

NOTE TO GRADERS: there will be *no way* for students to distinguish between two symmetric cases, i.e. a mutation correctly placed an one end could either be toward the 5' or the 3' end of the gene. For this reason, when displaying actual results (select the student, go to the 'Modeling' area

#8 Demonstrate codons are 3 bases

The underlying rationale is part of the trick here. I have periodically taken advantage of the fact that several of our teaching rooms have tiled floors to demonstrate. Using one color of take, mark off every third tile. Then use a second piece to indicate +1, then +2, then +3 and note that we are back in frame.

Alternatively, the more verbal presentation of "If words are 3 letters long, how many words have we added when 1 letter has been added?" (1/3). If two letters? (2/3). If three letters? (3/3 = 1). Thus, *however many additions it takes* to add a whole word is the measure of a word. If 2 additions restore function, words are 2 'wide'; if 4 additions, that is your word size.

There are several components necessary to putting this together. First is gathering a pool of frameshift mutations all of the same 'sign' (+1 or -1). This can be readily done because all the mutations that restore function to any given *starting* frameshift mutation must all be of the *opposite* sign of the initial, and thus the *same* sign as each other. Thus, Scenario #2.

From here, we must extract the novel mutations (through recombination with wild type; Scenario #4), and from there, we must determine which of the recovered single mutations are the same as the generative initial mutation (boring; discard) and which are the new, opposite-signed mutation (overall, essentially scenario #3).

Now we must engage in construction of the triple mutant in 2 steps. Step one is the combination of two like-signed (+1 and +1 *or* -1 and -1) mutations to create a double. Yes, Scenario #6. Note, however, that in order to make the final construction step work, the double mutant should NOT be the one involving the outermost 2 mutations, as creating a triple will then require a *double* crossover event. So we need to know… which of two is in the middle (Scenario #5).

Most students will proceed in the 'obvious' way now, combining a double mutant with the remaining mutation to yield a triple, which hopefully will show function. Despite their training, many will forget that wild type will also be generated! A very useful intermediate discussion here is to ask "How many of the wild type *phenotyped* individuals will be 'true wildtype' and how many will be triple mutant? A little drawing and thinking should allow them to conclude (or rather, visualize) that the exact same recombination event generates both, so the numbers should be essentially equal. How to test for the triple vs. true wild type? When combined with true wild type (from stock collection), the triple should recombine to yield several different flavors of single and double mutant recombinants, whereas wild x wild should yield new mutants only at the (very rare) frequency of new mutation.

It must be pointed out that there is a vastly superior approach, and the one that was actually used by Crick et al. Consider 3 (+1) mutations, designated A, B and C, with B in the middle. If one creates TWO double mutants, A-B and B-C, then one is in a really neat place. Putting A-B and B-C into a condition where they are able to recombine generates NO true wild

type; any phenotypically wild type strains must be A-B-C (or, of course and always, newly arising mutations A-B-D* or B-C-E*). The numbers game argues for A-B-C to be most common; alternatively, recombining the candidates with wild type should yield multiple instances of only-A and only C, as well as A-B and B-C, each of which can be tested to see if they are 'same as' our existing stocks of A, C (scenario #3) or contain two (scenario #6). This can get tedious, but the bottom line is we have the tools to detect whether the phage of wild-type phenotype indeed contains each of our desired 'ingredient' mutations.

Delivery

My personal view is that for this scenario, the DOING doesn't exhibit novel skills that the student has not already developed and demonstrated, so I consider the task here to be mapping out a clear, complete plan, referencing the 'where I learned it' aspects (see above for referencing each scenario) and recognizing and working around pitfalls (i.e. with the 2 + 1 strategy, true wild type will arise).

Of course, having the students actually DO this does carry a certain savor with it, and cements techniques as well as providing the harsh overseer of reality vs. the merely aspirational goal above :-). Just note that the program has a capacity of only 500 phage, after which nothing can be done other than providing the user with a fresh account to start over. However, several aspects can be useful in strategizing—picking mutations that are not too close (thus ensuring that recombinants will not be excessively rare) but not too far (increases the risk of intervening stop codons) and pursuing 2-3 combinations for the triple are all wise moves.

Demonstrate the location of a <STOP> codon

One of the amazing things about this work is how much juice the investigators sucked out of their data. The fact that restorative mutations were common suggested that one could read a *relatively* long distance in the wrong frame, which indicated that most 'mis-spelled' codons nonetheless 'meant' something, which indicated that most of the 64 possible codons must translate as amino acids, which meant that there must be many synonyms coding for the 20 amino acids. The fact that one could not ALWAYS do so indicated that not ALL of the 64 possible codons were words—there were indeed 'stops' among us.

The key idea here resembles the old poem about the wind: "Who has seen the wind? Neither you nor I/But when the trees bow down their leaves/the wind is passing by". Without sequencing technology, without the ability to purify and examine proteins, there is no hope of *directly* observing a stop codon or its effects. However, the location of a stop codon can be *inferred* using techniques our students have mastered.

Briefly, a stop codon defines *the point past which revertants can no longer be identified in a given frame*. In other words, starting with a +1 frameshift mutation, any -1 frameshift before or after it will restore reading frame. However, if, for example, there is a 'lurking' STOP codon in the +1 frame then translation will stop at that point, and subsequent restoration of frame will be inconsequential. Graphically put:

_____ *wild type b/c STOP is out of frame*
_____(+1)_____[STOP]_____ *fails bc of frameshift & stop*
_____*(+1)*_____*(-1)*_____ *works bc frame restored PRIOR to STOP*
_____(+1)_____[STOP]___(-1)____ *fails bc of stop acts before frame restored*

Thus, another definition of the location of a stop codon is "the point past which revertants can no longer be recovered in a given frame". Of course, this is somewhat complicated in the *rIIb* case because a *second* limitation is that frame must be restored by the time the 'critical' regions of *rIIb* are reached.

Same or different *kinds [sign]* of mutation

This one is not on the intellectual path of the actual discovery or publication, and so is just an intellectual 'toy', hence the Honors designation. The principle here is simple: there are only two possible types of frameshift: +1 and -1. If you're not one, you're the other.