

RECONSTRUCTING THE EVOLUTION OF CAULIFLOWER AND BROCCOLI

For TAs & Instructors: Teaching Information and a Key for the Lab Handout

The Big Picture

This week we'll talk about **how it is possible for large morphological changes to arise from relatively "small" mutations in transcriptional regulators**. We'll talk about one particular variety of these: **heterochronic mutations**, which change the timing of developmental events (an example: broccoli flowering).

What We'll Do

1. Identify different plant structures on a variety of Brassicas, and try to figure out how they might have been caused by differences in timing of developmental events.

We will have broccoli, cauliflower, red cabbage, kohlrabi, collards, Brussels sprouts, and kale in the lab for students to observe. We will have some wild cabbage plants, dried Brussels sprouts shoots, and possibly an *Arabidopsis* plant with a *CAL* mutation. Each group will receive their own (fresh) Brussels sprout, broccoli floret, and cauliflower floret. The other veggies will be shared in the lab section. TAs can walk around and show students the variation in forms, and help them think about what differences in developmental timing could have caused the variety. In the process, the students will learn some basic plant anatomy.

In addition, students will take a closer look at their broccoli and cauliflower under the dissecting scope, trying to identify flower parts (these are identifiable in broccoli but not cauliflower). Please ask students to rinse & dry the fine forceps after use, to put the point-protectors back on them, and to put them back in whatever case they came from.

Students will fill out Table 1 (p. 2) in the lab manual. An example of how they might fill this out is given below, along with some helpful information you might point out to the students as you go around the room. Keep in mind that the main point is to get across the idea that the plants are similar in form, and may only differ because of how much cell division occurs in a certain area at a certain developmental point.

<u>Internode Length</u> <i>(long? short?)</i>	<u>Stem Width</u> <i>(wide? narrow?)</i>	<u>Axillary Buds?</u> <i>(yes/ no)</i>	<u>Petiole Length</u> <i>(long? short?)</i>	<u>Leaf Shape</u> <i>(describe)</i>
--	---	---	--	--

Brussels sprouts	short	wide, cone-shaped?	yes	short	curved in, round
-------------------------	--------------	---------------------------	------------	--------------	-------------------------

Students can cut these down the center, like we do the cabbage, and then peel leaves off to look for axillary buds. Point out that in the photos at the benches and the dried stalks, the parts of the plant they have look like overgrown axillary buds. You can sometimes find axillary buds on the Brussels sprouts starting to expand into leaves. There must be a relatively large amount of cell division taking place near the main vein of the leaf when compared to cell division at the margins of the leaf—this is true of cabbage as well, and the inverse of the pattern in the more curly-leafed kale. The petiole (leaf stalk) is very short in these and cabbage, but a bit longer on the whole plant.

collards	a little longer?	wide	yes	longer	fairly flat
-----------------	-------------------------	-------------	------------	---------------	--------------------

The lab section will share a single collard greens plant, and leaves can be taken off from the bottom to see axillary buds. The petioles are longer than many of the others in lab.

domestic cabbage	short	wide, cone-shaped	yes	short	curved in, round
-------------------------	--------------	--------------------------	------------	--------------	-------------------------

There will be one cabbage for each lab section, so you'll want to cut it down the center and then take (or pass) it around to show the students. You can take the outer leaves off one by one to see the axillary buds and underlying structures. The same pattern of cell division which takes place in Brussels sprouts leaves also occurs in cabbage leaves. Unlike Brussels sprouts, which has many "heads" on a single plant, a single cabbage plant produces a single head of cabbage.

kale	a little longer?	wide	yes	a bit longer	curly at edges
-------------	-------------------------	-------------	------------	---------------------	-----------------------

Again, the lab section will share a single kale plant, and leaves can be taken off from the bottom to see axillary buds. Kale's pattern of leaf growth is that there is much more cell division at the margins of the leaves compared to near the center of the leaf, which gives the leaves the curly appearance. Kale leaves have a longer petiole (leaf stalk) than cabbage or Brussels sprouts. On the kale plant, it is usually the most clear that the leaves spiral off the stem.

	<u>Internode Length</u> (long? short?)	<u>Stem Width</u> (wide? narrow?)	<u>Axillary Buds?</u> (yes/ no)	<u>Petiole Length</u> (long? short?)	<u>Leaf Shape</u> (describe)
kohlrabi	a little longer?	very wide, round	yes or no	very long	fairly flat

There is one Kohlrabi for each lab section. You probably won't want to remove the leaves. The stem is modified into a fleshy, round organ. The petioles (leaf stalks) are quite long. There may or may not be axillary buds present.

wild cabbage	a little longer?	wide	yes	a bit longer	fairly flat; a bit wavy
---------------------	-------------------------	-------------	------------	---------------------	--------------------------------

There are four wild cabbage plants, growing in pots, for the lab sections to share. These may be flowering, so students can also observe and compare the flower structure to that of broccoli and cauliflower. Do not remove the leaves from the wild cabbage plants. Students should compare this plant to each of the others in the lab.

2. View flower structures of broccoli and cauliflower on the SEM.

Students will take a "field trip" to the school's scanning electron microscope, where the technician will show them that it's very clear under the SEM that broccoli has identifiable flower parts, while most of cauliflower's inflorescence meristems never go on to produce flower parts. Send students down in large groups to the SEM room; usually this happens around an hour into the lab—that gives students time to look at the florets under the dissecting scopes first.

3. Using Biology Student Workbench (BSW), investigate one of the transcription factor genes (*CAL*) which has been implicated in cauliflower's phenotype.

There are several steps in this process.

Students will first look at a cDNA sequence of the normal BoCAL gene (this stands for *Brassica oleraceae* CAL) and compare it to the BobCAL gene (*B. oleraceae* *ssp. botrytis*) found in cauliflower. They will look for the ORFs using BSW, and find that there is a stop codon in the BobCAL cDNA which leads to a truncated protein. They will be asked to form a hypothesis about the cause of cauliflower's phenotype (most likely, they will say that the gene is responsible for the arrested inflorescence development (lack of flower parts) in cauliflower; possibly, they will hypothesize that the truncation causes the proliferation of inflorescence meristems common to both cauliflower and broccoli).

They will then download CAL sequences from several *B. oleraceae* subspecies. They will use these to build phylogenetic trees and try to determine where the cauliflowers fall within the species. They will find that cauliflower and broccoli actually fall out together on the tree.

To investigate this further, they will be asked to check the DNA sequences for the stop codon. This will be a little tricky, because the second set of sequences they download are NOT cDNA's, but full DNA sequences (including introns). They will get practice with a real gene dealing with both intron and exon regions. The lab handout has specific guidance for them to do this. When they look for the stop codons, they will find that both cauliflower and broccoli have stop codons, while the other subspecies do not. They may need to adjust their hypothesis about CAL to indicate that maybe the effect of truncation explains the proliferation of inflorescence meristems rather than the arrested development of flower parts.

Attached is a key for the last several pages of this week's lab in the lab manual.

fewest stop codons (the longest Open Reading Frame, or ORF), listed a second time. Import the first listing with the fewest stop codons (not surprisingly, it has only one): click the checkbox next to the sequence, then click the “Import Sequences” button.

At this point, Student Biology Workbench will take you to the Protein Tools page, since you now are importing a protein sequence.

- d. Repeat this process by searching for the *BobCAL* sequence, taken from *B. oleraceae* ssp. *botrytis* (cauliflower). You will need to **return to the Nucleic Tools page** to do this. Use the same database you used in #2b. Select the sequence which says “mRNA, complete cds” rather than one of the variants (“cds” stands for “coding sequences”). Using the SIXFRAME tool, what is the minimum number of stop codons you find? Import this protein sequence (again, not just the longest ORF) as well.
- e. You can compare these two protein sequences by selecting them both on the Protein Tools page and clicking the “View” button. The program will inform you there are “Undefined Characters,” which just means the asterisks representing the stop codons are still present.
- f. After viewing these sequences, formulate a hypothesis for why cauliflower might have a different morphology than wild type *Brassica oleraceae*. In other words, make an educated guess about what effect this change in protein sequence has on the morphology (or phenotype) of cauliflower when compared to wild cabbage. Re-read pages 9-3 & 9-4 if you need to be more “educated.” Write your hypothesis in the space below; talk about it with your TA or lab instructor to make sure it is specific enough.

My hypothesis:

The cauliflower sequence has an extra stop codon, causing the CAL protein to be non-functional. My hypothesis is that this prevents floral development from occurring. (Sometimes students suggest that the stop codon causes the white color of the cauliflower head, and

sometimes they say it causes the proliferation of inflorescence meristems. Other hypotheses are possible.)

3. To begin to test your hypothesis, you can use the *CAL* sequences from several different *Brassica oleraceae* subspecies to make a phylogenetic tree.

If your hypothesis is correct, where would you expect representatives of the *botrytis* (cauliflower) subspecies to be on a phylogenetic tree? Would they be grouped with any of the other *B. oleracea* subspecies? If so, which one(s)?

My prediction:

All the sequences from cauliflower plants should be isolated together on the phylogenetic tree, away from other members of the species.

Unlike the two sequences you’ve been working with, these sequences you’ll be using are full DNA sequences, and include the intron regions of the gene.

- a. First, return to the Nucleic Tools Page. You can download the DNA sequences by typing in the accession numbers. These accession numbers are available in the scientific paper where the sequences are first reported. Here are the eight numbers we’ll be using:

AF241113	AF241129
AF241114	AF241130
AF241123	AF241140
AF241127	AF241149

To download these efficiently, you can type all the numbers into the search box on the Nucleic Tools (not the Protein Tools) page at once; you must separate them with “OR” (e.g. “AF241113 OR AF241114 OR...”). Again, use the GenBank Plant Sequences database. The search should return eight results; if you have fewer, check the numbers in the search against the numbers listed above and make sure they were all entered correctly.

Import the sequences into your database. Use the first two columns of the Table 4 on page 9-11 and the list on the Nucleic Tools page to fill in Column 3 of the data table. (You should have filled out the second column of Table 4 already, as described on page 9-1.) You’ll find this information handy when you look at your

tree. Note that all eight sequence names end with “cauliflower,” which refers to the name of the gene, not the subspecies, in this case.

The subspecies name can be found following the abbreviation “var.,” which stands for “variety” and is used synonymously with subspecies in plants.

- b. Next, you will align the sequences to get ready to make a phylogenetic tree. The computer module which aligns the sequences simply puts them next to each other so that all the similar regions in the sequences line up, accounting for any missing base pairs in the process. Select the eight new sequences on the Nucleic Tools page and click the “CLUSTALW” button to align the sequences.

When you look at the alignment, you will see each of the eight sequences lined up; they are labeled on the left by their ID # (which starts with “101806”). The sequences are too long to fit on a single line, so they scroll down the page. The blue letters in the sequence represent bases which are identical in all eight sequences. The black letters in the sequence represent bases which differ in at least one of the sequences. A dash represents a base pair that is not present in some of the sequences; this may be because of a deletion or an insertion. Note how similar the sequences are for the different subspecies. Note also that these sequences are much longer than the cDNA sequences you looked at earlier; there are several intron regions present.

- c. Import the alignment by clicking the “Import Alignment” button, then select the alignment on the Alignment Tools page. Click the “DRAWTREE” button to see a phylogenetic tree based on the *CAULIFLOWER* sequence.

This tree is unrooted (we aren’t using an outgroup), and so is based only on the proportional difference between each subspecies. Longer lines represent more differences between organisms. You should still be able to get an idea of how the organisms are related to one another, even though the format is different.

You’ll need to use the Table 4 on page 9-11 to determine which subspecies each number corresponds to. Sketch your tree below Table 4 using the common or subspecies names instead of the numeric identifiers.

*There are seven additional *B. oleraceae* ssp. *italica* sequences available from the same paper which are all identical to the *italica* sequence on your tree; we have not asked you to include them because they would be difficult to decipher on the tree. Realize, though, that the position of the *italica* on your tree is not a fluke.*

4. Based on your tree and your earlier prediction, do you need to revise your hypothesis about the relationship between the *CAULIFLOWER* gene and cauliflower’s phenotype? (Again, you may find pages 9-3 & 9-4 helpful.)

My revised hypothesis:

At first, I thought that the stop codon in cauliflower must cause the arrest in floral development. But since broccoli comes off at the same point in my tree as cauliflower, perhaps the stop codon causes a proliferation of inflorescence meristems (the big bunchy head thing that both broccoli and cauliflower have).

5. To determine if the data support your revised hypothesis, you will go back and check the status of that extra stop codon in each of the sequences you downloaded from the different *B. oleraceae* subspecies. Based on your revised hypothesis, which sequence(s) do you predict will have the stop codon?

My prediction:

The cauliflower and broccoli sequences should have the stop codon, but the other sequences should not.

Remember that the second set of sequences you downloaded by accession number are full gene sequences, not copies made from mRNA. You need some way to find out where the extra stop codon is located in the gene.

To do this, first go back to your translated *BoCAL* and *BobCAL* cDNA sequences on the Protein Tools page. View the sequences by selecting them and clicking the “View” button. You want to know exactly where the stop codon

is located, so start by determining how many amino acids are present before the stop codon. You could count all those letters, but it is much easier to just view the sequence in a different format. From the “Format” box, select “PIR/CODATA” and click the “Change Format” button. What is the number of the last intact amino acid *before* the first stop codon in the *BobCAL* sequence?

Last amino acid #: 150.

Using this information, you can figure out exactly where to look for a stop codon in the nucleic acid sequence. Use what you know about the relationship between amino acid sequences and nucleic acid sequences to calculate which bases you want to look at. (You shouldn’t need a calculator, although you may need a piece of paper.)

The base #s I should look at in the cDNA nucleic acid sequence for the stop codon:

451 452 453.

Check your answer with your instructor or TA before continuing.

Go back to the *BoCAL* and *BobCAL* cDNA sequences on the Nucleic Tools page. Select them and click the “View” button to view the nucleic acid sequences this time. Change the format as before, so that the bases are all numbered. Locate the codon indicated above in both sequences, and record your results below. Using the codon tables available in lab, determine the amino acid coded for in *BoCAL* and confirm that the *BobCAL* codon signals a stop.

BoCAL codon: G A G.

codes for: glutamate.

BobCAL codon: T A G.

codes for: stop codon.

Remember that the goal here is to find out what is happening at this location in the other eight sequences you downloaded. You now know which base to look for in the cDNA, but not in the whole gene sequence. To determine this, you need to know something about the structure of the gene, particularly where the introns and exons are.

Luckily, this information is available from the database where we downloaded the sequences. Here is an example of how you would solve this sort of problem, given a gene with three exons and a desire to find the 50th base of the cDNA sequence in the gene sequence which includes introns.

In our hypothetical example, you are looking for the 50th base in the coding sequence. From the GenBank database, you find out that exon 1 of the full gene sequence consists of bases 1-40, exon 2 is bases 100-129, and exon 3 is bases 162-227. We filled this information into the second column of Table 2 below for you. The first row of Table 2 includes a structural diagram of the gene, and gives you a visual way to compare the gene with introns to the cDNA sequence. The numbered boxes represent exons, and the lines between them represent introns (this is just an easy way to think about the structure: it’s really just a string of bases, like any strand of DNA).

Table 2. Example of determining locations of exons in cDNA, given exon information in the DNA sequence including introns.

	base # in DNA including introns (copy from “View Record” exon information)	base # in cDNA (calculate)
exon 1	1 - 40	1 - 40
exon 2	100 - 129	41 - 70
exon 3	162 - 227	71 - 136

How do you determine which base in the full DNA sequence to look at when all you know is the base number in the cDNA? There are several ways to figure this out; here's one suggestion.

- 1) **Determine the exons' corresponding bases in the cDNA.** We've filled in the third column of Table 2 for you. To do this, we used the information in the second column and the knowledge that the cDNA begins with the first base of the first exon. Since we know (from the second column) that the first exon contains bases 1-40 in the full sequence, we can write that in the third column as well. For exon 2, we know (since there are no introns in the cDNA sequence) that it must start with base 41. If we subtract 100 from 129 (the range of bases for exon 2), we get 29. If we add 29 bases to base 41 in the cDNA, we find that the range for exon 2 is 41-70. We can do the same thing for exon 3. It must start with base 71, and (since $227-162=65$, and $71+65 = 136$) go until base 136.
- 2) **Determine which exon contains the base you're interested in.** Remember that we are looking for base 50 in the cDNA. This is in exon 2, which contains the bases 41 through 70. (*You might note that we didn't actually need to fill out the final row of the third column to determine this.*)
- 3) **Find how far into that exon the base is** (the first base? the nineteenth?). 50 (the base we want) minus 41 (the first base of the cDNA exon) is 9, so we know we want the ninth base after the first base of the second exon.
- 4) **Use this information and the number of the first base in the exon to get the answer.** Exon 2 in the DNA sequence including introns begins with base 100; $100+9=109$. So now we know that to find the 50th base in the coding sequence, we should look for the 109th base in the gene sequence which includes introns.

There are eight exons in the *CAULIFLOWER* gene. To find out the specifics of the gene's structure, go to the Nucleic Tools page, select your *B. oleraceae* ssp. *oleracea* sequence whose ID# ends with "93" and click the "ViewRecord" button. Note that more information is available

with this tool than with the "View" tool. Here you can find what organism the gene comes from, a full classification of the organism, and a list of references to scientific journal articles related to the sequence (the length of this list varies depending on which sequence you are looking at). The information you are looking for is found in the "Exon" rows (it is repeated in the "mRNA" row and the "CDS" row, but you may find the "Exon" rows the most clear).

Unfortunately, we hit a snag with the particular sequences we downloaded. The first exon present in these sequences is actually exon #3; exons 1 and 2 are not part of the DNA that was sequenced. By going back to the original paper, it is possible to determine that the length of exons 1 and 2 combined is 264 bases long. Now, looking back at the “View Record” page about the *B. oleraceae* ssp. *oleracea* sequence, we find that exon 3 is listed as “37...107,” meaning that in the gene sequence we downloaded, exon 3 begins at base 37 and ends at base 107. Based on this piece of information and the fact that exon 1 and 2 add up to 264 bases, we can say that this exon (“37...107”) corresponds to bases 265-335 in the cDNA sequence. We have filled this information into Table 3 for you.

You should now be able to follow the procedures above to determine where you need to look for that STOP codon in the gene sequence including introns. Remind yourself which bases you are looking for in the cDNA sequence (these will be the three numbers you listed in the left-hand column of page 9-8:

I am looking for the cDNA base #'s:

451 452 453 .

Now fill out Table 3 (or as much of it as you need to) and determine which bases these correspond to in the sequence you downloaded.

I am looking for the *CAL* gene sequence including introns base #'s:

482 483 484 .

451 - 436 = 15

(bp we want) (first bp in exon 5)

15 + 467 = 482

The location of these bases will be identical for seven of the eight sequences: the kale #46 sequence is off by one base, because its exon 3 range goes from 36-106. Once you know where to look, “View” each nucleic acid sequence and figure out what bases are present. Write the bases for each sequence in Column 4 of Table 4 on page 9-11.

Don't forget to change the format under “View” to PIR/CODATA. You can view more than one sequence at a time by selecting multiple sequences before clicking the “View” button.

Finally, determine what amino acid (if any) is coded for by each codon you found in Column 4 of Table 4. Write this information in Column 5 of Table 4.

Do these results support your hypothesis? Make sure you can explain. Check your answer with your lab instructor or TA before leaving today.

Table 3. *CAL* gene structure data table, based on sequence #93.

	base # in DNA including introns (copy from “View Record” exon information)	base # in cDNA (calculate)
exons 1 & 2	not present	1 - 264
exon 3	37 - 107	265 - 335
exon 4	<u>237</u> - 336	336 - <u>435</u>
exon 5	<u>467</u> - 508	** <u>436</u> - 477 **
exon 6	<u>856</u> - 897	<u>478</u> - 519
exon 7	<u>1514</u> - 1647	<u>520</u> - 653
exon 8	<u>1897</u> - 1975	<u>654</u> - 732

Table 4. Data Table for Cauliflower and Broccoli Evolution Lab.

<i>Column 1</i>	<i>Column 2</i>	<i>Column 3</i>	<i>Column 4</i>	<i>Column 5</i>
Common Name	<i>Brassica oleracea</i> Subspecies Name	Two Digits After "101806"	Bases Present	Codon Codes For
wild cabbage	<i>oleracea</i>	93	GAG	glutamate
		76	GAG	glutamate
cauliflower	<i>botrytis</i>	64	TAG	stop codon
		62	TAG	stop codon
		58	TAG	stop codon
kale	<i>viridis</i> (also called <i>acephala</i>)	46	GAG	glutamate
		44	GAG	glutamate
broccoli	<i>italica</i>	65	TAG	stop codon
Brussels sprouts	<i>gemmifera</i>			
cabbage	<i>capitata</i>			
kohlrabi	<i>gongylodes</i>			
Jersey kale	<i>palmifolia</i>			

Phylogenetic Tree based on *CAL* Sequences (use common names):

