# RECONSTRUCTING THE EVOLUTION OF CAULIFLOWER AND BROCCOLI

## Project Goals

We want you to apply your understanding of nucleic acids, proteins, transcription factors, development, evolutionary selection, and the species concept to a specific evolution of development problem: the origin of domesticated cauliflower and broccoli. A second goal is to develop an understanding of basic flowering plant morphology.

## *Brassica oleracea* Subspecies

Crop species provide dramatic examples of how selection pressure can affect the evolution of plant development and lead to novel morphologies (forms). The species *Brassica oleracea* has several morphologies including cauliflower, broccoli, kale and wild cabbage (Figure 1). Wild cabbage is a perennial plant found in southern Britain, western France, northern Spain, and along rocky cliffs overlooking the Mediterranean. Domestication most likely occurred in this region as people selected variants of the wild species. The result, over thousands of years, is *B. oleracea* plants with such distinct morphological forms. Despite the morphological differences, all these plants are considered to be members of the same species by systematists (biologists that classify organisms based on evolutionary relationships). To distinguish between the forms, subspecies (ssp.) names have been assigned. The wild *B. oleracea* is named *B. oleracea* ssp. *oleracea*. Cauliflower is *B. oleracea* ssp. *botrytis*, and broccoli is *B. oleracea* ssp. *italica*. There are other subspecies (see Figure 1) that have intriguing morphologies as well.

Before you start the lab, fill out the second column of Table 4 on page 11 (really, you'll want this later!).

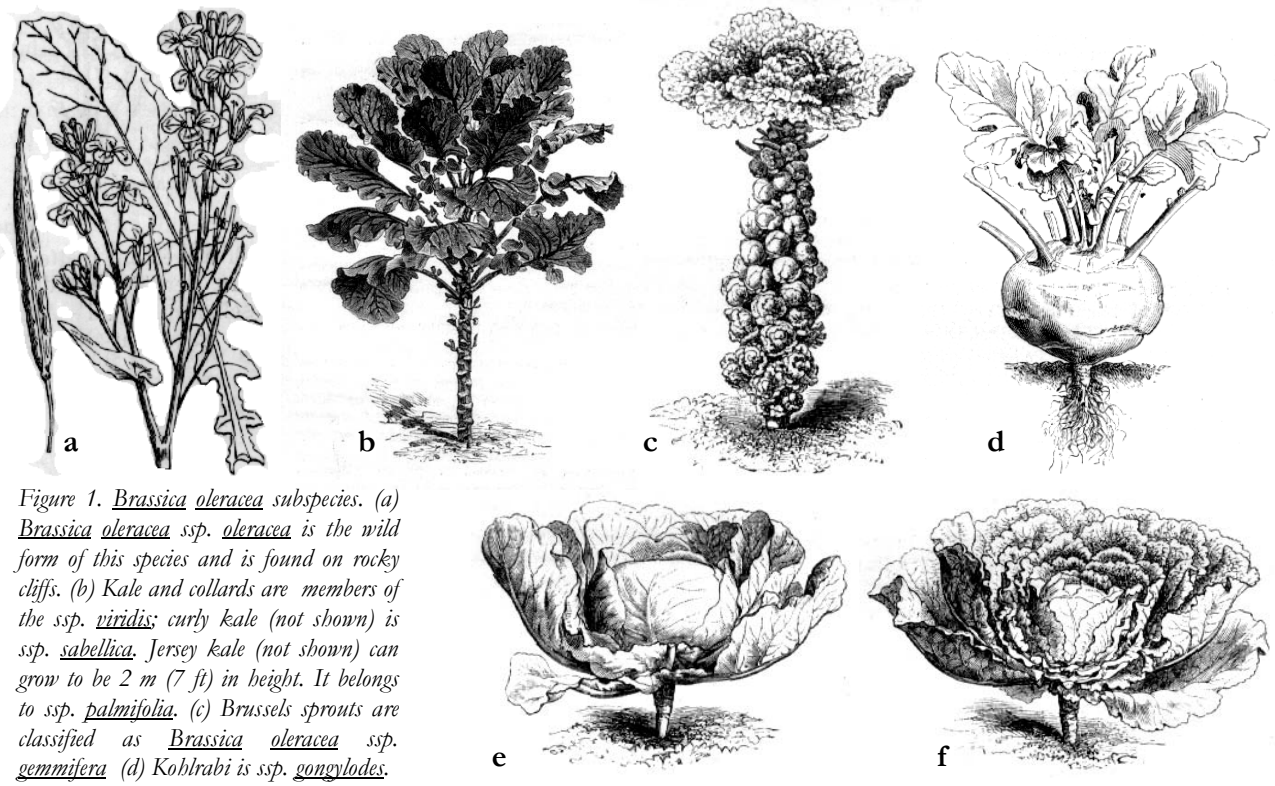



*Figure 1. Brassica oleracea subspecies. (a) Brassica oleracea ssp. oleracea is the wild form of this species and is found on rocky cliffs. (b) Kale and collards are members of the ssp. viridis; curly kale (not shown) is ssp. sabellica. Jersey kale (not shown) can grow to be 2 m (7 ft) in height. It belongs to ssp. palmifolia. (c) Brussels sprouts are classified as Brassica oleracea ssp. gemmifera (d) Kohlrabi is ssp. gongylodes. (e) Brassica oleracea ssp. capitata includes red and green cabbage. The English word cabbage comes from the French word caboche which means head. (f) Savoy cabbage belongs to Brassica oleracea ssp. sabauda.*

*Just how is it that a single species of plants can exhibit so many different forms?*

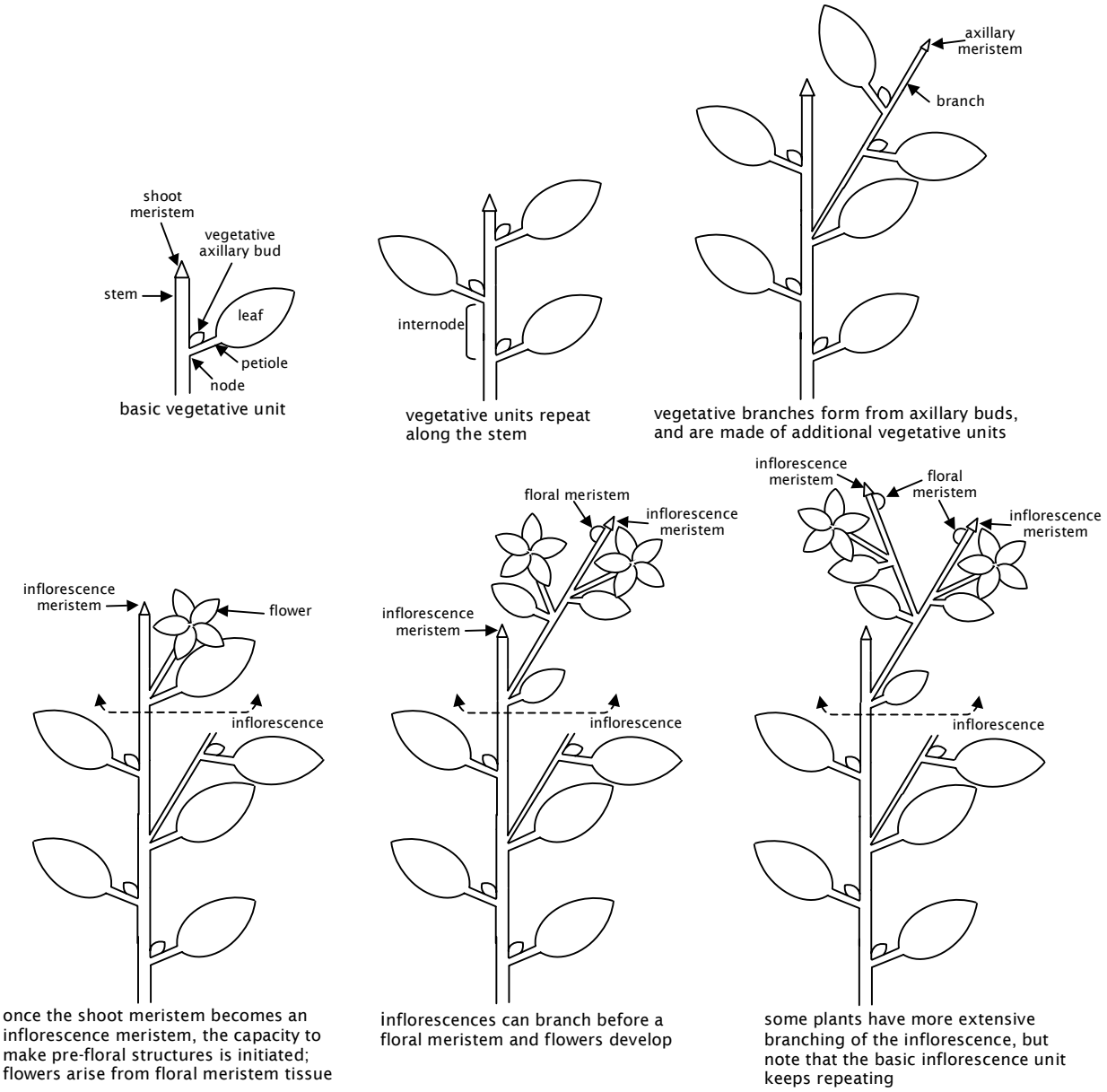



*Figure 2. Schematic of a flowering plant showing the iterative growth pattern.*

## Observations

### *Plant Morphology: Vegetative Structures*

Before we can understand how the differences in the subspecies of *Brassica oleracea* arose, we need to understand how angiosperms (flowering plants) are "built." A **meristem** is a type of plant tissue which is capable of forming many different types of plant structures. Meristem cells are analogous to animal stem cells. When a meristem cell divides, one daughter cell maintains its ability to divide and does not differentiate, or specialize. The other daughter cell differentiates and contributes to the plant body: for example, it may differentiate as a leaf cell. This leaf cell may subsequently divide, but it will only make more leaf cells.

Initially, as a plant grows, the **shoot meristem** produces stem and leaf structures, which are considered **vegetative structures** (as opposed to reproductive structures). The shoot meristem is a cluster of cells located at the growing tip of the plant. Plant body plans are iterative in nature, meaning they are built on repeating units or sets of structures. In addition to producing leaves and the stem, the shoot meristem produces **axillary buds**;

*Table 1. Comparing morphology of vegetative structures in Brassica oleraceae subspecies.*

| | Internode Length *(long? short?)* | Stem Width *(wide? narrow?)* | Axillary Buds? *(yes/ no)* | Petiole Length *(long? short?)* | Leaf Shape *(describe)* |
|---|---|---|---|---|---|
| Brussels sprouts | | | | | |
| collards | | | | | |
| domestic cabbage | | | | | |
| kale | | | | | |
| kohlrabi | | | | | |
| wild cabbage | | | | | |

these buds are found at the junction where leaf meets stem, and actually contain additional meristematic tissue (Figure 2). The meristems in vegetative axillary buds are called **axillary meristems**, and can create additional stem and leaf structures: this is how branching in plants can occur. If there is cell division in an axillary bud, and a branch is formed, there will be axillary buds on the branch as well, which allows for further branching of the plant.

The stem between two adjacent leaves is called an **internode**; plants can differ markedly in their internode length, leaf size, leaf shape, and stem diameter. Differences in the amount of cell division occurring in these structures, or parts of these structures, can help explain differences in form. Using Figure 2, identify the vegetative structures (shoot meristem, stem, leaves, internodes, axillary buds) in the Brussels sprouts, collards, cabbage, kale, kohlrabi, and wild cabbage in lab. (The petiole is the part of the leaf where it narrows and attaches to the stem.) Fill in Table 1 with your observations (no need for measurements—just make rough comparisons). Think about how changes in the timing of developmental events might have led to these differences in form.

## *Plant Morphology: Reproductive Structures*

Developmentally, the major differences between the wild cabbage, broccoli and cauliflower involve when and where the transition to reproductive development begins and arrests.

The transition to the reproductive phase of development results in a modification of the iterative units produced by shoot meristems. At some point in the plant's life (often after a certain number of axillary buds have been formed on the stem), the shoot meristem may be transformed into an **inflorescence meristem**. The inflorescence meristem does not directly produce flower parts, only the branching stem structure which supports the flowers; this branching stem structure is called the **inflorescence** (Figure 2). An inflorescence meristem can transform into a **floral meristem** and form flower structures (sepals, petals, stamens, and carpels) (Figure 3). An inflorescence meristem can also create additional inflorescence meristems.

**Cauliflower** is the result of an evolutionary change in the inflorescence which produces many, many inflorescence meristems without stem elongation (Figure 4). Less than 10% of these inflorescence meristems initiate floral meristems that go on to produce flower parts. This densely packed inflorescence is the cauliflower curd. **Broccoli** has the same compact and extensive inflorescence branching as cauliflower, but the floral meristems initiated by the inflorescence meristems begin initiating floral parts before they are developmentally arrested.

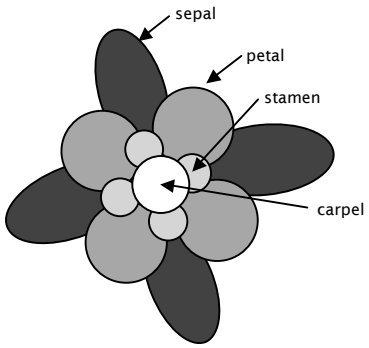



*Figure 3: Organization of flower parts. The outermost whorl is the sepals. The petals are next, followed by the stamens (male reproductive structures) and carpel (female reproductive structure).*

Get a broccoli floret and a cauliflower floret from your instructor or TA. To view these under the dissecting scope, insert the "stem" into the slit of the black rubber stopper at your lab bench. This should hold your broccoli or cauliflower so you can look for flower structures under the scope. Using the fine forceps at your bench, try to dissect apart flowers and identify structures (Figure 3).

*Bioinformatics Investigation of the CAL gene*

The rest of the lab will be devoted to developing a supportable hypothesis for the evolution of broccoli and cauliflower at the level of DNA sequences. This will involve using a site called Student Interface to the Biology Workbench: (http://bighorn.animal.uiuc.edu/cgi-bin/sib.py).

Rather than considering all possible changes in the genome, we will focus on alleles of a particular gene which was first identified in another plant, *Arabidopsis thaliana*. *Arabidopsis* is a member of the family Brassicaceae, as is *B. oleracea*. Due to its small genome and rapid lifecycle, *Arabidopsis* has become a model system for the study of plant development. The entire genome has been sequenced and the roles of many genes in the shoot development, including reproductive development, are well understood. The *CAULIFLOWER* (*CAL*) gene was identified through its mutant phenotype in *Arabidopsis*; the mutant form resembles cauliflower (figure 5). (Remember that genes are often named for the appearance of the organism when the gene is non-functional.) The **ortholog** (same gene in different species) of *CAL* has been cloned from 37

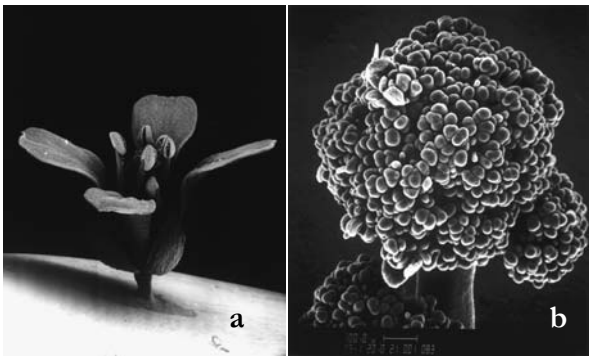



*Figure 5. Scanning electron micrographs of Arabidopsis thaliana flowers. This member of the Brassicaceae family has a very small genome and was the first plant to have its entire genome sequenced. The CAL gene we will be exploring today was first identified in Arabidopsis. (a) A normal, wild type flower in A. thaliana. In combination with a mutation in a closely related gene, AP1, a mutation in CAL results in the cauliflower phenotype seen in (b). Images courtesy National Science Foundation.*

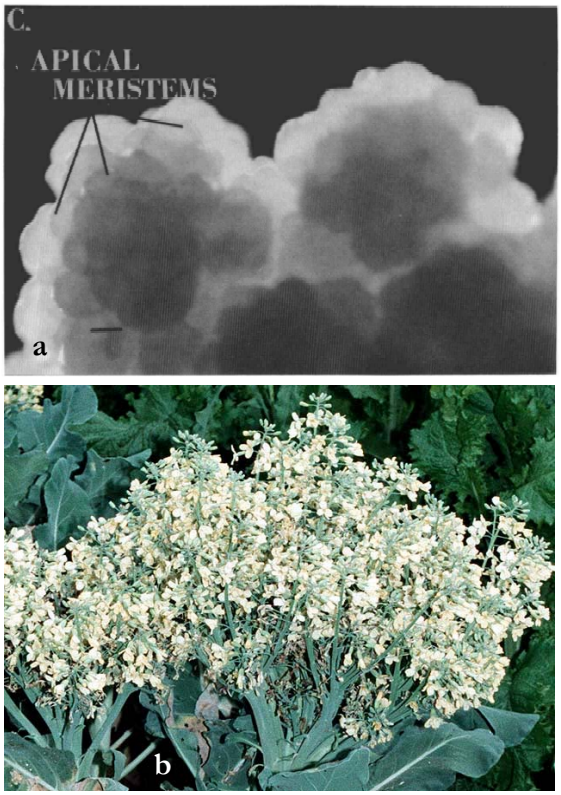



*Figure 4. Cauliflower develops as far as the inflorescence meristem stage and then arrests development (a), while broccoli has initiated floral parts when it arrests. This micrograph (a) shows the many meristems of cauliflower that have ceased growth before inititiating flower parts (source: Medford, J.I., J.S. Elmer, and H. J. Klee 1991 Plant Cell 3:359-370). Unlike cauliflower, broccoli (b) begins to initiate floral parts before development is arrested. Under warm conditions, broccoli flowers will develop.*

different populations representing several of the *Brassica oleracea* subspecies listed in Table 4 on page 11 These genomic DNA sequences have been entered in a public database called GenBank; you will be working with some of these sequences in lab today.

*In Arabidopsis, a second gene called APETALA1 (AP1) is also involved in the phenotype which causes a cauliflower-like appearance. AP1 is a gene that codes for a transcription factor that initiates gene expression necessary for flowering. Among the eudicots (a huge group of flowering plants),* AP1 *is highly conserved. That is,* AP1 *has been found in all of the eudicots that have been screened thus far. This gene is a member of the MADS box transcription gene family (M for mouse, A for Arabidopsis, D for Dictyostelium which is a cellular slime mode, and S for Saccharomyces which is yeast). This gene family has evolved, in part, through gene duplication and divergence. CAL, in fact, is a "new" gene which resulted from the duplication of AP1.*

1. When researchers publish information about a new DNA sequence, they enter those data into GenBank (a freely available database) and include the GenBank accession number in their article so that others have access to that information. We will be working with those sequences in lab today. To obtain your sequences, start at the Student Interface to Biology Workbench (SIB) site.

   a. Go to the SIB - http://bighorn.animal.uiuc.edu/cgi-bin/sib.py

   b. Click the REGISTRATION button and register with a username and password you will remember; this will allow you to save information about your sequences on the SIB web server.

> You may find it helpful to choose one username and password for all web sites where security is not crucial, so you have fewer passwords to remember. This would be an appropriate site to use such a password. Try to make this different from the password you use to access school accounts (which will change regularly) or web sites where you purchase things (this should be more secure).

   *You will only be asked to provide your name, email address, a username, and a password. The next time you go to the SIB, you can just type in your username and password to access your data.*

   c. You should be on the Preferences page. Start a new session by clicking on the NEW button; create a name for the session (like "Brassica") and click the "Start New Session" button. You will be returned to the Preferences page, and your new session will be highlighted at the bottom of the page.

   *There are four different primary pages in SIB; here is a list, including commonly used features:*

   *Preferences: start or resume sessions*
   *Protein Tools: search for, get information on, and align protein sequences*
   *Nucleic Tools: search for, get information on, and align DNA or RNA sequences*
   *Alignment Tools: compare sequences and make trees*

   *Creating new sessions can help you keep your data organized; for today, you can probably keep all your data in one session. To choose a particular session, go to the Preferences page, click the button beside the session you want (at the bottom of the page), and click on the RESUME button. The highlighted session is your current session.*

2. In order to develop a hypothesis for exactly what the *CAL* gene has to do with cauliflower's phenotype, you will first download sequences which are ready to be translated into protein. These DNA sequences have been copied from the mRNA, and are called "cDNA" for "copy DNA." Realize that this means that the intron sequences have been removed.

   a. Click the "Nucleic Tools" button at the top of the Preferences page. Scroll through the Nucleic Tools page to get an idea of the options that are there.

   b. Type "BoCAL" (without the quotes) into the top text field under "Multiple database search for nucleic sequences." In the same box in the table, select "GenBank Plant Sequences" from the list of possible databases. Click the "Ndjinn" button in the right hand column of the table to begin the search.

   *Ndjinn is pronounced "engine" (as in search engine).*

   *"BoCAL" is the name of the version of the CAL gene found in Brassica oleracea ssp. oleracea, or wild cabbage. This is the wild type version of the gene.*

   c. On the results page, click the checkbox next to the wild type sequence of the *CAULIFLOWER* gene found in *Brassica oleraceae* ssp. *oleracea* (wild cabbage). Choose the sequence which says "mRNA, complete cds," rather than one of the variants ("cds" stands for coding sequences). Click the "Import Sequences" button to save this sequence in your workbench. You should be returned to the Nucleic Tools page.

   d. Your sequence is at the bottom of the Nucleic Tools page; scroll down to find it. Select your nucleic acid sequence by clicking the checkbox next to it. Then view the sequence by clicking the "View" button. After you've looked it over, click the "Return" button to go back to the Nucleic Tools page.

e. Next, translate your DNA sequence into a protein sequence by selecting the sequence on the Nucleic Tools page and clicking on the "SIXFRAME" button. This will return all the possible protein sequences based on this DNA sequence (why are there 6?). Each amino acid is represented by a single letter, standard abbreviation (see box below). The asterisks are places where no amino acid is coded for (due to a stop codon in the DNA sequence). The six protein sequences are listed, followed by the sequence with the 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.

| Abbreviation | Amino Acid |
|---|---|
| A | alanine |
| C | cysteine |
| D | aspartic acid |
| E | glutamic acid |
| F | phenylalanine |
| G | glycine |
| H | histidine |
| I | isoleucine |
| K | lysine |
| L | leucine |
| M | methionine |
| N | asparagine |
| P | proline |
| Q | glutamine |
| R | arginine |
| S | serine |
| T | threonine |
| V | valine |
| W | tryptophan |
| Y | tyrosine |

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 3 & 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:

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:

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 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 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 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 3 & 4 helpful.)

My revised hypothesis:

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:

   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 #:_____

   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:

   _____ _____ _____

   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: ____ ____ ____

   codes for:_____

   *BobCAL* codon: ____ ____ ____

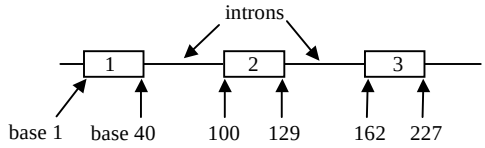
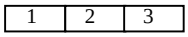
   codes for:_____

   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 50$^{th}$ base of the cDNA sequence in the gene sequence which includes introns.

   In our hypothetical example, you are looking for the 50$^{th}$ 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 8:

I am looking for the cDNA base #'s:

_____ _____ _____

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:

_____ _____ _____

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 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** | | 336 - |
| **exon 5** | | |
| **exon 6** | | |
| **exon 7** | | |
| **exon 8** | | |

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

| *Column 1* | *Column 2* | *Column 3* | *Column 4* | *Column 5* |
|---|---|---|---|---|
| **Common Name** | ***Brassica oleracea* Subspecies Name** | **Two Digits After "101806"** | **Bases Present** | **Codon Codes For** |
| **wild cabbage** | | | | |
| | | | | |
| **cauliflower** | | | | |
| | | | | |
| | | | | |
| **kale** | | | | |
| | (also called *acephala*) | | | |
| **broccoli** | | | | |
| **Brussels sprouts** | | | | |
| **cabbage** | | | | |
| **kohlrabi** | | | | |
| **Jersey kale** | | | | |

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