Lab, Week 5
From Gene to Protein to Behavior

ANALYSIS OF THE BASE SEQUENCE IN THE DNA OF THE *ADE1* AND *ADE2* GENES

In the first laboratory period you examined the effect of UV light on *S. cerevisiae*. Many of you were able to note that this treatment resulted in a number of different types of variants, including colonies that appeared red. That the red colony types arose from the white suggests that the two are highly related (highly similar). Are there any other differences between red and white strains in addition to the difference in color? You tested the growth properties of the red and white yeast under two conditions, with adenine provided in the agar and without adenine provided in the agar. The two types of yeast showed a difference with respect to their growth. Thus the two types of yeast, although presumably highly similar, seem to differ with respect to two characters, colony color and nutritional requirements.

So in those colonies that were red that came from a white progenitor, what changed? What is the underlying difference that leads to the difference in colony color?

In the third week of laboratory you prepared DNA from the red type of yeast. We are testing the hypothesis that the difference between the red and the white yeast is a difference in one of two genes. The names of the two genes are *ADE1* and *ADE2*. Genes are segments of DNA and therefore in order to test this hypothesis we need to obtain and characterize the DNA from the red and the white yeast. We will examine *ADE1* and *ADE2* genes (there are a total of 6000 genes in the yeast) because in previous cases red colonies were associated with changes in one of these two genes. You noted this in the computer work you did last week in lab. By the end of the third laboratory period you obtained DNA from the red type of yeast.

> (There are no set rules for naming of genes but it is typical that the name of a gene reflects the function of the protein encoded by the gene. Recall that in the first week you showed that the red version of the yeast required adenine in order to grow while the white version of yeast did not require adenine to grow. Previous researchers have shown that in versions of yeast where cells are red and adenine is required to grow there is a change in one of two genes. Because changes in each of these genes results in a requirement for adenine the genes were named *ADE1* and *ADE2*.)

In the fourth week of laboratory you performed the polymerase chain reaction (PCR) on the DNA from the red yeast strain. This DNA contains the entire repertoire of 6000 yeast genes. Remember, the hypothesis that we are testing is that the difference between the standard (wild-type) white yeast strain and the red yeast strain is due to differences in the base sequence of either the *ADE1* or *ADE2* genes. The PCR was done in order to restrict our analysis to only two of the 6000 genes and to obtain sufficient quantities of the DNA of the two genes (*ADE1* and *ADE2*) so that the base sequence of each of these two genes could be determined. Also remember that we already know the sequence of the two genes in the standard white type of yeast. This is what you found on the computer last week.
We have taken those DNA molecules generated by PCR and sent them to a DNA sequencing facility. Such facilities have the equipment to determine the base sequence in the DNA. A brief animation of how this sequencing technology works will be shown.

The data have come back from the sequencing facility and we will spend today's laboratory period analyzing and interpreting data through the use of computer software. The data analyzed will be from only one of the red mutant yeast strains. These data will then be compared to data you obtained last

week on the standard (wild-type) white strain. Will there be any differences between the data we have collected from the red strain when compared to the sequence data of the white strain? If so, will differences be detected in both the *ADE1* and *ADE2* genes?

Before accomplishing these goals we will first check to see if the polymerase chain reaction that you did last week was successful. The amplification of the DNA from the *ADE1* and *ADE2* genes of the red yeast is required in order to have enough material to determine the base sequence of these two genes. In order to determine if the PCR was successful, we will use the same technique that we used last week, agarose gel electrophoresis.

Agarose gel electrophoresis of the PCR reactions

- Retrieve your samples from the PCR experiment that you did last week. You will analyze these two samples at the same time.

- Cut a square of parafilm approximately 3 inches x 3 inches. Pipet 3 microliters of DNA loading dye on two well-separated spots on the waxy surface of the parafilm.
  Set a P20 pipetman to 15 µl. Withdraw 15 microliters of the sample from the tube that contains the *ADE1* PCR reaction. Then pipet the contents of the tip into one of the 3 microliter drops of loading dye. Mix the sample with the loading dye. To do this, pipet the sample up into the pipet and then back onto the parafilm. Repeat this several times. By now, the blue color of the loading dye should be well mixed with your sample. Eject the pipet tip and place a new tip on the pipetman.

  The loading dye serves two purposes. First, it gives the sample a density greater than water. This means when you add your sample to an aqueous medium it will sink. Second, the migration of the dye solution will give an indication of how long the sample should remain in the electric field.

- Again using a pipetman (P20) withdraw 15 microliters of the *ADE2* PCR sample and pipet it into the other 3 microliter drop of loading dye. Mix the sample with the loading dye by pipetting up and down several times as described above.

- The agarose gel will be prepared by your instructor. When the gel is ready, **carefully** take your piece of parafilm that has your two samples mixed with dye to the electrophoresis chamber. The gel will be loaded as a group effort. Take notes so that you know which lane has your sample and which samples were loaded in the other lanes.

- After all of the samples are loaded on the gel, a current will be passed between the two electrodes. This will start the process of electrophoresis. After about 75 minutes the samples should be ready to visualize.

Any sample that has detectable amounts of DNA will be made visible when the agarose gel is viewed under UV light. If your PCR reaction(s) did not work then you will not see any bands that fluoresce under UV light. If your PCR reaction did work there should be a single discrete band ("pile" of DNA). You should be able to compare the size of this band to the size of the bands in the molecular weight stands to get an **approximate** size of the DNA made in the PCR reaction. Note that this is an **approximation** of the size, because DNA sometimes runs a bit differently than the size standards when there is a lot of DNA and the bands are thick, as yours are likely to be. For example, you might be off in your estimate by a 100 bases or so.

Estimate the size of your DNA fragments generated by PCR. Remember you can only estimate the size---you should be able to state with confidence that the size is within a certain range. How do these sizes compare to the sizes of the ADE1 *and* ADE2 *genes from white yeast that you calculated in the computer lab last week (this calculation is exact---to the base-pair)?*

Compare the distance traveled by the DNA from each of the two PCR samples. Additionally, compare these distances with the distance traveled by the DNA you recovered from the red yeast in the third week of lab. Based on the gel, are the ADE genes smaller or larger than the total yeast DNA pieces you isolated in the third week of lab? Does this make sense?

DNA Sequence Analysis

The technology of DNA sequencing has revolutionized biology. This technique is described in your textbook; you will also work through some critical thinking questions to better understand the technique. In addition, a short animation describing how DNA is sequenced will be shown. This will give you an indication of how the following sequence data (see Step 1) was obtained. The computer files that you will need are on the CD labeled "principles- ade exercise." The computer programs you will need are on the desktop or in the applications folder on the hard drive or on the CD. These programs are Editview and DNA Strider. Note that each question that is both underlined and in italics will be questions for the lab assignment.

During today's laboratory period you will use the computer to analyze the *ADE1* and *ADE2* genes from the red yeast and compare the data with the *ADE1* and *ADE2* genes of the standard white yeast. One person in each lab group will study the *ADE1* gene while the other person will study the *ADE2* gene. You will then share your results with your lab partner. If you will be studying *ADE1*, start reading below. If you will be studying *ADE2*, skip about 8 pages and start on the page at the top of which reads "Below are the objectives and procedures for studying the *ADE2* gene".

Below are the objectives and procedures for studying the *ADE1* gene.

1) **Examining the primary data of the *ADE1* gene**
2) **Comparing the nucleotide sequence of the *ADE1* gene from the red yeast to the *ADE1* gene from the white yeast**
3) **Determining the amino acid sequence encoded by the *ADE1* gene**
4) **Comparing this amino acid sequence to the sequence from the standard (wild-type) white strain**
5) **Examining the database for a human gene that encodes a similar (although weakly similar) protein that carries out the same function**

1) Examining the primary data of the *ADE1* gene

Open up the file labeled "ADE1f.seq" using the program "Editview". To do this double-click on the "Editview" icon, and under the pulldown heading of "File" select "Open". Then select the file name "ADE1f.seq". This computer file contains the sequence data we obtained from the PCR product generated with the ADE1 primer set in the previous lab session. This is the sequence of bases corresponding to the *ADE1* gene from the red yeast strain. At this point you should see peaks of four different colors. There is also an X-axis that consists of a time unit of measurement, a nucleotide number, and a nucleotide (designated A, C, G, or T). There are a few things to note:

i) at many points there is only one peak and that the color of the peak determines the nucleotide indicated on the X-axis.

ii) at some points, there is more than one peak. This is due to technical problems and makes it difficult or impossible to predict the base that is at that position on the DNA strand. This is why there is occasionally a "N" instead of an A, C, G, or T. "N" designates that the identity of the base at this position was not unambiguously determined.

iii) it is technically more difficult to determine the base sequence near the beginning and thus there are many "N's" in this area. It is also technically difficult to determine the base sequence much more than 500 nucleotides from the start. Therefore we will limit our analysis to the base sequence between position 36 and 579.

In the bottom left corner of the window there are four icons. One of them reads "CATG". Click once on this icon. This changes the information that you were looking at into text only, rather than the multicolored peaks with the various other annotations. Notice all of the N's at the beginning and a few of them near the end of the text string. Highlight the characters between position 36 and 579 (from TATTAA to CTGCCCA) and then select "copy" from the "Edit" pull-down menu. You will now move this sequence of bases into a program that allows a search of the *Saccharomyces* database for any similar sequences.

Deleted: (from TATTAA to CCAAGGC)(from TATTAA to CCAAGGC)

2) Comparing the sequence of the *ADE1* gene from the red yeast to the sequence of the *ADE1* gene from the white yeast. In the following steps you will search SGD, the database that has the DNA sequence of all genes from the standard (wild-type) white yeast that you used last week, to see if any have the same, or similar, sequence of bases as the sequence you have determined for the *ADE1* gene from the red yeast.

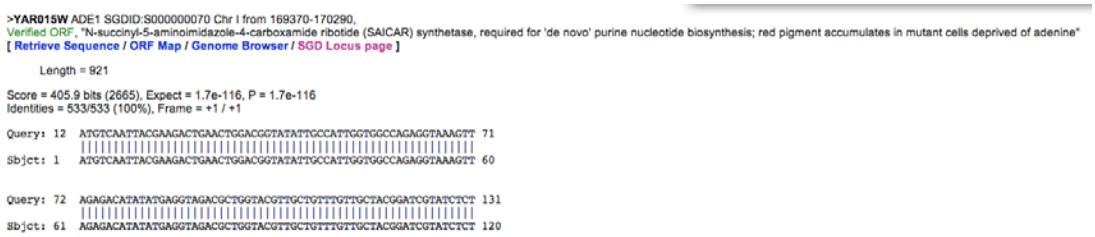
a) Accessing the database via the internet

Open Netscape and type in the following address
http://www.yeastgenome.org
At this website, the entire base sequence of the standard (wild-type) white strain of *Saccharomyces cerevisiae* is stored. This represents about 10 million base pairs. Nested somewhere in this 10 million base pairs is the sequence for the *ADE1* gene (only about 1000 base pairs long). At the top of this web page within the beige bar select "BLAST". This program will search the entire sequence of bases of the *Saccharomyces cerevisiae* in attempt to find some sequence that is either identical or similar to the "query" sequence (BLAST is an acronym for Basic Local Alignment Sequencing Tool). The query sequence is that which you type or paste into the window, as in the following step. In this case, the query sequence is the *ADE1* base sequence from the red yeast strain.

b) Searching the DNA database: Now place the cursor into the window above which is written "Type or Paste a Query Sequence" and paste the sequence that you copied in step 1. Use either the Edit menu or hold down the apple key and the "V" key to paste. In the window "Choose one or more Sequence Data Sets" select "Open Reading Frames (DNA or Protein)". Next select the button "Run WU-BLAST". The query sequence (the *ADE1* sequence from the red yeast strain) is now being compared to all 10 million base pairs in the database of the standard (wild-type) white yeast strain to find the sequence or sequences that are most similar.

c) Matches to the query DNA sequence: After a short time you will see the results of this search although it will be difficult to interpret. Near the top of the page you will see what looks like a number line. This line represents the sequence you put in, with the numbers indicating the nucleotide number. Under this numbered line you will see one or more colored thick lines (actually they are arrows). Each thick line represents a match to the query sequence. In this case there is one match (there is one line), represented by a thick green arrow. Click on the line and the similarity between the Query sequence (the ADE1 gene from the red yeast) and the Subject sequence (the ADE1 gene from the white type of yeast) will be shown. Note the name for this Subject sequence, YAR015W ADE1 SGDID:S000000070. The name suggests that this is the ADE1 gene for the wild-type yeast (and there is an identification number analogous to a catalog number). Below the name for the Subject sequence is some information, including a P value. The higher the negative exponent (the number after the "e") of this P value, the better the match. 1.7e-116 is a very high level match. In order to see how closely the two base sequences match scroll down; you should see lines of base sequence (strings of A, C, G, and T) and hash-

marks between them. Note the similarity in the DNA sequence between the query (the *ADE1* gene of the red yeast) and the DNA sequence found in the database (labeled "Sbjct"; this is the *ADE1* gene of the white [wild-type] yeast). The query sequence starts with coordinate 12 and goes to coordinate 544; the "Sbjct" sequence starts at coordinate 1 and goes to coordinate 533. Note that each line of sequence alternates between the query and the sbjct sequence. A hash-mark between the query and sbjct sequences indicates that the base sequence is identical between the query and the sbjct at that coordinate. The top portion of the screen will look like this:



d) But is it an exact match? Recall that our hypothesis is that the difference between the white and red yeast strain is attributable to difference in the base sequence of the *ADE1* or *ADE2* genes. Such a difference could be fairly small so check carefully to see if the query sequence and the Sbjct sequence are exact matches.

i. *Do you see any differences? Any difference would mean that the base sequence of the* ADE1 *gene of the red yeast is different from the base sequence of the* ADE1 *gene of the white yeast. If a difference was observed, what treatment of the yeast in week 1 led to this difference in the base sequence?* As a reminder, the base sequence in the database (the "Sbjct") came from the standard (wild-type) white yeast.

Summary of Steps 1&2. At this point you have determined the sequence of bases that make up the *ADE1* gene from the red yeast, searched a database and compared the sequence of bases of the *ADE1* gene from the white yeast to the *ADE1* gene of the red yeast. In the next step you will use the base sequence of the *ADE1* gene from red yeast to determine the amino acid sequence of the protein encoded by the *ADE1* gene from the red yeast. You will then compare the amino acid sequences from the red and white yeasts.

3) Determining the amino acid sequence of the Ade1 protein from the red yeast. In this section you will use the program DNA Strider to determine the amino acid sequence of the protein encoded by the *ADE1* gene.

a) What's in a DNA sequence? Recall our definition of a gene: a gene is a segment of DNA that codes for a protein. As was the case last week, the computer shows you the sequence of the non-template strand of DNA, the strand that looks like the mRNA, except for the presence of T's instead of U's. This strand can be directly analyzed using a genetic code table to decipher what amino acids are encoded. The DNA code is written in groups of three bases. Note that the "reading frame" of the sentence is quite important. For instance, if I wrote a sentence of three letter words in the midst of a random assortment of letters it would be nonsense unless you started in the correct reading frame. The three reading frames of the sentence TETHECATATETHEBIGRATTT would be:

Reading frame (1) TET HEC ATA TET HEB IGR ATT T

(groups of three start at the first letter)

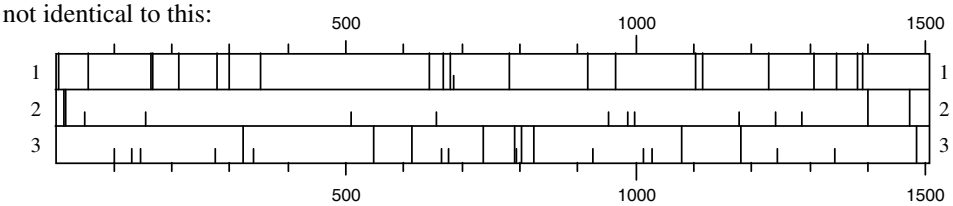
Reading frame (2) T ETH ECA TAT ETH EBI GRA TTT
(groups of three start at the second letter)

Reading frame (3) TE THE CAT ATE THE BIG RAT TT
(groups of three start at the third letter)

There is only one of three reading frames in which you will be able to decipher the meaning of this sentence. So in this example, unless you start reading in the "3rd" reading frame, this string of letters would have no meaning. Likewise, to decipher the amino acid sequence of a gene the information must be read in the correct "reading frame". The computer program DNA Strider will examine each of the three possible reading frames to determine which one has a "sentence". In this case a sentence means a long string of amino acids that are encoded without any "stop" codes. Recall from the homework assignment that you did last week that some 3-letter combinations represent a "stop"---that is, rather than an amino acid this stop signal results in the termination of translation. The codes for stop are TGA, TGG, and TAG.

b) What and where is an ORF? An "ORF" is an abbreviation for an open reading frame and is metaphorically the same as finding a sentence in the midst of a number of letters strung together. The computer program "DNA Strider" has tools designed for this purpose. Open DNA Strider by double-clicking on the DNA Strider icon in the "Applications" folder or on the CD. Note that it appears that nothing has happened because there is no open window but the program is open. There is now a menu bar that should not look familiar. Under the pull-down "File" select "New" and under "New" select "DNA". A window will open into which you will paste the DNA sequence of the *ADE1* gene from the red yeast. This DNA sequence should still be in the "clipboard" of the computer so that you should be able to simply "Paste" by simultaneously hitting the "Apple" and "V" keys. If this did not work then go back to the program "Editview" and reselect nucleotides 36-579, "copy" and then "paste" into the DNA Strider window.

Next, under the pull down "AA" (which stands for amino acids), select "ORF" and "3-phase". This operation will "decode" (translate) your DNA sequence into an amino acid sequence by analyzing the nucleotide sequence in groups of three, noting any cases of stop codes. The map that you are currently looking at is plotting this translation in each of the three possible reading frames the presence of "periods" (stop codes; TGA, TGG, and TAG). Each of the three reading frames is indicated by a number next to a rectangle. The rectangle represents the DNA sequence. Wherever stop codes occur in the sequence they are plotted as a solid vertical line from the top to the bottom of the reading frame's rectangle. An example of the output on a randomly chosen gene (i.e., not *ADE1*) is shown below. Your output should look similar but not identical to this:



An open reading frame would be a large area in which there are no long vertical lines. In the example provided above, the ORF is in the second reading frame. Do you see an ORF in the

ADE1 sequence? If you do, this means that this segment of DNA has the potential to code for a protein because there are no stops, just amino acids, encoded.

c) The amino acid sequence of the protein encoded by the *ADE1* gene of the red yeast. If you do see an ORF this indicates the potential for this sequence of bases to encode a protein. You will now determine the sequence of amino acids that make up the protein that is encoded by the ORF. The program DNA Strider will allow us to see this as well. Position the cursor between the second and third base of the sequence (between the A and the T so that the first three bases selected will be TTA) in the DNA Strider window. Highlight all of the bases starting from the T. Then under the "Conv" (Convert) pulldown select "Prot 5' → 3'". This is an abbreviation for "Protein sequence encoded by the DNA in the 5' to 3' direction". A new window will appear that is filled with an amino acid sequence. This is the sequence of amino acids in the protein encoded by the red yeast *ADE1* DNA that you sequenced. The amino acids are given in one-letter abbreviations. For example, the amino acid serine is abbreviated with the letter S and the amino acid leucine is abbreviated with the letter L. If you have succeeded then the last five amino acids using the one letter abbreviation should be "NISPA".

ii. Given what you learned in part 3(b), why was the cursor placed between the 2nd and 3rd base of the sequence? Would the result have been different if you had placed the cursor between the first and second bases? [Also see example under part 3(a)].

Deleted: *Would the result have been different if you had placed the cursor between the first and second bases?*

Deleted: *Would the result have been different if you had placed the cursor between the first and second bases?*

4) Comparing the amino acid sequence encoded by the *ADE1* gene from the red yeast to the amino acid sequence encoded by the *ADE1* gene from the white yeast: The above sequence of amino acids is not by itself very informative. It is more important to compare the amino acid sequence of this protein in the red yeast to the amino acid sequence in the standard white yeast. This can be done using a program that was used before, BLAST. BLAST can be used to search databases for either DNA sequences (as in step 2) or protein sequences (what we are about to do). To do this, highlight and copy the amino acid sequence in DNA Strider from step 3. Now go back to the http://www.yeastgenome.org web address and choose BLAST, or use your back button to back up one page. Delete any text that might be in the "Type or Paste Query Sequence" window. Next paste the amino acid sequence that you just copied into this window. You will now search the database for any protein from the standard (wild-type) white yeast whose amino acid sequence is identical or similar to the one that you just determined was encoded by the *ADE1* gene of the red yeast. Under the heading of "Choose the appropriate BLAST program" select "BLASTP". Under the next heading "Choose one or more Sequence Datasets" select "Open Reading Frames (DNA or Protein). These steps allow you to compare the red yeast protein sequence encoded by the *ADE1* gene to all of the standard (wild-type) white yeast proteins in the database. Then select the "Run WU-BLAST" option.

Results of the search: The results of this search are depicted in a way that is similar to step 2. The results are shown schematically in the new window. The thick lines are color-coded to indicate the degree to which the query sequence and the target sequence match. In this case, there is only one "good" match shown in green—each blue line corresponds to an amino acid sequence whose similarity is so weak as to be of no or little significance. The amino acid sequence of the Ade1 protein from the red yeast and the match to that sequence from the database are shown lower on the page---you can get there simply by clicking on the thick green line. At the top of this page is a title of the Subject sequence (the one that was found in the database). Notice that ADE1 is in the title (as well as a catalog number; this verifies that the sequence that has the best similarity to the Ade1 amino acid sequence from red yeast is the Ade1 protein from the white yeast). Just below the title of the Subject sequence you will see an alignment of letters, presented as sets with three lines of

letters each. Again, each letter is the single letter abbreviation for an amino acid. Thus, there are 20 different letters that you could see. In this case, hash marks are not used to indicate amino acids that are identical between the two sequences. Instead the top line of each set of three is your query sequence – the protein encoded by the *ADE1* gene from the red yeast. The bottom line of each set of three is the "Sbjct" sequence, the protein encoded by the white yeast DNA in the database. The middle line indicates the amino acids that are shared by both the query and Sbjct sequences – i.e., if both Sbjct and query have a "V" (for the amino acid valine) then a "V" will be shown in the middle line. Gaps in the middle line indicate that the two sequences do not match at this position.

iii. ***Are the two amino acid sequences identical or are there differences?*** *If there are differences then address the following questions: what amino acid number in the chain is different? Provide the single letter abbreviation for the name of the amino acid that is present at that position in the ADE1 gene from the red yeast. In order to figure this out you need to go back to the program DNA Strider and under the pulldown "AA" select "Genetic Code". A table will appear that has the one letter designations and the corresponding amino acid. What would happen to the amino acid chain at this position of the protein? Would this mutation affect the size of the protein? If so, how large would the ADE1 protein be for the red type of yeast? How large is the ADE1 protein for the white type of yeast? The homework that you did last week categorized mutations as "nonsense"," missense", "silent", or "frameshift". Which category does this fit into? Is the mutation likely to harm the ability of the ADE1 protein to carry out its function? Explain. Is the difference in the amino acid sequence of the Ade1 protein in the red strain compared to the white strain a reflection of a difference in the DNA sequence between the ADE1 gene of the red strain compared to the white strain? Explain your answer.*

If the two amino acid sequences are identical, is it likely that the difference in color between the red and white strains is attributable to a difference in the ADE1 gene or corresponding protein? Explain.

5) Examining the database for a human gene that encodes a (weakly) similar protein that carries out the same function. We are using a lowly single-celled organism to learn about adenine biosynthesis---that is, how the critical compound adenine is made inside a cell. One reason to use such an organism is that it offers many advantages over more complex ones. For instance, it grows exceedingly fast. In 24 hours one cell will grow to more than 60,000 cells. This makes it simple to gather sufficient material on which to do experiments. In cases where one is examining different conditions for growth or absence of growth the experiment can be done in one day. Another advantage is the ability to treat the organism in a variety of ways. Consider the treatments you have done to the yeast during the last several weeks. It is unlikely that you could (or should) treat a vertebrate organism in the same way.

But do yeast cells have anything in common with cells from more complex organisms? We can make one attempt at answering this question by asking whether cells from humans have a protein that is similar to the protein encoded by the *ADE1* gene of yeast. If the answer to this question is "yes" then it seems reasonable to conclude that both yeast cells and human cells make their adenine in similar ways. Furthermore, this would then suggest that studying adenine biosynthesis in yeast is an adequate way to begin to understand adenine biosynthesis in humans.

In order to determine if human cells have a protein that is similar to the protein encoded by the *ADE1* gene of yeast we will search a database that has amino acid sequences of proteins that have determined by scientists thus far in any type of organism. In your web browser go click on the link "SGD locus" (this is just below the title of the sequence, above the Ade1 amino acid alignment). This should take you to the page titled ADE1/YAR015W that you saw last week. On the right hand side of this webpage are a series of titles in bold font, each with a pull down menu from which to select. You will have to scroll down a bit to see the option "Comparison Resources"; under this option select "BLASTP at NCBI". Now click the "View" button. The computer is now poised to look for any amino acids sequences from any organism which are similar to the yeast Ade1 protein. We would like, however, to limit our search to proteins that are found in humans. To do this, go to the section "Choose Search Set" and in the window titled "Organism" type "Homo sapiens". There will then appear two options from which to choose---select the one "Homo sapiens (taxid:9606)". Now scroll down a bit and click on the blue "BLAST" button. The search is now on for any amino acid sequences from mammals that bear a resemblance to the amino acid sequence of *ADE1*. The page will refresh one or more times; it may take a little while for the search to be completed. When it is complete there will be a page that is titled "ref|NP_009409.1| (306 letters)" with a blue box as a background. In order to see the degree of similarity of the yeast protein compared to the human version scroll down until you get to the section labeled "Alignments". Below this is the alignment between the yeast Ade1 protein (Query) and the human version of this protein (Subject). As before, the "query" sequence is the sequence we entered (in this case Ade1 protein from the white yeast) and the "Sbjct" is a sequence in the database (in this case the closest match in humans to the yeast *ADE1*). The middle line contains some letters and some plus signs. Note that the letters appear at each position where the two sequences are identical. Although it is less obvious unless you know the structures of the amino acids, a "+" appears where the amino acids are similar to each other in structure – an L and a V, for example. So the areas with the highest density of letters and + symbols are the areas that are the most similar between the yeast and human proteins. Note that the degree of similarity is not the same throughout the protein – the beginning (N-terminus), for example, is fairly similar between the two sequences (GK++++YE+). The most similar regions are usually especially important in the function of the protein.

Examine the comparison. Near the end of the yeast protein amino acid sequence of *ADE1* there is a sequence that is DKQFLRD. There is a similar sequence in the human protein DKQSYRD. One could argue that the similarity between the yeast and human proteins is due to chance. But notice the "P" value associated with the match between the yeast and human amino acid sequence. This value estimates the probability that such an amino acid sequence match could be found by random chance. This value is fairly low ($4e^{-08}$; = one in 25 million) suggesting that there is some significance to this match.

What is known about this human protein? In order to find out click on the link "GENE ID: 10606 PAICS". At the top of the new screen is the name of the human protein. It has alternative names but one is phosphoribosylaminoimidazole succinocarboxamide synthetase which is itself an alternative name for SAICAR synthetase. You saw last week that SAICAR synthetase is the name for the protein encoded by the yeast *ADE1* gene. Furthermore, the text of the Summary section notes that this protein "catalyzes steps…in purine biosynthesis". Therefore we have found that the protein from humans that is similar in acid <u>sequence</u> to the yeast Ade1 protein also has the same <u>function</u> – synthesizing the intermediate SAICAR in the pathway that manufactures adenine. Thus experiments investigating the yeast enzyme might tell us important things about the human enzyme.

You have now completed an analysis of the *ADE1* gene from the red yeast and compared it to the *ADE1* gene from the white yeast. You have identified the base sequence of the *ADE1* gene from the red yeast, you have identified the *ADE1* base sequence from the standard (wild-type) white yeast, and you have compared these two base sequences. You have also determined the sequence of amino acids that is encoded by the *ADE1* gene from the red yeast and compared it to the amino acid sequence of the protein encoded by the *ADE1* gene from the white yeast.

Below are the objectives and procedures for studying the *ADE2* gene.

1) Examining the primary data of the *ADE2* gene
2) Comparing the nucleotide sequence of the *ADE2* gene from the red yeast to the *ADE2* gene from the white yeast
3) Determining the amino acid sequence encoded by the *ADE2* gene
4) Comparing this amino acid sequence to the sequence from the standard (wild-type) white strain
5) Examining the database for a human gene that encodes a similar (although weakly) protein that carries out the same function

1) Examining the primary data of the *ADE2* gene

Open up the file labeled "ADE2f.seq" using the program "Editview". To do this, double-click on the "Editview" icon, and under the pulldown heading of "File" select "Open". Then select the file name "ADE2f.seq". This computer file contains the sequence data we obtained from the PCR product generated with the ADE2 primer set in the previous lab session. This is the sequence of bases corresponding to the *ADE2* gene from the red yeast strain. At this point you should see peaks of four different colors. There is also an X-axis that consists of a time unit of measurement, a nucleotide number, and a nucleotide (designated A, C, G, or T). There are a few things to note:

i) at many points there is only one peak and that the color of the peak determines the nucleotide indicated on the X-axis.

ii) at some points, there is more than one peak. This is due to technical problems and makes it difficult or impossible to predict the base that is at that position on the DNA strand. This is why there is occasionally an "N" instead of an A, C, G, or T. "N" designates that the identity of the base at this position was not unambiguously determined.

iii) it is technically more difficult to determine the base sequence near the beginning and thus there are some "N's" in this area. It is also technically difficult to determine the base sequence much more than about 500 nucleotides from the start. Therefore we will limit our analysis to the base sequence between position 14 and 630.

In the bottom left corner of the window there are four icons. One of them reads "CATG". Click once on this icon. This changes the information that you were looking at into text only, rather than the multicolored peaks with the various other annotations. Notice all of the N's at the beginning and a few of them near the end of the text string. Highlight the characters between position 14 and 630 (from GTTGAGG to GTTATGC) and then select "copy" from the "Edit" pull-down menu. You will now move this sequence of bases into a program that allows a search of the *Saccharomyces* database for any similar sequences.

2) Comparing the sequence of the *ADE2* gene from the red yeast to the sequence of the *ADE2* gene from the white yeast. In the following steps you will search SGD, the database that has the DNA sequence of all genes from the standard (wild-type) white yeast that you used last week, to see if any have the same, or similar, sequence of bases as the sequence you have determined for the *ADE2* gene from the red yeast.

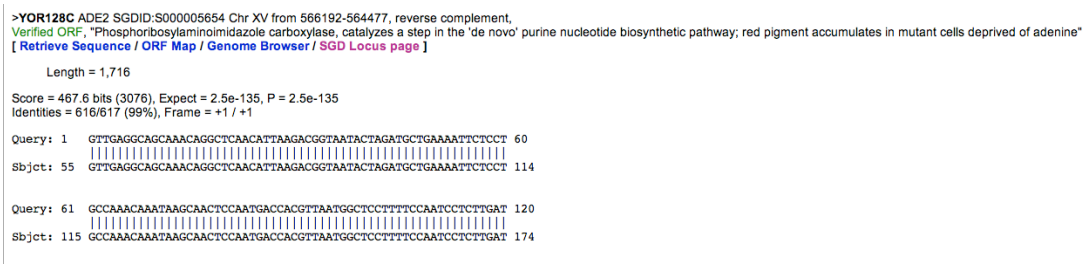
a) Accessing the database via the internet

Open Netscape and type in the following address
http://www.yeastgenome.org

At this website, the entire base sequence of the standard (wild-type) white strain of *Saccharomyces cerevisiae* is stored. This represents about 10 million base pairs. Nested somewhere in this 10 million base pairs is the sequence for the *ADE2* gene (only about 2000 base pairs long). At the top of this web page within the beige bar select "BLAST". This program will search the entire sequence of bases of the *Saccharomyces cerevisiae* in attempt to find some sequence that is either identical or similar to the "query" sequence (BLAST is an acronym for Basic Local Alignment Sequencing Tool). The query sequence is that which you type or paste into the window, as in the following step. In this case, the query sequence is the *ADE2* base sequence from the red yeast strain.

b) Searching the DNA database: Now place the cursor into the window above which is labeled "Type or Paste a Query Sequence" and paste the sequence that you copied in step 1. Use either the Edit menu or hold down the apple key and the "V" key to paste. Under the heading "Choose one or more Sequence Datasets" choose the option "Open Reading Frames (DNA or Protein). Next select the button "Run WU-BLAST". The query sequence (the *ADE2* sequence from the red yeast strain) is now being compared to all 10 million base pairs in the database of the standard (wild-type) white yeast strain to find the sequence or sequences that are most similar.

c) Matches to the query DNA sequence: After a short time you will see the results of this search although it will be difficult to interpret. At the top of the page you will see a thick purple line. This line is a graphic representation of a segment of DNA that closely matches the query sequence. If you can mouse over the purple line you will see in the box above it a description of the nature of the DNA that closely matches the query sequence. This description starts with "p=2.5e-135 s=3076 YOR128C|ADE2 SGDID:S000005654…". The "2.5e-135" is a representation of how closely the query sequence and the sequence in the database match. The higher the negative exponent (the number after the "e"), the better the match. This is an excellent match. Note that in this description there is also the text "ADE2" suggesting that the Subject sequence is indeed the ADE2 gene from the white type of yeast. In order to see how closely the two base sequences match scroll down; you should see lines of base sequence (strings of A, C, G, and T) and hash-marks between them. Note the similarity in the DNA sequence between the query (the *ADE2* gene of the red yeast) and the DNA sequence found in the database (labeled "Sbjct"; this is the *ADE2* gene of the white [wild-type] yeast). The query sequence starts with coordinate 1 and goes to coordinate 617; the "Sbjct" sequence starts at coordinate 55 and goes to coordinate 671. Note that each line of sequence alternates between the Query and the Sbjct sequence. A hash-mark between the Query and Sbjct sequences indicates that the base sequence is identical between the Query and the Sbjct at that coordinate. The top portion of the screen should look like this:



d) But is it an exact match? Recall that our hypothesis is that the difference between the white and red yeast strain is attributable to difference in the base sequence of the *ADE1* or *ADE2*

genes. Such a difference could be fairly small so check carefully to see if the Query sequence and the Sbjct sequence are exact matches.

i. *Do you see any differences? Any difference would mean that the base sequence of the ADE2 gene of the red yeast is different from the base sequence of the ADE2 gene of the white yeast. If a difference was observed, what treatment of the yeast in week 1 led to this difference in the base sequence?* As a reminder, the base sequence in the database (the "Sbjct") came from the standard (wild-type) white yeast.

Summary of Steps 1&2. At this point you have determined the sequence of bases that make up the *ADE2* gene from the red yeast, searched a database and compared the sequence of bases of the *ADE2* gene from the white yeast to the *ADE2* gene of the red yeast. In the next step you will use the base sequence of the *ADE2* gene from red yeast to determine the amino acid sequence of the protein encoded by the *ADE2* gene from the red yeast. You will then compare the amino acid sequences from the red and white yeasts.

3) Determining the amino acid sequence of the Ade2 protein from the red yeast. In this section you will use the program DNA Strider to determine the amino acid sequence of the protein encoded by the *ADE2* gene.

a) What's in a DNA sequence? Recall our definition of a gene: a gene is a segment of DNA that codes for a protein. As was the case last week, the computer shows you the sequence of the non-template strand of DNA, the strand that looks like the mRNA, except for the presence of T's instead of U's. This strand can be directly analyzed using a genetic code table to decipher what amino acids are encoded. The DNA code is written in groups of three bases. Note that the "reading frame" of the sentence is quite important. For instance, if I wrote a sentence of three letter words in the midst of a random assortment of letters it would be nonsense unless you started in the correct reading frame. The three reading frames of the sentence TETHECATATETHEBIGRATTT would be:

Reading frame (1) TET HEC ATA TET HEB IGR ATT T
(groups of three start at the first letter)

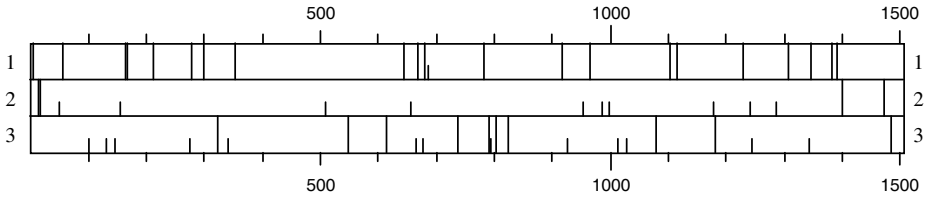
Reading frame (2) T ETH ECA TAT ETH EBI GRA TTT
(groups of three start at the second letter)

Reading frame (3) TE THE CAT ATE THE BIG RAT TT
(groups of three start at the third letter)

There is only one of three reading frames in which you will be able to decipher the meaning of this sentence. So in this example, unless you start reading in the "3rd" reading frame, this string of letters would have no meaning. Likewise, to decipher the amino acid sequence of a gene the information must be read in the correct "reading frame". The computer program DNA Strider will examine each of the three possible reading frames to determine which one has a "sentence". In this case a sentence means a long string of amino acids that are encoded without any "stop" codes. Recall from the homework assignment that you did last week that some 3-letter combinations represent a "stop"---that is, rather than an amino acid this stop signal results in the termination of translation. The codes for stop are TGA, TGG, and TAG.

b) What and where is an ORF? An "ORF" is an abbreviation for an open reading frame and is metaphorically the same as finding a sentence in the midst of a number of letters strung together. The computer program "DNA Strider" has tools designed for this purpose. Open DNA Strider by double-clicking on the DNA Strider icon in the "Applications" folder or on the CD. Note that it appears that nothing has happened because there is no open window but the program is open. There is now a menu bar that should not look familiar. Under the pull-down "File" select "New" and under "New" select "DNA". A window will open into which you will paste the DNA sequence of the *ADE2* gene from the red yeast. This DNA sequence should still be in the "clipboard" of the computer so that you should be able to simply "Paste" by simultaneously hitting the "Apple" and "V" keys. If this did not work then go back to the program "Editview" and reselect nucleotides 14-630, "copy" and then "paste" into the DNA Strider window.

Next, under the pull down "AA" (which stands for amino acids), select "ORF" and "3-phase". This operation will "decode" (translate) your DNA sequence into an amino acid sequence by analyzing the nucleotide sequence in groups of three, noting any cases of stop codes. The map that you are currently looking at is plotting this translation in each of the three possible reading frames the presence of "periods" (stop codes; TGA, TGG, and TAG). Each of the three reading frames is indicated by a number next to a rectangle. The rectangle represents the DNA sequence. Wherever stop codes occur in the sequence they are plotted as a solid vertical line from the top to the bottom of the reading frame's rectangle. An example of the output on a randomly chosen gene (i.e., not *ADE2*) is shown below. Your output should look similar but not identical to this:



An open reading frame would be a large area in which there are no long vertical lines. In the example provided above, the ORF is in the second reading frame. Do you see an ORF in the *ADE2* sequence? If you do, this means that this segment of DNA has the potential to code for a protein because there are no stops, just amino acids, encoded.

c) The amino acid sequence of the protein encoded by the *ADE2* gene of the red yeast. If you do see an ORF this indicates the potential for this sequence of bases to encode a protein. You will now determine the sequence of amino acids that make up the protein that is encoded by the ORF. The program DNA Strider will allow us to see this as well. Position the cursor at the start of the sequence (the cursor should be at the very start so that the first three nucleotides are GTT) in the DNA Strider window. Highlight all of the bases starting from the G. Then under the "Conv" (Convert) pulldown select "Prot 5' → 3'". This is an abbreviation for "Protein sequence encoded by the DNA in the 5' to 3' direction". A new window will appear that is filled with an amino acid sequence. This is the sequence of amino acids in the protein encoded by the red yeast *ADE2* DNA that you sequenced. The amino acids are given in one-letter abbreviations. For example, the amino

acid serine is abbreviated with the letter S and the amino acid leucine is abbreviated with the letter L. If you have succeeded then the last five amino acids using the one letter abbreviation should be "CDLCY".

ii. Given what you learned in part 3(b), why was the cursor placed at the start of the sequence? Would the outcome have differed if the cursor had been placed between the first and second characters (the G and the T)? [Also see example under part 3(a)].

4) Comparing the amino acid sequence encoded by the *ADE2* gene from the red yeast to the amino acid sequence encoded by the *ADE2* gene from the white yeast: The above sequence of amino acids is not by itself very informative. It is more important to compare the amino acid sequence of this protein in the red yeast to the amino acid sequence in the standard white yeast. This can be done using a program that was used before, BLAST. BLAST can be used to search databases for either DNA sequences (as in step 2) or protein sequences (what we are about to do). To do this, highlight and copy the amino acid sequence in DNA Strider from step 3. Now go back to the http://www.yeastgenome.org web address and choose BLAST, or use your back button to back up one page. Delete any text that might be in the "Type or Paste Query Sequence" window. Next paste the amino acid sequence that you just copied into this window. You will now search the database for any protein from the standard (wild-type) white yeast whose amino acid sequence is identical or similar to the one that you just determined was encoded by the *ADE2* gene of the red yeast. Under the heading of "Choose the appropriate BLAST program" select "BLASTP". Under the heading "Choose one or more Sequence Datasets" choose "Open Reading Frames (DNA or Protein). These steps allow you to compare the red yeast protein sequence encoded by the *ADE2* gene to all of the standard (wild-type) white yeast proteins in the database. Then select the "Run WU-BLAST" option.

Results of the search: The results of this search are depicted in a way that is similar to step 2. The results are shown schematically at the top of the page. The thick lines are color-coded to indicate the degree to which the query sequence and the target sequence match. In this case, there is only one "good" match shown in purple—each blue line corresponds to an amino acid sequence whose similarity is so weak as to be of no or little significance. The amino acid sequence of the Ade2 protein from the red yeast and the match to that sequence from the database are shown lower on the page---you can get there simply by clicking on the thick purple line. At the top of this page is a title of the Subject sequence (the one that was found in the database). Notice that ADE2 is in the title (as well as a catalog number; this verifies that the sequence that has the best similarity to the Ade2 amino acid sequence from red yeast is the Ade2 protein from the white yeast). Just below the title of the Subject sequence you will see an alignment of letters, presented as sets with three lines of letters each. Again, each letter is the single letter abbreviation for an amino acid. Thus, there are 20 different letters that you could see. In this case, hash marks are not used to indicate amino acids that are identical between the two sequences. Instead the top line of each set of three is your query sequence – the protein encoded by the *ADE2* gene from the red yeast. The bottom line of each set of three is the "Sbjct" sequence, the protein encoded by the white yeast DNA in the database. The middle line indicates the amino acids that are shared by both the query and Sbjct sequences – i.e., if both Sbjct and query have a "V" (for the amino acid valine) then a "V" will be shown in the middle line. Gaps or a "+" sign in the middle line indicate that the two sequences do not match at this position.

iii. Are the two amino acid sequences identical or are there differences? If there are differences, then address the following questions: What amino acid number in the chain is

different? Provide the single letter abbreviation for the name of the amino acid that is present at that position in the ADE2 gene from the red yeast. In order to figure this out you need to go back to the program DNA Strider and under the pulldown "AA" select "Genetic Code". A table will appear that has the one letter designations and the corresponding amino acid. Note that AMB, OCH, and OPA are odd names for stop codons (amber, ochre, and opal) and are also represented by the single letter code "Z". What would happen to the amino acid chain at this position of the protein? Would this mutation affect the length of the protein? If so, how large would the Ade2 protein be for the red type of yeast? How large is the Ade2 protein for the white type of yeast? Would this mutation be likely to harm the Ade2 protein's ability to carry out its function? Explain. The homework that you did in a previous week categorized mutations as "nonsense"," missense", "silent", or "frameshift". Which category does this fit into? Is the difference in the amino acid sequence of the Ade2 protein in the red strain compared to the white strain a reflection of a difference in the DNA sequence between the ADE2 gene of the red strain compared to the white strain? Explain your answer.

If the two amino acid sequences are identical, is it likely that the difference in color between the red and white strains is attributable to a difference in the ADE2 gene or corresponding protein? Explain.

5) Examining the database for a human gene that encodes a (weakly) similar protein that carries out the same function. We are using a lowly single-celled organism to learn about adenine biosynthesis---that is, how the critical compound adenine is made inside a cell. One reason to use such an organism is that it offers many advantages over more complex ones. For instance, it grows exceedingly fast. In 24 hours one cell will grow to more than 60,000 cells. This makes it simple to gather sufficient material on which to do experiments. In cases where one is examining different conditions for growth or absence of growth the experiment can be done in one day. Another advantage is the ability to treat the organism in a variety of ways. Consider the treatments you have done to the yeast during the last several weeks. It is unlikely that you could (or should) treat a vertebrate organism in the same way.

But do yeast cells have anything in common with cells from more complex organisms? We can make one attempt at answering this question by asking whether cells from humans have a protein that is similar to the protein encoded by the *ADE2* gene of yeast. If the answer to this question is "yes" then it seems reasonable to conclude that both yeast cells and human cells make their adenine in similar ways. Furthermore, this would then suggest that studying adenine biosynthesis in yeast is an adequate way to begin to understand adenine biosynthesis in humans.

In order to determine if human cells have a protein that is similar to the protein encoded by the *ADE2* gene of yeast we will search a database that has amino acid sequences of proteins that have been determined by scientists thus far in any type of organism. In your web browser near the top of the alignment (just below the title), click on the link "SGD Locus page". Scroll down on the new page and on the right side of the page look for the heading "Comparison Resources". In that window select from the pulldown menu "BLASTP at NCBI". Then select "View". We would like, however, to limit our search to proteins that are found in humans. To do this, go to the section "Choose Search Set" and in the window titled "Organism" type "Homo sapiens". There will then appear two options from which to choose---select the one "Homo sapiens (taxid:9606)". Next, scroll down to where it says "Filters and Masking" and

uncheck the box that reads "Low complexity regions". You will now have to scroll back up a bit and click on the blue "BLAST" button. The search is now on for any amino acid sequences from mammals that bear a resemblance to the amino acid sequence of Ade2. The page will refresh one or more times. After a few seconds a new screen will load; the new page will have at the top "ref|NP_014771.1| (571 letters)" with a light blue box as background. You will need to scroll down to the "Alignments" section to find the comparison between the yeast Ade2 protein (Query) and a human protein that is similar (Subject). This comparison is just below the link "GENE ID: 10606 PAICS."

What is known about this human protein that is similar to the yeast Ade2 protein? To find out, click on the link "GENE ID: 10606 PAICS". This new page has information about the human protein. At the top the name of the protein is provided in bold: phosphoribosylaminoimidazole carboxylase. This is the same name as the protein encoded by the yeast *ADE2* gene and tells us that this human protein catalyzes the same chemical reaction that corresponding yeast protein does. Similarly the text under the heading "Summary" states that this protein "catalyzes steps…of purine biosynthesis." Thus this alignment indicates that we have found a protein from humans which has the same name (and the same function) as the enzyme encoded by the *ADE2* gene of yeast. Thus experiments investigating the yeast enzyme might tell us important things about the human enzyme.

Hit the back button on your browser to return to the alignment page. Scroll down on the resulting page until you see the alignment of the yeast and the human protein. As before, the "query" sequence is the sequence we entered (in this case Ade2 protein from the white yeast) and the "Sbjct" is a sequence in the database (in this case the closest match in humans to the yeast Ade2). The middle line contains some letters and some plus signs. Note that the letters appear at each position where the two sequences are identical. Although it is less obvious unless you know the structures of the amino acids, a "+" appears where the amino acids are similar to each other in structure – an L and a V, for example. So the areas with the highest density of letters and + symbols are the areas that are the most similar between the yeast and human proteins.

Examine the comparison. One could argue that the similarity between the yeast and human proteins is due to chance. But notice the "expect value" associated with the match between the yeast and human amino acid sequence. This value estimates the probability that such an amino acid sequence match could be found by random chance. This value (1e-04) is fairly low suggesting that there might be some significance to this match. Note that the degree of similarity is not the same throughout the protein – the region ++MGS-SDL near the beginning of the alignment, for example, is quite similar between the two sequences. The most similar regions are usually especially important in the function of the protein.

***Before you leave the lab*, exchange data with your lab partner by filling out the sheet on the next page and having the instructor check it. Having data for both the *ADE1* and *ADE2* genes will be crucial in order to write the lab report.**

SUMMARY: You and your partner have examined the base sequence of two genes *ADE1* and *ADE2* from the red yeast. The reason that you examined these two genes is that previous

research has indicated that changes in either of these two genes can give rise to yeast cells that have a red pigment. Could this be true in the case of the red yeast that you observed after ultraviolet light treatment? You have compared the sequence of the *ADE1* gene and the *ADE2* gene from the red yeast to the corresponding genes from the white yeast. Were there differences between the red and white yeast in either *ADE1* or *ADE2*? In both *ADE1* and *ADE2*? You have also examined the sequence of amino acids encoded by the *ADE1* and *ADE2* genes from the red yeast and compared these sequences to the corresponding amino acid sequences from the white yeast. Were there amino acid sequence differences between proteins encoded by the *ADE1* and *ADE2* genes in red and white yeast?

QUESTIONS FOR DISCUSSION: <u>Before you leave the lab</u>, please meet with your lab partner to answer the following questions. Write the answers below and keep this page for your future reference after showing it to the instructor or TA.

1) You should have seen some difference between the red yeast DNA and the white yeast DNA in the database. What difference did you see? Write here:

The gene that had the difference (was it *ADE1* or *ADE2*?):

The position of the base that was changed and what that change was:
Base number:

Base seen at that position in white yeast DNA in the database.

Base seen at that position in red yeast DNA that you sequenced:

2) Was there a corresponding alteration in the protein encoded by the altered gene? What difference did you see? Write here:

The position of the amino acid that was altered and the type of alteration:
Amino acid number:

Amino acid seen at that position in white yeast in the database:

Amino acid seen at that position in the red yeast:

Nature of the mutation (silent, missense, nonsense, frameshift):

Instructor signature:_____

Lab assignment
Due at the start of lab in week 6
10 points

Using a word processor (no handwritten assignments will be accepted), address each of the following questions. Each student will have to address these questions for both *ADE1* and *ADE2*. Because during the computer lab you only examined one of these two genes you will have to confer with your lab partner to answer all of these questions. These questions were taken from the text of the lab instructions. Look there to find the proper context of the question. This assignment is valued at 10 points and is due at the start of lab in week 6.

For the agarose gel:

i. Estimate the size of your DNA fragments generated by PCR. Remember you can only estimate the size---you should be able to state with confidence that the size is within a certain range. How do these sizes compare to the sizes of the ADE1 and ADE2 genes from white yeast that you calculated in the computer lab last week (this calculation is exact---to the base-pair)?

ii. Compare the distance traveled by the DNA from each of the two PCR samples. Additionally, compare these distances with the distance traveled by the DNA you recovered from the red yeast in the third week of lab. Based on the gel, are the ADE genes smaller or larger than the total yeast DNA pieces you isolated in the third week of lab? Does this make sense?

For ADE1:

i. *Do you see any differences? Any difference would mean that the base sequence of the ADE1 gene of the red yeast is different from the base sequence of the ADE1 gene of the white yeast. If a difference was observed, what treatment of the yeast in week 1 led to this difference in the base sequence?* As a reminder, the base sequence in the database (the "Sbjct") came from the standard (wild-type) white yeast.

ii. Given what you learned in part 3(b), why was the cursor placed between the 2nd and 3rd base of the sequence? [Also see example under part 3(a)].

iii. ***Are the two amino acid sequences identical or are there differences?*** *If there are differences then address the following questions: what amino acid number in the chain is different? Provide the single letter abbreviation for the name of the amino acid that is present at that position in the ADE1 gene from the red yeast. In order to figure this out you need to go back to the program DNA Strider and under the pulldown "AA" select "Genetic Code". A table will appear that has the one letter designations and the corresponding amino acid. What would happen to the amino acid chain at this position of the protein? Would this mutation affect the size of the protein? If so, how large would the ADE1 protein be for the red type of yeast? How large is the ADE1 protein for the white type of yeast? The homework that you did last week categorized mutations as "nonsense"," missense", "silent", or "frameshift". Which category does this fit into? Is the mutation likely to harm the ability of the ADE1 protein to carry out its function? Explain. Is the difference in the amino acid sequence of the Ade1 protein in the red strain compared to the white strain a reflection of a difference in the DNA sequence between the ADE1 gene of the red strain compared to the white strain? Explain your answer.*

If the two amino acid sequences are identical, is it likely that the difference in color between the red and white strains is attributable to a difference in the ADE1 gene or corresponding protein? Explain.

For ADE2

i. *Do you see any differences? Any difference would mean that the base sequence of the ADE2 gene of the red yeast is different from the base sequence of the ADE2 gene of the white yeast. If a difference was observed, what treatment of the yeast in week 1 led to this difference in the base sequence?* As a reminder, the base sequence in the database (the "Sbjct") came from the standard (wild-type) white yeast.

ii. Given what you learned in part 3(b), why was the cursor placed at the start of the sequence? Would the outcome have differed if the cursor had been placed between the first and second characters (the G and the T)? [Also see example under part 3(a)].

iii. Are the two amino acid sequences identical or are there differences? If there are differences, then address the following questions: What amino acid number in the chain is different? Provide the single letter abbreviation for the name of the amino acid that is present at that position in the ADE2 gene from the red yeast. In order to figure this out you need to go back to the program DNA Strider and under the pulldown "AA" select "Genetic Code". A table will appear that has the one letter designations and the corresponding amino acid. Note that AMB, OCH, and OPA are odd names for stop codons (amber, ochre, and opal) and are also represented by the single letter code "Z". What would happen to the amino acid chain at this position of the protein? Would this mutation affect the length of the protein? If so, how large would the Ade2 protein be for the red type of yeast? How large is the Ade2 protein for the white type of yeast? Would this mutation be likely to harm the Ade2 protein's ability to carry out its function? Explain. The homework that you did in a previous week categorized mutations as "nonsense"," missense", "silent", or "frameshift". Which category does this fit into? Is the difference in the amino acid sequence of the Ade2 protein in the red strain compared to the white strain a reflection of a difference in the DNA sequence between the ADE2 gene of the red strain compared to the white strain? Explain your answer.

If the two amino acid sequences are identical, is it likely that the difference in color between the red and white strains is attributable to a difference in the ADE2 gene or corresponding protein? Explain.