

Name: _____

Cellulose Synthase Genes: Where do we start?

Table 1: One of the conserved amino acid sequence of all cellulose synthase genes is DDG. Use the information in Figure 1 and write the full name for each of these amino acids.

1-letter abbreviation	D	D	G
Amino acid name			

Table 2: Possible codons which could product the indicated amino acids.

	Aspartic Acid	Glycine
List all possible codons for these amino acids:		

Table 3: Determining a DNA sequence that is complementary to the given piece of RNA which would code for the production of the DDG portion of a CESA polypeptide.

Amino acid sequence	Aspartic Acid—Aspartic acid ---Glycine
One possible RNA sequence	GACGAUGGG
Complementary DNA sequence	

Table 4: For any section of DNA sequence submitted to one of the databases, the position of the proper reading frame is initially unknown. Until the sequence is analyzed, it is also unknown whether the sequence is from the sense or antisense strand of the DNA molecule. You will analyze a small section to determine the proper reading frame and if it is the sense or antisense strand of DNA. Follow the models. Use the codon chart in Figure 2 to determine the amino acids.

The following is a small sample sequence of nucleotides was submitted to a database:

GGCTGCTACCCT

In the spaces below, translate the six possible amino acid sequences for which this might code. Remember, when reading the codon chart, substitute uracil where you see thymine:

Reading frame 1, as if this was the sense strand
GGC TGC TAC CCT → glycine – cysteine – tyrosine - proline
Reading frame 2, as if this was the sense strand
G GCT GCT ACC CT →
Reading frame 3, as if this was the sense strand
GG CTG CTA CCC T →

The complement of the submitted sequence, Reading frame 1, as if this was the sense strand

CCG ACG ATG GGA→

The complement of the submitted sequence, Reading frame 2, as if this was the sense strand

C CGA CGA TGG GA →

The complement of the submitted sequence, Reading frame 3, as if this was the sense strand

CC GAC GAT GGG A →

Now, if someone submitted a query for the amino acid sequence DDG, which of these would code for that?

BLAST Questions:

1. Are there any of the results that show 100% identity to the sequence from your query? Why or why not? If so, what species are they from?

2. Look at the list and choose one result that does not have 100% identity. Write the species name here:

Look up the common name of this plant. Write it here:

3. Click on the name of the plant. This takes you to a screen where we can learn more about the part of the DNA sequence that was found to align with the sequence that we queried. Here there are also links to additional information. Click on "GenBank". What kind of information is available here?

Why do you think that this information must be submitted whenever sequence is submitted to the Gene Bank?