

# CS313 Exercise 1 Cover Page

Name(s): \_\_\_\_\_

In the *TIME* column, please estimate the time you spent on the parts of this exercise. Please try to be as accurate as possible; this information will help us to design future exercises.

<b>PART</b>	<b>TIME</b>	<b>SCORE</b>
Exercise		



		Second letter				
		U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U	
	UUC } Phe	UCC } Ser	UAC } Tyr	UGC } Cys	C	
	UUA } Leu	UCA } Ser	UAA Stop	UGA Stop	A	
	UUG } Leu	UCG } Ser	UAG Stop	UGG Trp	G	
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U	
	CUC } Leu	CCC } Pro	CAC } His	CGC } Arg	C	
	CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg	A	
	CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg	G	
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U	
	AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser	C	
	AUA } Ile	ACA } Thr	AAA } Lys	AGA } Arg	A	
	AUG Met	ACG } Thr	AAG } Lys	AGG } Arg	G	
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U	
	GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly	C	
	GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly	A	
	GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly	G	

- 4) In addition to the gene coded on the *top* strand of the DNA molecule above, suppose a second gene was coded on the *bottom* strand of the DNA molecule. Would the second gene read in the same direction as the first or in the opposite direction?

## Task 2: Evolutionary relationships: Homologs, paralogs, orthologs, and functional similarity

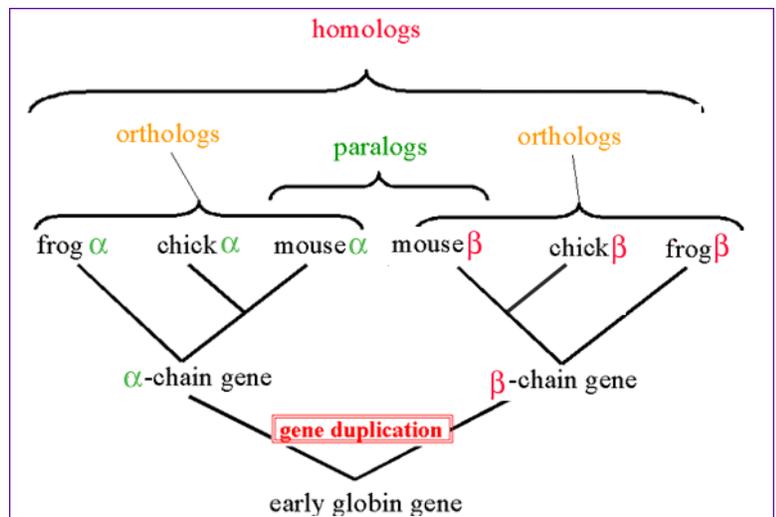
Proteins are linear chains of amino acid residues that fold into complex 3D structures that carry out cellular functions. Proteins that have similar linear sequences of amino acid residues often fold into similar 3D shapes and have similar functions.

In class, we discussed the term “homology” in the context of comparative genomic analysis. Since homology plays an important role in many bioinformatics analyses, let’s explore what it means for genes to be homologous.

**Genes are said to be homologs if they are derived from a common ancestral gene.** Because of this common ancestry, homologous genes often encode structurally similar proteins. Two genes are either homologous or they are not; homology is a Boolean property.

**Homologous genes in different species are called orthologs.** Because they share a common ancestry, orthologs often, but not always, have conserved functions. Functional conservation is a consequence of orthology, not part of its definition.

**Homologous genes in the same species that arose via gene duplication are called paralogs.** Paralogous genes usually diverge functionally, or only one copy of the gene retains function.



- 1) Orthologous proteins don't always have completely overlapping functions. In fact, orthologs may no longer share the same functions. Why is this?
  
- 2) Generally speaking, would you expect the nucleotide sequences of orthologous genes to be more or less similar than the amino acid sequences of the orthologous proteins they encode? Why?
  
- 3) When trying to determine whether two genes are orthologous, one must consider the possibility that two different genes are similar because, over time, their sequences converged towards one another instead of sharing similarity because they diverged from a common ancestral gene. If two genes have evolved convergently, would you expect them to be more or less similar to each other than their ancestral sequences are to one another? What if the two genes have evolved divergently?
  
- 4) Paralogous genes (homologs in the same species that arose as the result of a gene duplication event) often diverge functionally. It is common for only one copy to remain active, while the other copy becomes inactive (called a pseudogene). Why might paralogs diverge functionally?

**Task 3: Online genomic resources: the Yeast Genome Database ([www.yeastgenome.org](http://www.yeastgenome.org)).**

Please fill in the table below as you work your way through the questions in tasks 3 and 4.

Organism	Total # of chromosomes in genome	Gene name	Chromosome for gene	Strand	Coordinates of gene in chromosome	Number of amino acids in protein	Upstream gene and its strand	Downstream gene and its strand
Yeast		<i>HXK1</i>						
Fruit fly		<i>Hex-A</i>						
Human						X	X	X

- 1) Most genes code for proteins and most proteins are enzymes, i.e., they help carry out reactions in the cell. One enzyme is hexokinase. Hexokinase participates in the metabolism of glucose by setting glucose up for glycolysis. Glycolysis is the process of using glucose to make ATP, which stores the energy used by cells. In yeast (*Saccharomyces cerevisiae*), one of the hexokinase enzymes is named *HXK1* or YFR053C. Find and explore the web page for this hexokinase protein at the yeast genome database: <http://www.yeastgenome.org/>
- 2) In the table above, indicate the chromosome in the yeast genome that *HXK1* resides in.
- 3) In the table, indicate the strand, Watson (+) or Crick (-), that *HXK1* resides on.
- 4) In the table, indicate the chromosomal coordinates of the gene.
- 5) In the table, indicate the number of amino acids in the protein.
- 6) In the table, indicate the upstream gene (left) and downstream gene (right) of *HXK1* in the yeast genome and indicate the strand of the upstream and downstream genes (trying clicking on JBrowse and ignore dubious open reading frames, putative proteins, and non-protein-coding genes).
- 7) In the table, indicate the total number of chromosomes in the yeast genome (you may have to find this information on a different web site).

#### **Task 4: Comparative genomics: yeast, fruit fly, and human orthologs**

Access the fruit fly genome database: <http://flybase.org/>

Search for the fruit fly ortholog, *Hex-A*, of the yeast hexokinase gene in the *Drosophila melanogaster* genome database.

- 1) In the table above, fill in the chromosome for the *Hex-A* gene in the fruit fly genome, the strand, the coordinates of the gene, and the number of amino acids in the protein (explore the Gene Model and Products field and look for polypeptide Hex-A-PA in the sub-menu).
- 2) Closely related organisms often have a similar linear arrangement of genes on their chromosomes. Regions of the genome that are highly similar between two organisms are commonly said to be **syntenic** or **share synteny**. In the table above, indicate the upstream gene (left) and downstream gene (right) of *Hex-A* in the fruit fly genome and indicate the strand of the upstream and downstream genes (ignore dubious open reading frames and non-protein-coding genes). Are the two flanking genes similar to the genes that surround the yeast ortholog? What can you conclude about the degree of synteny between the yeast genome and fly genome immediately adjacent to these hexokinase gene orthologs? Is this a surprise? Why or why not?
- 8) In the table, indicate the total number of chromosomes in the *D. melanogaster* genome (you may have to find this information on a different web site).
- 9) You may have noticed in the schematics (when exploring GBrowse or JBrowse) that some of the fruit fly genes have multiple blocks interrupted by thin lines. This indicates that some sequences in the initial RNA transcript are removed (spliced out) before the protein is translated. The pieces of RNA that are removed from the original transcript are called **introns**, while the pieces of the original transcript that remain in the final mRNA are called **exons**. The number of introns and exons in genes is generally correlated with genome size and complexity. Humans have many introns in their genes, fruit flies often have few introns in their genes, yeast rarely have introns in their genes, and bacteria do not have introns in their genes. How many exons does the yeast gene *EFB1* have? How many exons does the fruit fly gene *Karl* have?
- 10) On the page for the *Hex-A* gene in the fruit fly genome, explore the Orthologs field and find the orthologous gene, glucokinase, in humans. In the table, indicate for the human glucokinase ortholog the chromosome number, the strand, and the coordinates of the gene on the chromosome.
- 11) In the table, indicate the total number of chromosomes in the human genome (you may have to find this information on a different web site).

12) Is it surprising that a hexokinase gene in the yeast genome is conserved in both the fruit fly genome and the human genome? Why or why not?