CSE 111 Bio: Program Design I
Lecture 5: Genbank, Lists, More Slicing

Tanya Berger-Wolf (CS) & Boris Igić (Bio)
University of Illinois, Chicago
September 6, 2016

Survey!

CS 1 + Bio:
https://uic.qualtrics.com/SE/?SID=SV_6R7FQcseBNY87Mp

Diversifying CS with a Biology-themed Introductory CS Course at a Large, Diverse Public University

"As members of this class, you are asked to participate in a survey to better understand student perceptions of computer science. This survey is being given by Robert Sloan, a Computer Science Professor at UIC. The survey is accessible via https://uic.qualtrics.com/SE/?SID=SV_6R7FQcseBNY87Mp. The survey will ask you about your experiences and perceptions of CS 111 and computer science in general, and should take between 10-20 minutes. Taking the survey is purely optional, and will not affect your grade in this course in any way. You may opt to stop taking the survey at any time. If you have questions about this survey, please feel free to contact Dr. Robert Sloan at sloan@uic.edu."

Alternate Problem If Not Taking the Survey

Find the *reverse complement (opposite strand sequence) for this DNA string:*

- 5' – CTCAGGAGTCAGGTGCACCAT – 3'

What is the transcribed mRNA string for this (reverse complemented) sequence?
What is the translated protein sequence?

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine	TCA		TAA	Stop (*Ochre*)	TGA	Stop (*Opal*)	A
	TTG		TCG		TAG	Stop (*Amber*)	TGG	(Trp/W) Tryptophan	G
C	CTT		CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

Announcements

- Assignment 2 is out on BB, due tomorrow. Corrected wording
- Prof. Berger-Wolf's office hours **TODAY ONLY** are 3:30-4:30
- Prof. Igić's office hours location will change. Watch this space.

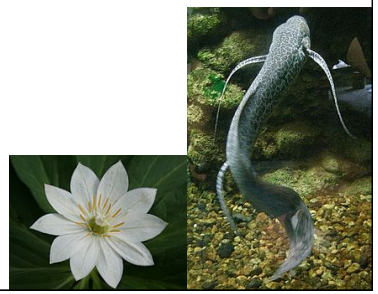
Introduction to GenBank

- *GenBank is a public database of nucleotide sequences and supporting information*

- *It is hosted at the National Center for Biotechnology Information (NCBI), associated with the US National Institutes of Health (NIH)*

Introduction to GenBank

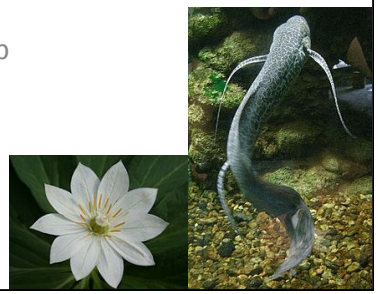
http://www.ncbi.nlm.nih.gov/ or
http://www.ncbi.nlm.nih.gov/genbank



Introduction to GenBank

Protopterus aethiopicus, 130Gbp

Paris japonica, 150Gbp



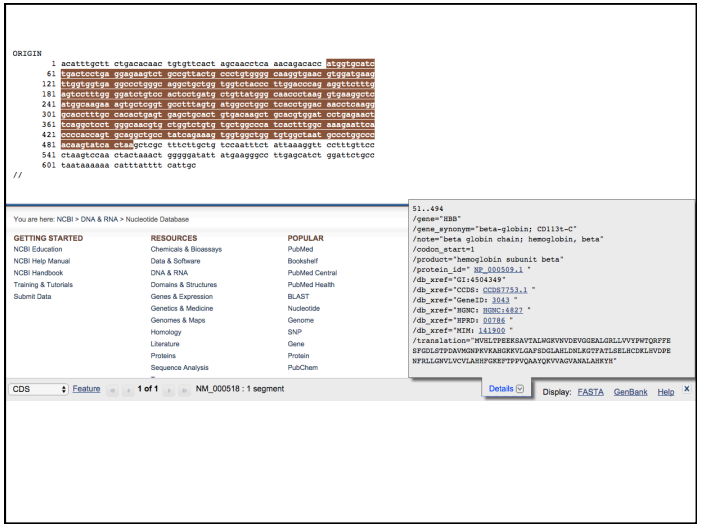
Introduction to GenBank

ß-globin or ***hemoglobin beta chain (HBB)***: a crucial protein involved in binding oxygen and providing it to all of your cells.

https://en.wikipedia.org/wiki/HBB

http://www.ncbi.nlm.nih.gov/

https://www.ncbi.nlm.nih.gov/nucleotide/49456780



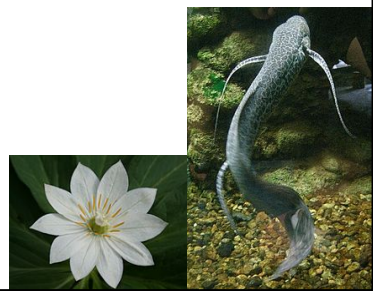
Introduction to GenBank

- *GenBank is a public database of nucleotide sequences and supporting information*

- *It is hosted at the National Center for Biotechnology Information (NCBI), associated with the US National Institutes of Health (NIH)*

Introduction to GenBank

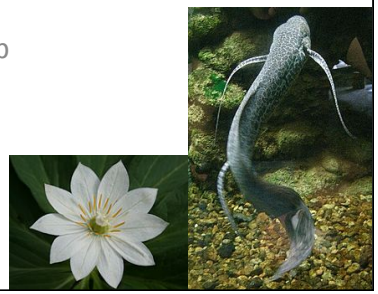
http://www.ncbi.nlm.nih.gov/ or
http://www.ncbi.nlm.nih.gov/genbank



Introduction to GenBank

Protopterus aethiopicus, 130Gbp

Paris japonica, 150Gbp



Lists

```
>>> primes = [2,3,5,7,11]

>>> biologists = ["james","francis","rosalind"]

>>> L = [2, "zebra", 11]

>>> M = [2, "zebra", 11, ["spam","spamity","spam"] ]
```

Lists index/slice the same as strings

```
          0      1        2          3
>>> M = [2, "zebra", 11, ["spam","spamity","spam"] ]

>>> len(M)
4

>>> M[2]
11

>>> M[3]
['spam', 'spamity', 'spam']

>>> M[3][0]  A. ["spam","spamity","spam"], B. [`spam']
             C. `spam'   D. Error    E. No clue

>>> M[2:]    A. 11 B.[11] D.[11, ["spam","spamity","spam"]]
             E. [11, "spam","spamity","spam"] F. No clue
```

The `range` function

```
>>> range(0,25)
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24]

>>> range(25)
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19, 20, 21, 22, 23, 24]

>>> range(3,9)
[3, 4, 5, 6, 7, 8]

>>> range(3,15,3)
[3, 6, 9, 12]
```

Addition in lists

```
>>> myList = [42, 47, 23]
>>> newList = myList + 100
BARF!
>>> newList = myList + [100]
>>> newList
[42, 47, 23, 100]
>>> myList
[42, 47, 23]
>>> newList = newList + newList
>>> newList
```

A. [42, 47, 23, 100, 42, 47, 23, 100]
B. [42, 47, 23, 100]
C. "42, 47, 23, 100"
D. No clue

Mutability

```
>>> myList = [42, 47, 23]
>>> myList = myList + [3]
>>> myList
[42, 47, 23, 3]
>>> myList.append(15)
>>> myList
[42, 47, 23, 3, 15]
>>> x = myList
>>> x.append(234)
>>> x
[42, 47, 23, 3, 15, 234]
>>> myList
```

A. [42, 47, 23, 3, 15] **B. x**
C. [42, 47, 23, 3, 15, 234] **D. No clue**

Types of data in Python

```
>>> goodNum = 42
```
← Integer

```
>>> pi = 3.1415926
```
← "Floating point" number

```
>>> special = [2.718, 3.141, 42]
```
← List

```
>>> okFood = "spam"
>>> greatFood = 'chocolate'
```
← String (single or double quotes work)

```
>>> 5 > 6
>>> True
```
← Boolean