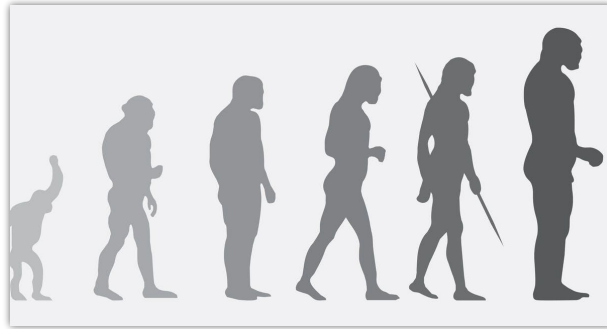


# Codon Usage Bias for Evolution (C.U.B.E)

02-601 Programming for Scientists



# Degeneracy in the Genetic Code leads to CUB

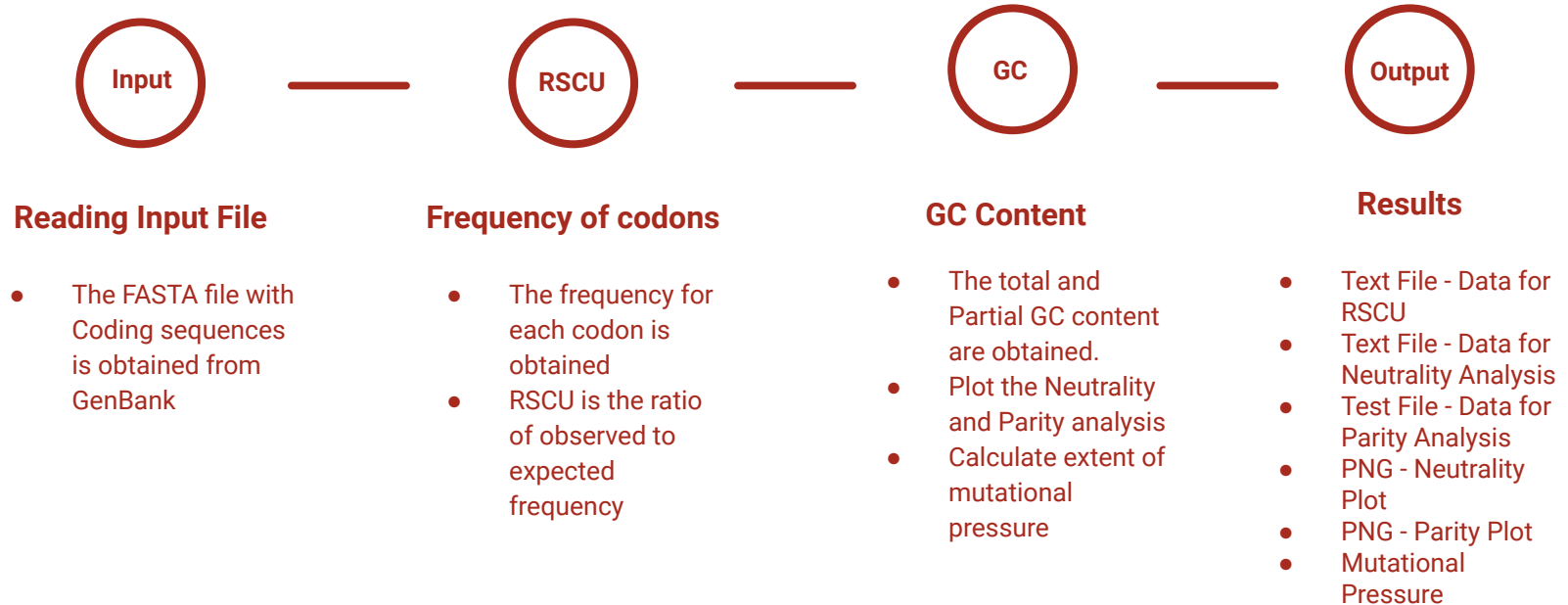
		Seond letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C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

# Statistical Analysis to determine Codon Usage Bias

1. Relative Synonymous Codon Usage
2. GC content
3. Neutrality plot
4. Parity plot

Synonymous Codons ( $X_{ij}$ )						$\sum X_{ij}$	Degeneracy ( $n_i$ )	$RSCU_{ij} = \frac{n_i \cdot X_{ij}}{\sum X_{ij}}$					
GAT	GAC					1	2	0.0	2.0				
0	1												
TCT	TCG	TCA	TCC	AGC	AGT	1	6	0.0	0.0	0.0	6.0	0.0	0.0
0	0	0	1	0	0								
AAC	AAT					2	2	2.0	0.0				
2	0												
CCT	CCG	CCA	CCC			1	4	0.0	0.0	4.0	0.0		
0	0	1	0										
AAG	AAA					3	3	0.0	2.0				
0	3												
GCA	GCC	GCG	GCT			6	4	2.0	0.0	0.0	2.0		
3	0	0	3										
GGA	GGC	GGG	GGT			2	4	0.0	4.0	0.0	0.0		
0	2	0	0										
TTA	TTG	CTA	CTG	CTT	CTC	5	6	0.0	0.0	0.0	4.8	0.0	1.2
0	0	0	4	0	1								
GTA	GTC	GTG	GTT			5	4	1.6	0.8	0.0	1.6		
2	1	0	2										
GAG	GAA					1	2	0.0	2.0				
0	1												
TAT	TAC					1	2	0.0	2.0				
0	1												

# Code Workflow



# Key Results

## Input Text File



test1.txt

```
>AJ007765.1 Naja naja atra mRNA for snake venom
AGATGAAACTCTGCTGCTGACCTTGGTGGTGGTGACAATTGTGCGCTGGACTTAGGATACACCTTAAA
ATGTGCACACACACAGCTTCTTTTCATCTATAAGACTTGTCCAGAAGGGAAGAACTTATGCTTTAAAGCG
ACTTTGAGGAAGTTTCCATTGAAATTTCTGTCAAAGGGGATGTGCTGATAATTGCCCTAAAAACAGTG
CCCTATTGAAGTATGTGTGTTGCAGCAGACAGACAAATGCAACTGATAGCTCTACGAGTGGCTAAATTCGCT
GAG

>AJ007766.1 Bungarus multicinctus mRNA for snake venom, clone pGEMT-BGTV31
AGATGAAACTCTGCTGCTGCTTGGTGGTGGTGACAATCGTGTGCTGGACTTAGGATATACCATCGT
ATGCCACACACAGCTACTTCGCCTATTAGCGCTGTGACTTGTCCACCTGGGAGAACCTATGCTATAGA
AAGATGTGGTGTGATGATTCTGTTCCAGCAGAGGAAAGGTAGTGAATTGGGGTGTGCTGCTACTTGCC
CTTCAAAGAAGCCCTATGAGGAAGTTACCTGTTGCTCAACAGACAAGTGCAACCCACATCCGAAACAGAG
ACCTGGTTGAGTTTGTCTCTCTCCATCATGGACCATCC

>AJ007764.1 Bungarus multicinctus mRNA for snake venom, clone pGEMT-BMNTL4
AGATGAAACTCTGCTGCTGACCTTGGCGGTGGTGACAATGGTGTGCATGGACTTAGGATACACCACAAT
ATGTTACAAACCACTGTCAAGGACTCCTGAAACCACTGAAATTTGTCCAGATTCATGGTACTTTTGCTAT
AAAATCTCTTTGGCTGATGGAATGATGTAAGAATCAAAAGGGGATGTACTTTTACTTGCCTGAATTGA
GACCCACTGGGAAATATGTTTACTGTTGCAGAAGAGACAAATGTAACCAATAGCTCTACGGGTGGCTAAA
TTCCTTGAATTTTGCTCTCTCAATCATGGACCATCCA
```

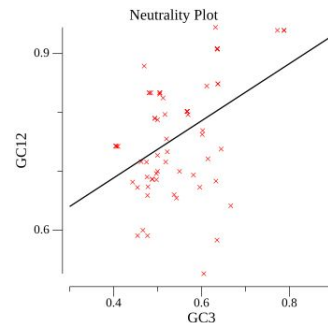
## Output Text File

```
RSCU data for accession ID : lcl|AB984614.1_cds_BAP39945.1_1
TCA : 1.764706
CCT : 1.560976
GGG : 0.640000
CAC : 0.526316
CGG : 0.000000
GAG : 0.677419
ATG : 1.000000
AAG : 0.733333
GAT : 1.457627
CCG : 0.195122
GCC : 1.049180
CTA : 0.896552
CTG : 1.241379
TAC : 0.652174
CGA : 0.731707
CAG : 0.971429
TCT : 1.235294
AGT : 1.500000
GCA : 1.442623
CCA : 1.463415
```

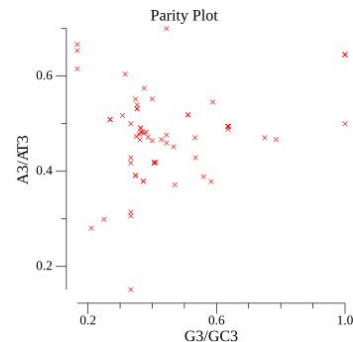
## Mutational Pressure

```
kghatole@98ketaki14 Project % ./Project test3.txt
The effect of mutational pressure is 22.17379544235057
```

## Neutrality Plot



## Parity Plot



# Applications

1. Translation dynamics
2. Gene expression levels
3. Codon Optimization

# Future Directions

- There are several other statistical tools to understand the codon usage bias like:
  1. Effective Number of Nucleotides
  2. Codon Adaptation Index
- User Interface
- Flexibility with graphs