

 INSTITUTO FEDERAL Brasília	Instituto Federal de Brasília Campus Taguatinga Superior em Computação	Nota
	Atividade Prática 2 Expressões Regulares em Bioinformática 1/2019	Data

Nome: _____

- 1) O arquivo [toxinsNCBI.fna](#) contém 22 sequências genômicas de mRNAs que codificam toxinas em bactérias. Essas sequências residem na fita codificadora (5' → 3').
 - a) Crie um *script* python (ou a linguagem que preferir), que receba o arquivo [toxinsNCBI.fna](#) e retorne um arquivo [toxins_3-5.fna](#) que contém a sequência da fita molde 3' → 5' (lembre-se que a sequência deve ser salva na direção 5' → 3').
 - b) Agora crie um *script* que receba [toxins_3-5.fna](#) e transforme as sequências genômicas em sequências de mRNA (na direção 5'-3') que devem ser salvas no arquivo [mRNA_toxins.fna](#)
 - c) Por fim crie um script que receba o arquivo [mRNA_toxins.fna](#) e transforme suas sequências de mRNA em sequências de aminoácidos, salvando estas no arquivo [toxins.faa](#).

Dica: para transformar as sequências de mRNA em sequências de RNA, é necessário consultar o código genético:

Standard genetic code									
1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA		UCA		UAA	Stop (Ochre) ^[B]	UGA	Stop (Opal) ^[B]	A
	UUG		UCG		UAG	Stop (Amber) ^[B]	UGG	(Trp/W) Tryptophan	G
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG ^[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Consulte também o site <https://www.bioinformatics.org/sms/iupac.html> para confirmar o símbolo canônico para cada aminoácido.