

Noțiuni de bază legate de obținerea proteinelor pornind de la ADN

Dogma centrală a geneticii spune că informația genetică merge unidirecțional:



Prima etapă presupune replicarea, care permite celulelor să genereze material genetic nou, utilizând ADN-ul original drept șablon. A doua etapă este transcripția, în care celulele creează ARN-ul mesager bazându-se pe secvența de ADN ca șablon. Urmează faza de splicing, în care sunt eliminate regiunile non-codante, numite introni. ARN-ul mesager matur, rezultat în urma acestui proces va fi utilizat drept șablon pentru legarea aminoacizilor care vor forma proteina, în urma procesului de translație.

Informații legate de program-ul creat și utilizarea lui

Inițial utilizatorul este rugat să introducă secvențele de ADN pe care vrea să le analizeze. Acestea sunt văzute ca string-uri formate combinând literele A, C, T, G, care sunt abrevieri pentru bazele azotate care formează secvența de ADN. Dacă introduce greșit de 3 ori secvența de ADN, se va ieși din program.

Odată ce cele două secvențe de ADN au fost introduse conform cerințelor (fără virgule sau spații între litere), utilizatorului i se permite accesarea meniului. Acesta cuprinde următoarele opțiuni:

1. Determinarea frecvențelor de apariție ale bazelor azotate
2. Transcripția secvențelor de ADN în secvențe de ARN
3. Translația în secvențe de aminoacizi
4. Determinarea frecvenței de apariție a aminoacizilor
5. Calculul distanței Hamming pentru secvențele de ADN
6. Determinarea celui mai lung subșir comun al secvențelor de ADN
7. Calculul distanței de editare a secvențelor de ADN
8. Ieșire din program

1. Determinarea frecvențelor de apariție ale bazelor azotate

Utilizatorul este rugat să introducă literele, care reprezintă abrevierile pentru bazele azotate de interes. Rezultatele vor fi afișate drept un dicționar în care sunt specificate frecvențele de apariție a bazelor azotate în fiecare dintre cele două secvențe de ADN.

2. Transcripția secvențelor de ADN în secvențe de ARN

În urma transcripției se obține ARN-ul mesager. Acesta se obține astfel:

Pe pozițiile din secvența de ADN în care stă adenina (A), în ARNm va sta uracil (U). Acolo unde avem citozină (C), va sta guanina (G). Acolo unde apare guanină (G), în ARN va sta citozină (C). Pe pozițiile în

care în ADN se găsește timină (T), în ARNm va sta adenina (A). Observăm că se respectă complementaritatea bazelor azotate și că secvențele de ARN conțin uracil în loc de timină.

3. Translația în secvențe de aminoacizi

Înainte de translație are loc procesul de splicing, în care sunt eliminate regiunile non-codante din ARN-ul mesager. Acest proces este complex, însă pentru simplificare am presupus că el elimină ultimele baze azotate, dacă este necesar, pentru a obține un număr de baze azotate care este divizibil cu 3.

Apoi urmează înlocuirea codonilor (grupuri formate din 3 baze azotate consecutive) cu aminoacizii care formează proteina, conform următoarei imagini:

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

<https://www.nature.com/scitable/topicpage/nucleic-acids-to-amino-acids-dna-specifies-935/>

4. Determinarea frecvenței de apariție a aminoacizilor

Selectarea acestei opțiuni va genera un grafic în care sunt ilustrate frecvențele de apariție ale aminoacizilor în cele două proteine. De asemenea aceste date vor fi trecute și într-un fișier csv.

5. Calculul distanței Hamming pentru secvențele de ADN

Distanța Hamming reprezintă numărul de poziții în care diferă simbolurile din două string-uri de dimensiuni identice. Pe lângă aceasta, vor fi afișate pozițiile în care diferă bazele azotate ale celor două secvențe de ADN, permițând astfel o vizionare mai rapidă a diferențelor.

6. Determinarea celui mai lung subșir comun al secvențelor de ADN

Utilizând programarea dinamică, se va calcula dimensiunea celui mai lung subșir comun. De asemenea, se va afișa și conținutul acestuia.

7. Calculul distanței de editare a secvențelor de ADN

Utilizând programarea dinamică, se va determina distanța de editare, adică numărul minim de operații de deleție, inserție sau substituție necesare pentru a transforma o secvență de ADN în cealaltă.