

C++프로그래밍및실습

DNA 해독기

프로젝트 제안서

제출일자 : 2023. 10.

제출자명 : 백성준

제출자학번 : 175989

1. 프로젝트 목표

1) 배경 및 필요성

"Chat gpt"와 같은 대화형 인공지능의 출현으로 세계의 기술적 환경은 크게 변화하였다. 평범한 사람들도 이제는 그 변화를 피부로 느끼고 있다. 당장 이러한 대화형 인공지능을 사용해본 사람들은 그 용이성과 위기감을 느낄 수 있다. 머지않아 대부분의 단순 노동은 대체될 것이며, 가장 인간적이라 생각한 부분도 대체될 것이다. 그리고 실제로 지금도 대체가 일어나고 있다.

이러한 변화 속에서 인공지능 또는 자동화 기술을 다양한 학문과 연계한다면 큰 경쟁력을 획득할 수 있을 것이다. 본 프로젝트에서는 생물학에서의 자동화를 목표로 한다.

2) 프로젝트 목표

이 프로젝트에서는 생명의 가장 기초적이지만 근본적인 DNA 염기서열을 자동으로 처리하는 프로그램을 다루고자 한다. DNA 염기서열에서 ORF(Open Reading Frame)을 찾아내고 CDS(Coding sequence)를 최종적으로 얻어내는 것을 목표로 한다. 또한, 이 CDS를 단백질로 변환하는 작업까지 수행한다. 이 때, 사람의 DNA라 가정하며 가장 자주 발견되는 consensus sequence를 따른다고 가정한다. 또한, codon table은 보편적인 table을 사용한다.

3) 차별점

기존의 염기서열 정보를 단백질 서열로 변환하는 프로그램에는 단순 변환만 존재하였다. 여기에서 더 나아가 DNA에서 얻어낸 단백질의 정보가 저장된 ORF(Open Reading Frame)에서 intron을 제

거해낼 수 있다. 또한, 얻어낸 서열 정보의 분석을 진행한다.

4) 배경지식

사람의 유전자는 A,T,G,C의 염기를 가지고 있다. 이 염기가 3 개 모여서 단백질 정보를 이루는 것을 codon이라 하고, codon은 20 개 단백질 정보를 담고 있다. 각 codon에 하나의 단백질이 상응하며 단백질의 시작이 되는 start codon과 stop codon이 각각 존재한다. Start codon에서 stop codon 까지를 ORF(Open Reading Frame)라 한다.

DNA는 RNA로 변환되어 splicing이라는 intron이 제거되는 과정을 거친다. Intron이 제거된 RNA가 ribosome에서 단백질로 변환된다. 이때의 RNA서열에 상응하는 DNA를 CDS(Coding Sequence)라 한다.

2. 기능 계획

1) 사용자의 이름과 염기서열 정보 저장.

사용자의 이름을 입력할 수 있다.

염기서열의 정보를 입력할 수 있다.

2) 염기서열의 정보를 분석

① 염기서열에서 ORF를 저장

-6 가지의 가능한 reading frame을 저장한다.

-각 서열에서 ATG로 시작하고, TGA, TAA 또는 TAG로 끝나는 서열을 저장한다.

② 얻어낸 ORF에서 intron을 제거

- [GTA][AG] 또는 [GTG][AG]로 시작하며 [CAG]로 끝나는

서열을 제거한다. 이 때, 중간에 A가 하나 이상 존재해야 한다.

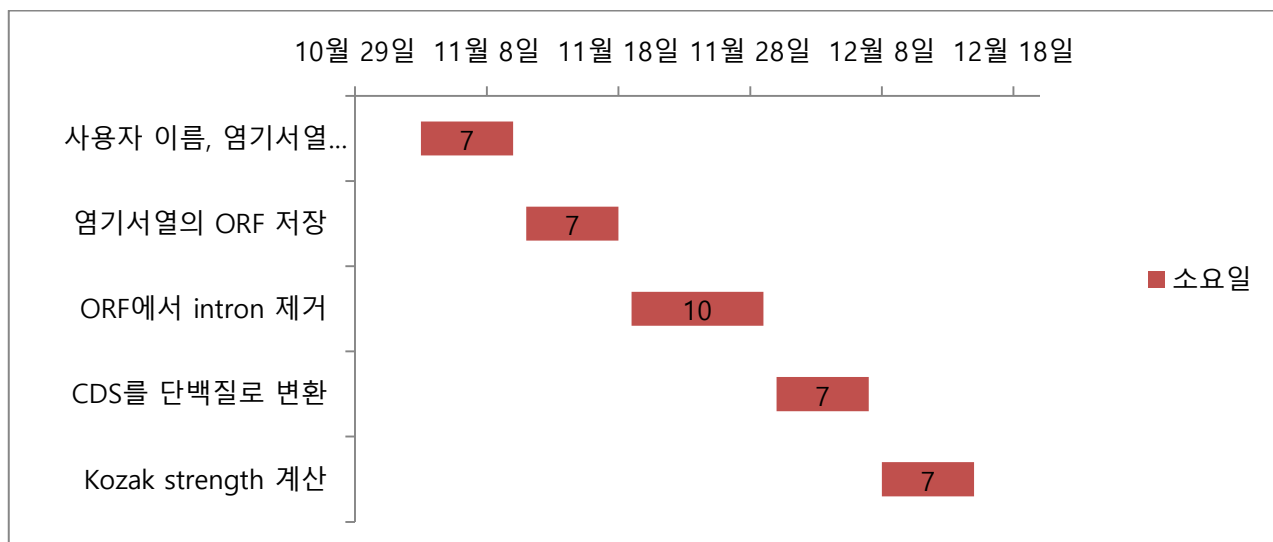
③ Intron이 제거된 CDS를 단백질 형태의 정보로 변환

-codon table에 맞게 염기서열을 단백질 형태로 변환한다.
즉, 각 codon에 맞는 amino acid로 변환한다.

④ Kozak sequence score를 계산

-염기서열에서 +4 위치에서 C 또는 G가 나오면 점수를 1 올리고, -2 위치에 A 또는 G가 존재하면 마찬가지로 1점 올리고, -3 에서 A 또는 G가 존재하면 +3점 증가한다. 이 점수를 합산하여 Kozak strength를 측정하여 출력시킨다.

3. 프로젝트 일정



일정명	시작일	소요일	마감일
사용자 이름, 염기서열 저장	11 월	7	11 월
	3 일		10 일
염기서열의 ORF 저장	11 월	7	11 월

	11 일		18 일
	11 월		11 월
ORF 에서 intron 제거	19 일	10	29 일
	11 월		12 월
CDS 를 단백질로 변환	30 일	7	7 일
	12 월		12 월
Kozak strength 계산	8 일	7	15 일