

Презентация проекта

Библиотека для поиска генетических заболеваний

Подготовил: Рахимов Тимур



Функции программы

- Чтение данных из файлов
- Обработка данных:
 - Исследование последовательностей на наличие изменений генов, вызывающих заболевания
 - Сохранение статистики в отдельный файл
- Скачивание данных из GenBank'a

Что можно найти сейчас?

- Болезнь Гентингтона
 - генетическое заболевание нервной системы
 - вызывается умножением кодона CAG в гене HTT

Classification of the trinucleotide repeat, and resulting disease status, depends on the number of CAG repeats^[25]

Repeat count	Classification	Disease status	Risk to offspring
<26	Normal	Will not be affected	None
27–35	Intermediate	Will not be affected	Elevated but <<50%
36–39	Reduced Penetrance	May or may not be affected	50%
40+	Full Penetrance	Will be affected	50%

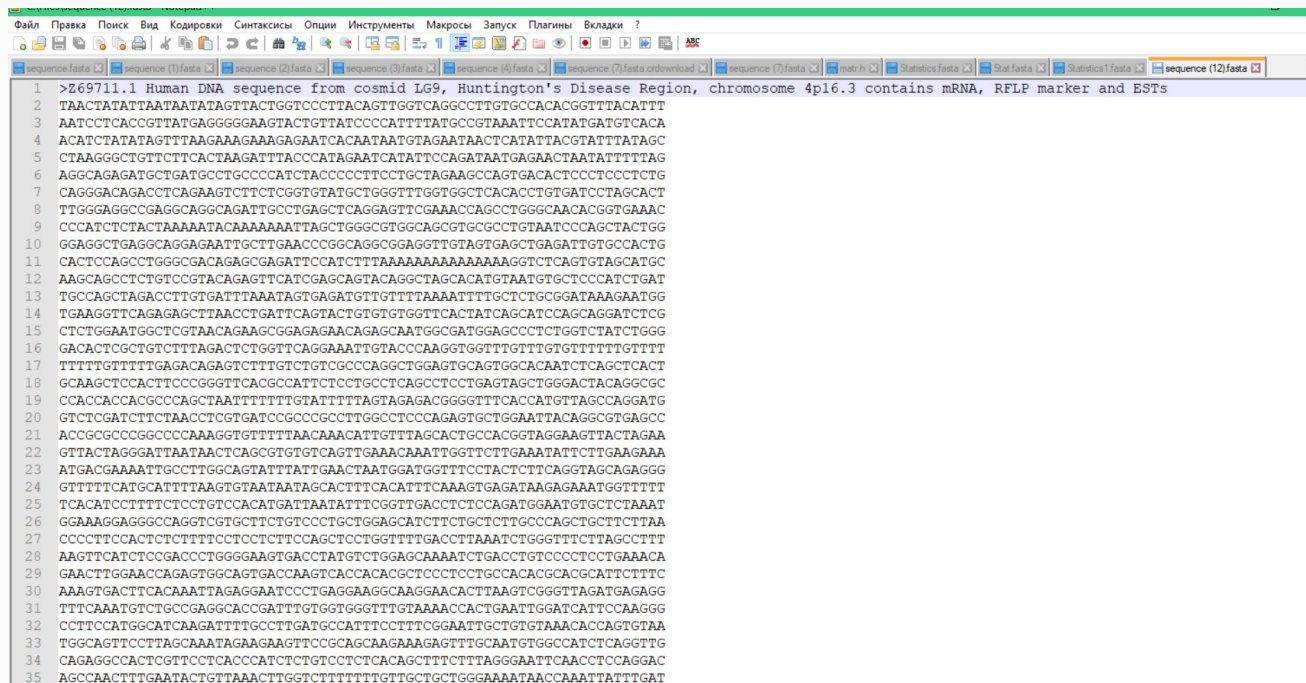
Архитектура проекта

- Проект разделен на два пакета: `fileService` и `diseases`
- `fileService`:
 - `FileDownloader.class` скачивание последовательностей по запросу из GenBank'a
 - `FileReader.class` чтение данных из файлов и директорий
 - `FileSaver.class` сохранение статистики в файл
- `diseases`:
 - `Huntington.class` поиск мутаций гена, характерных для болезни Гентингтона

Пример работы программы

489

последовательностей

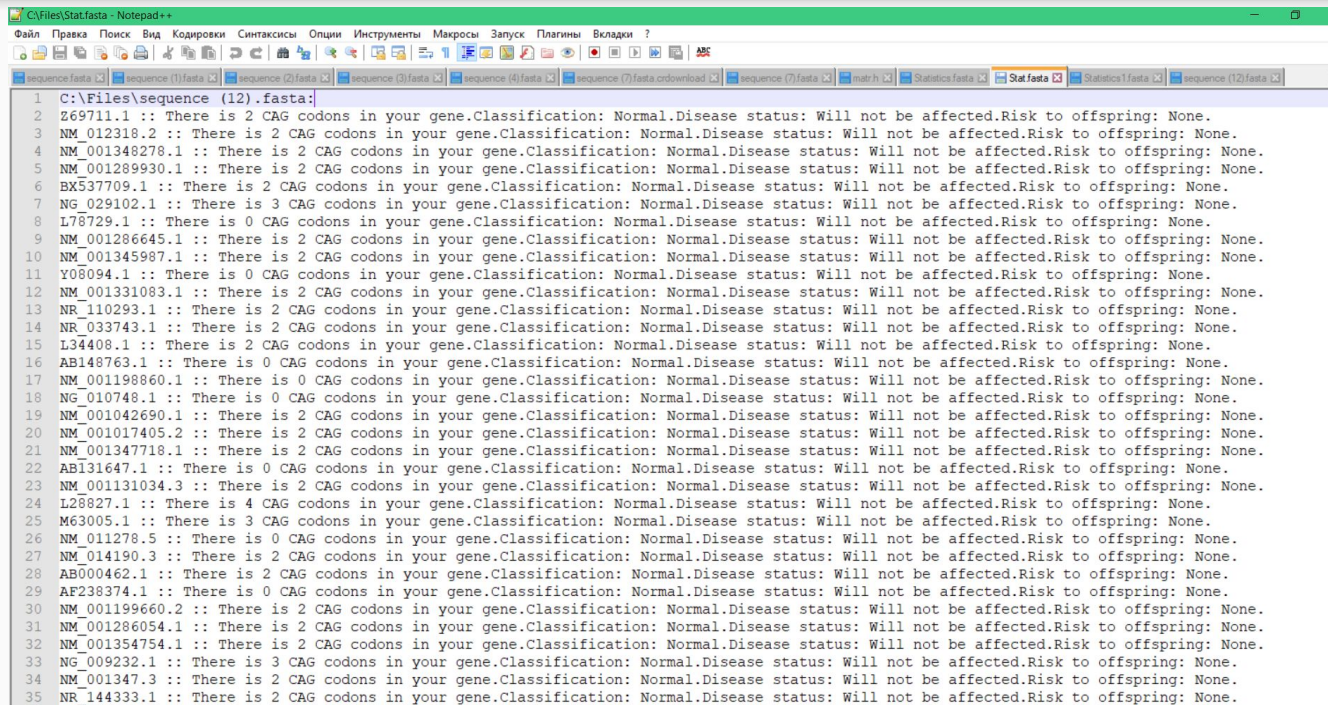


The screenshot shows a text editor window with a menu bar (File, Edit, Search, View, Encoding, Syntax, Options, Instruments, Macros, Run, Plugins, Windows, Help) and a toolbar. The editor contains a FASTA-formatted DNA sequence. The first line is a header: >Z69711.1 Human DNA sequence from cosmid LG9, Huntington's Disease Region, chromosome 4p16.3 contains mRNA, RFLP marker and ESTs. The sequence itself is a single line of uppercase letters (A, T, C, G) wrapped across 35 lines. The window title bar shows several open files, including 'sequence (1).fasta' through 'sequence (12).fasta'.

```
1 >Z69711.1 Human DNA sequence from cosmid LG9, Huntington's Disease Region, chromosome 4p16.3 contains mRNA, RFLP marker and ESTs
2 TAACATATATAATATAGTTACTGGTCCCTTACAGTTGGTCAGGCCCTTGTGCCACACGGTTTACATTT
3 AATCCTCACCGTTATGAGGGGGAAGTACTGTTATCCCCATTTTATGCCGTAAATTCATATGATGTCACA
4 ACATCTATATAGTTTAAAGAAAGAGAGAATCACAATAATGTAGAATAACTCATATTACGTAATTTATAGC
5 CTAAGGGCTGTTCTTCACTAAGATTACCCATAGAATCATATTCAGATAATGAGAATAATATTTTAG
6 AGGCAGAGATGCTGATGCCTGCCCTCATCCCCCTTCTCTGCTAGAAGCCAGTGACACTCCCTCCCTCTG
7 CAGGGACAGACCTCAGAAGTCTTCTGGGTGATGCTGGGTTTGGTGGCTCACACCTGTGATCCTAGCACT
8 TTGGGAGCCGAGGCAGGAGATTCCTGAGCTCAGGAGTTTGAAGCCAGCCTGGGCAACACGGTGAAGC
9 CCCATCTCTACTAAAAATCAAAAAAATTAGCTGGGCTGGCAGCGTGCCTGTAATCCAGCTACTGG
10 GGAGGCTGAGGCAGGAATTTGCTTGAACCCGCGAGCGGAGGTTGTAGTGAGCTGAGATTGTGCCACTG
11 CACTCCAGCCTGGGCGACAGAGCGAGATTCATCTTTAAAAAAGGTTCTCAGTGTAGCATCG
12 AAGCAGCCTCTGTCGTCACAGAGTTTCATCGACAGTACAGGCTAGCACATGTAATGTGCTCCCATCTGAT
13 TGCCAGCTAGACCTTGTGATTTAAATAGTGAGATGTTGTTTTAAAAATTTTGCTCTGGGATAAAGAAATGG
14 TGAAGGTTTCAGAGAGCTTAACCTGATTTCAGTACTGTGTGGTTTCACTATCAGCATCCAGCAGGATCTCG
15 CTCCTGGAATGGCTCGTAAAGAACGGGAGACAGAGCAATGGCGATGGAGCCCTCTGGTCTATCTGGG
16 GACACTCGCTGTCTTAGACTCTGGTTCAGGAATTTGACCCAGGTGGTTTGTGTGTTTTTGTGTTT
17 TTTTGTGTTTGTAGACAGAGTCTTGTCTGTGCCAGGCTGGAGTGCAGTGGCAATCTCAGCTCACT
18 GCAAGTCCACTTCCCGGTTTCAGGCCATTCTCTGCTCAGCCTCTGAGTAGCTGGGACTACAGGCGC
19 CCACACCACGCCAGCTAATTTTTTTTGTATTTTAGTAGAGACGGGTTTACCATTGTAGCCAGGATG
20 GTCTGATCTCTTAACCTCGTGATCGGCCGCTTGGCCTCCAGAGTGTGGAAATACAGGCGTGAGCC
21 ACCGCGCCCGCCCAAGGTTGTTTTAACAAACATTGTTTAGCACTGCCACGGTAGGAAGTTACTAGAA
22 GTTACTAGGATTAATAACTCAGCGTGTGTGAGTGAACAATTTGTTCTTGAATATTCTTGAAGAA
23 ATGACGAAATTCCTTGGCAGTATTATTGAACATAATGGATGTTTCTACTCTTCAGGTAGCAGAGGG
24 GTTTTTCATGCATTTAAGTGAATAATAGCACTTTCACATTTCAAAGTGAGATAAGAGAAATGGTTTTT
25 TCACATCCTTTTCTCTGTCACATGATTAATATTTCCGGTTGACCTCCAGATGGAATGTGCTCTAAAT
26 GGAAGGAGGGCCAGGTCGTGCTTCTGCTCCTGCTGAGCACTTCTGCTCTTCCGACGCTGCTCTTAA
27 CCCCCTCCACTCTCTTTTCTCTCTTCCAGCTCTGCTGTTTACCTTAAATCTGGGTTTCTTAGCCTTT
28 AAGTTCATCTCCGACCTGGGGAAGTGACCTATGTCTGGAGCAAAATCTGACCTGCTCCCTCCTGAACA
29 GAACTTGAACAGAGTGGCAGTGACCAAGTACCAACAGCTCCCTCCTGCCACAGCAGCATTCTTTC
30 AAAGTGACTTCACAAAATAGAGGAATCCCTGAGGAAGCAAGGAACCTTAAGTCGGGTAGATGAGAGG
31 TTTCAAATGTCTGCCAGGCACCGATTGTGGTGGGTTTGTAAACCACTGAATTTGGATCATTCGAAGG
32 CCTTCCATGGCATCAAGATTTTGCCTTGATGCCATTCTTTCGGAATTTGCTGTGAACACCACTGTAA
33 TGCCAGTTCTTAGCAAAATAGAAGAAGTTCCGACAGAAAGAGTTTGAATGTGGCATCTCAGGTTG
34 CAGAGGCCACTCGTTCCTCACCCTCTCTGCTCCTCAGAGCTTTCTTTAGGGAATCAACCTCCAGGAC
35 AGCCAACTTTGAATAGTGTAAACTTGGTCTTTTTTGTGCTGCTGGGAAATACCAAAATTTATTTGAT
```

Пример работы программы

Результат проверки



```
1 C:\Files\sequence (12).fasta:
2 Z69711.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
3 NM_012318.2 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
4 NM_001348278.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
5 NM_001289930.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
6 BX537709.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
7 NG_029102.1 :: There is 3 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
8 L78729.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
9 NM_001286645.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
10 NM_001345987.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
11 Y08094.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
12 NM_001331083.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
13 NR_110293.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
14 NR_033743.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
15 L34408.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
16 AB148763.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
17 NM_001198860.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
18 NG_010748.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
19 NM_001042690.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
20 NM_001017405.2 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
21 NM_001347718.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
22 AB131647.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
23 NM_001131034.3 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
24 L28827.1 :: There is 4 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
25 M63005.1 :: There is 3 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
26 NM_011278.5 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
27 NM_014190.3 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
28 AB000462.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
29 AF238374.1 :: There is 0 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
30 NM_001199660.2 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
31 NM_001286054.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
32 NM_001354754.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
33 NG_009232.1 :: There is 3 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
34 NM_001347.3 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
35 NR_144333.1 :: There is 2 CAG codons in your gene.Classification: Normal.Disease status: Will not be affected.Risk to offspring: None.
```

Перспективы

- Больше болезней
- Использование других баз данных для сбора статистики
- UI