Sequence API Assignment

Overview

Genetic variant classification systems predict the pathological consequences of a particular sequence variant (i.e. change in one, or several, basepairs of a genome) by comparing genomic changes caused by that variant to a reference genome. The goal of this assignment is to build an API which will provide programmatic access to the reference Human genome.

A copy of the reference genome (build GRCh37) is available here: https://s3.amazonaws.com/downloads.solvebio.com/sequence/genbank.GRCh37.fa.gz

Please note that the `.fa` extension denotes FastA file format. FastA is a very common file format for storing genomes; please feel free to use any third-party Python FastA libraries that you can find.

API Design

Your API should return a portion of the Human genome specified by an given genetic coordinate range. Genomic ranges are composed of `chromosome`, `start` and `stop`, where positions are given relative to the first basepair of a chromosome. **NOTE: genomic coordinates start at 1.**

Detailed specifications are given below.

Request

<table>
<tr><th>Method</th><th>Parameters</th></tr>
<tr><td>GET</td><td>(Query Parameters)
<pre>
{
  chromosome: <str>,
  start: <integer>,
  stop: <integer>
}
</pre></td></tr>
<tr><td>POST</td><td><pre>
[
  {
    chromosome: <str>,
    start: <integer>,
    stop: <integer>
  }, ...
]
</pre></td></tr>
</table>

`POST` should allow a user to retrieve multiple sequence ranges in the same request.

Parameter	Python Type	Value
`chromosome`	`str`	1-22, X, Y
`start`	`int`	[1, ...]
`stop`	`int`	[1, ...]

Response

Status Code	Body
`200`	a, or many (if `POST`), genetic sequence(s) (`str` containing `A, T, G, C's)`
`400`	detailed error response

Requirements

Your API should be built on Django and Django REST Framework, it should be REST-ful, should return JSON documents, and should conform to standard conventions and best-practices. Additionally, it should **only** return `HTTP 200` and `HTTP 400` status codes. All other responses should be handled server-side.

Additionally:

- the API should be one-based, fully-closed (i.e. it should support *inclusive* ranges)
- the API should support request ranges of up 500 basepairs; requests for longer ranges should result in an `HTTP 400`
- the API should not support negative ranges (i.e. [10, 8], ranges where start > stop)

Submission and Deliverables:

Please provide a link to your GitHub repository with the API, a README with complete documentation, and corresponding tests.

Bonus: Add support for additional genome builds (versions).