BIOEN 537
Computational System Biology
Final Presentation

Nello Gu
Dec 10th 2023

UNIVERSITY of WASHINGTON

Background

What is the problem

> Current software solutions for nanopore sequencing is not widely available with python.

> No clear guideline for installing the pipelines in python

> A need for a specialized toolkit that streamlines primer design and sequence alignment, specifically tailored for nanopore sequencing with python.

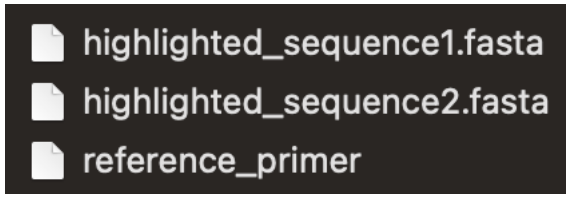
Use Case

> Laboratory Researchers:

- User will install the package with pip
- Quickly design primers for PCR amplification of specific DNA sequences (prior to Nanopore Sequencing).
- Import data from EPI2ME (official nanopore sequencing software) and align nanopore-sequenced fragments against known reference sequences.

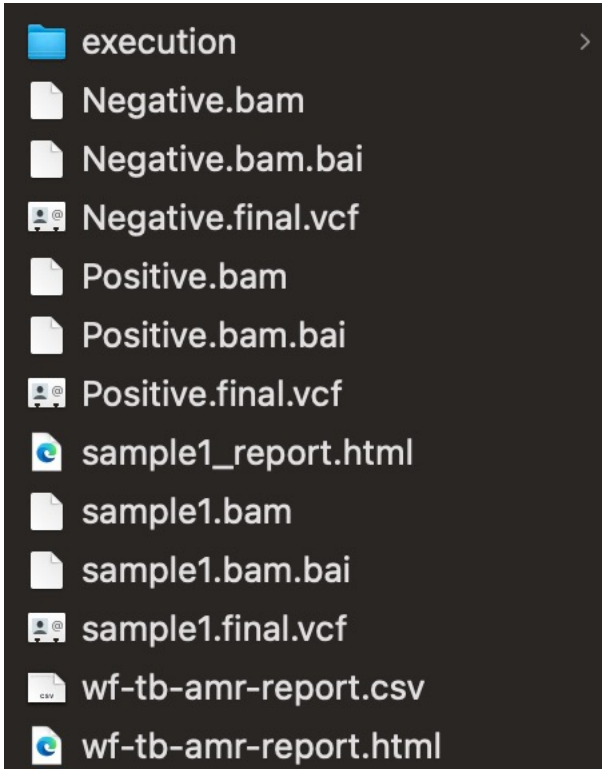
Demo – Primer Design

```
Primer: TTGACCGATGACCCCGGTTC
Sequence with Primer : TTGACCGATGACCCCGGTTCAGGCTTCACCACAGTGTGGAACGCGGTCGTCTCCGAACTTAACGGCGACC
Primer: AACTGGCTACTGGGGCCAAG
Sequence with Primer : AACTGGCTACTGGGGCCAAGTCCGAAGTGGTGTCACACCTTGCGCCAGCAGAGGCTTGAATTGCCGCTGG
```



Demo – Nanopore Sequence Alignment

```
executor >  local (24)
[c6/65f505] process > validate_sample_sheet          [100%] 1 of 1 ✔
[fd/2368f0] process > fastcat (2)                    [100%] 3 of 3 ✔
[e5/6351e5] process > pipeline:getVersions           [100%] 1 of 1 ✔
[da/ec30ec] process > pipeline:getParams             [100%] 1 of 1 ✔
[74/ffa312] process > pipeline:alignReads (3)        [100%] 3 of 3 ✔
[b1/608882] process > pipeline:mpileup (3)           [100%] 3 of 3 ✔
[44/d5f883] process > pipeline:whatshap (1)          [100%] 3 of 3 ✔
[23/ea9ab7] process > pipeline:countReadsRegions (3) [100%] 3 of 3 ✔
[b1/b3b301] process > pipeline:report (1)            [100%] 1 of 1 ✔
[0b/fa9009] process > pipeline:reportSingle (1)      [100%] 1 of 1 ✔
[b2/584306] process > output (4)                     [100%] 4 of 4 ✔
Completed at: 10-Dec-2023 19:18:32
Duration    : 1m 25s
CPU hours   : 0.1
Succeeded   : 24
```

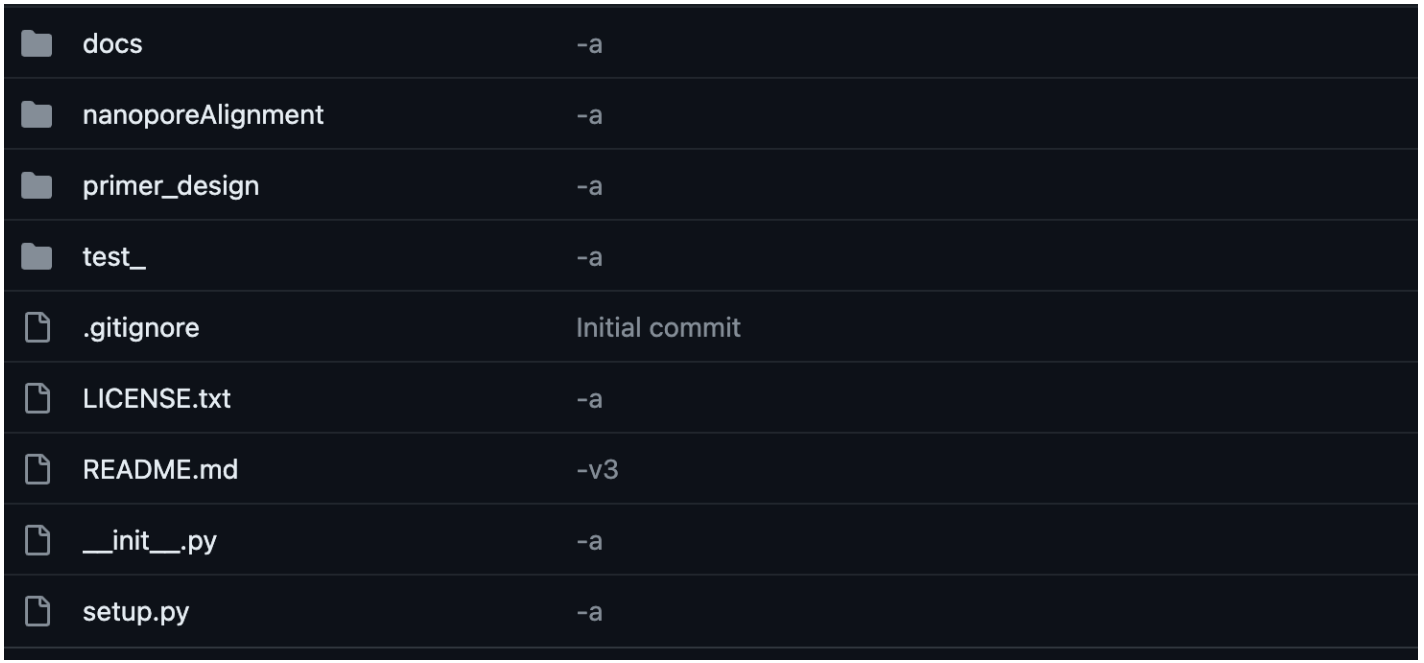


Design

- **Primer Design (primerdesign.py):**
  - Generates primers from input DNA sequences.
- **Sequence Alignment (nanoAlign.py):**
  - Aligns sequencing data and manages dependencies.
- **Interaction:**
  - Modules interact with the user's input data, processing it according to nanopore-specific algorithms.
  - Automated environment setup ensures smooth workflow execution.

Project Structure

- **GitHub Repository:**
  https://github.com/hgu1uw/UW_BIOEN537.git



Lessons Learned and Future Work

- **Modularity:**
  - Importance of designing independent, interchangeable modules.
- **Documentation:**
  - Well-documented code and usage instructions are crucial for user adoption and maintenance.
- **Future Work:**
  - Integration of basecalling and more quality control for nanopore sequencing
  - For primer design, implement automatic reading of all fasta files in the folder.
  - Expanding the toolkit to cover more use cases in nanopore sequencing.