PythoFragam: A Python-based Optimization Tool for DNA Fragment

This repository contains the code for DNA fragment assembly based on a metaheuristic* Overlap Layout Consensus Approach*: Restarting recentering hybrid genetic algorithm (RRHGA). Complete code is written in Python 3.

System requirements

1. **Python 3.4+**
2. **Swalign**
3. **Pandas**
4. **Numpy**

Datasets

http://chac.sis.uia.mx/fragbench/

1. GenFrag Instances
2. DNAgen instances

Main.py-This script contains the main functions for solving the problem of DNA fragment assembly. These are givens below.

- **Importing the dataset-**
  First, import a FASTA format file. The file is open and appends the fragments into a list name is sequences. The second is the CSV file includes the matrix, the value of the cell, is overlapping scores among the sequences.
- **Setting the parameters**
  The parameters such as Population size, Mutation rate, Number of Trans, and Cutoff value, the percentage of Trans is increased or decreased in case of improvement and no improvement.
- **run_2opt**
  This function takes the overlapping score matrix is an input, and find the initial Centre (optimal path-order of the fragments) for the genetic algorithm.

- **D_Rep**
  This function gets the Centre is an input, and generate the populations based on directed transpositions.
- **Genetic_alg**
  Once the populations are generated. The called function **genetic_alg** gets the populations, **num_parents_mating**, **n_generation**, **mutation_rate** is an inputs. The genetic_algo import partially_**cross** function for evolutionary operators such as crossover and mutation. However, the **array_contig** function used PALS as an evolutionary operator to order the fragments while minimizing the number of contigs. The **init_pop_score** function used the local alignment algorithm for the fitness value evaluation.
- **last_fit_score**

  This function gives the final output, by calculating the sum of overlap score, and the number of contigs of the best solution generated from RRHGA.

Contact: In case of any query, feel free to drop an email at uzma@giki.edu.pk

☺