

Multiple Sequence Alignment (MSA) di sequenze SARS-CoV-2

EDOARDO SILVA 816560
DAVIDE MARCHETTI 815990

A.A.: 2019/2020

1 Abstract

Il lavoro svolto è diviso in 2 parti e il codice in 4 sezioni:

1.1 Matrice delle variazioni

Qui si trovano le prime 2 parti del codice: il recupero per intero delle 14 sequenze usate e la lettura dei file di variazione generati in output nella partel

Listing 1: Porzione di ciclo

```
reference_id = load_fasta_id(os.path.join '..', '..', 'project
    ↪ -1', 'input', 'reference.fasta'))
sequence_ids = read_sequence_ids(paths=[
    os.path.join '..', '..', 'project-1', 'input', 'GISAID'),
    os.path.join '..', '..', 'project-1', 'input', 'ncbi'),
])
sequence_ids.insert(0, reference_id) #insert reference no
    ↪ variations

...

clustal_output = load_output('Clustal-NC_045512.2_2020-05-30_16
    ↪ -51.json')
variations = clustal_output['unmatches'].items()
```

Infine nella parte 3 del nostro codice generiamo la tabella delle variazioni **‘table.csv’** per variazione, per poi trasporla.

Listing 2: Porzione di ciclo

```
...  
  
for key, value in variations:  
    row = np.zeros(len(sequence_ids))  
    indexes.append('C{}'.format(counter))  
    for sequence in value['sequences']:  
        row[sequence_ids.index(sequence)] = 1  
    rows.append(row)  
    counter += 1  
  
...  
  
trait_matrix = pd.DataFrame(rows, index=indexes, columns=  
    ↪ sequence_ids, dtype=bool).transpose()  
trait_matrix = phylogeny.reorder_columns(trait_matrix, axis=0,  
    ↪ ascending=False)  
trait_matrix.to_csv(os.path.join '..', 'output', 'table.csv'))
```

1.2 Albero filogenetico

La quarta e ultima parte del codice consiste nel ricostruire l'albero filogenetico prendendo l'output del punto 3 e facendo estensivo uso di funzioni create per questo lavoro nel file **‘phylogeny.py’** per:

1. generare matrice compatibile con filogenesi perfetta.
2. assicurarsi che tutto abbia funzionato e non sia più presente matrice proibita.
3. generare e stampare in output l'albero delle sequenze.

Listing 3: Porzione di ciclo

```
candidate_matrix = get_perfect_phylogeny_character_matrix(  
    ↪ trait_matrix)  
if phylogeny.is_forbidden_matrix(candidate_matrix):  
    raise Exception('Invalid perfect phylogeny matrix')  
phylogeny.build_tree(candidate_matrix)
```

2 conclusione