

Progetto (seconda parte)

Creare una tabella in cui, per ognuna delle posizioni di variazione trovate sulla sequenza di riferimento, che cadono in uno dei suoi geni, viene fornito:

- ✓ il gene id del gene in cui cade la variazione, con lo start e l'end della sua CDS sulla reference sequence
- ✓ il codone (o i codoni) sulla reference che la variazione altera, con posizione di inizio rispetto alla CDS, sequenza del codone e amminoacido codificato
- ✓ per ognuna delle alternative della variazione (in genere sarà una), determinare il nuovo codone generato (o i nuovi codoni generati) specificando la sequenza del codone e il nuovo amminoacido codificato (che potrebbe anche rimanere invariato.

Progetto (seconda parte)

Creare una tabella in cui, per ognuna delle posizioni di variazione trovate sulla sequenza di riferimento, che cadono in uno dei suoi geni, viene fornito:

- ✓ il gene id del gene in cui cade la variazione, con lo start e l'end della sua CDS sulla reference sequence
- ✓ il codone (o i codoni) sulla reference che la variazione altera, con posizione di inizio rispetto alla CDS, sequenza del codone e amminoacido codificato

NB: potrebbe capitare che l'alternativa considerata alteri la lunghezza della CDS in maniera che non sia più un multiplo di 3. In tale caso segnalare semplicemente che i nuovi codoni non sono determinabili

Progetto (seconda parte)

```
/mol type="genomic RNA"
                /isolate="Wuhan-Hu-1"
                /host="Homo sapiens"
                /db_xref="taxon:2697049"
                /country="China"
                /collection date="Dec-2019"
5'UTR
                1..265
                266..21555
gene
                /gene="ORF1ab"
                /locus tag="GU280 gp01"
                /db xref="GeneID:43740578"
CDS
                join(266..13468,13468..21555)
                /gene="ORF1ab"
                /locus tag="GU280 gp01"
                /ribosomal slippage
                /note="pplab; translated by -1 ribosomal frameshift"
                /codon start=1
                /product="ORF1ab polyprotein"
                /protein id="YP 009724389.1"
                /db xref="GeneID:43740578"
                /translation="MESLVPGFNEKTHVOLSLPVLOVRDVLVRGFGDSVEEVLSEARO
                HLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGE
                TLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
                WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
                LDFIDTKRGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
                LNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHCGETSWQTG
                DFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEYHNESG
                LKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVGEGSEGLNDNL
                LEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVESCGN
                FKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAA
                ITILDGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKL
                KPVLDWLEEKFKEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLV
                NKFLALCADSIIIGGAKLKALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEII
                {\tt FLEGETLPTEVLTEEVVLKTGDLQPLEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEK}
```

Progetto (terza parte)

Date le variazioni visualizzate anche in termini di amminoacidi, scegliere un elenco di caratteristiche per le sequenze date e costruire una matrice binaria di caratteri, (1/0=c'è/non c'è il carattere). Costruire la filogenesi perfetta delle sequenze e confrontare l'albero ottenuto con quelli ottenuti in precedenza con i *tools* di allineamento.

Esempio di caratteri:

- avere acquisito una certa variante in un certo periodo di tempo
- abbinare alla variante la posizione geografica

NB: fate attenzione alla scelta, al fine di consentire la costruzione della filogenesi perfetta. Magari analizzate coppie di varianti alla volta verificando che non producano la matrice proibita.