
TP n° 4

Exercice 1 - Ecrire un programme C++ utilisant la STL qui permet, étant donné un ensemble de séquences d'ADN de calculer le score de parcimonie des séquences deux à deux.

On dispose en entrée d'un fichier de séquences au format suivant :

```
>S1
ACCGGTAGcggggGAGCGAGGGATG
ATGACGAGGGTAG
>S2
ACTTATATCTTAgtagATAGTCGATCgatAGCTAAGTA
...
>Sn
AAGGGCGTAGCTGGCTAGATCGATAGCTAGACTggtgg
```

Chaque séquence est identifiée par son nom (S_1, S_2, \dots, S_n). Le caractère > en début de ligne indique que les caractères qui suivent représentent un nom de séquence. Les caractères sur les lignes suivantes qui ne commencent pas par > correspondent au contenu ADN de la séquence.

Attention !

- au final toutes les séquences ont la même longueur
- les caractères ADN apparaissent en minuscule et majuscule

Vous pouvez récupérer un fichier exemple `sequences.txt` à l'adresse suivante :

<http://www.info.univ-angers.fr/pub/richer/ensmli.php>

Pour calculer le score de parcimonie de deux séquences, on utilise la formule suivante :

```
unsigned int parcimonie(Sequence& s1, Sequence &s2) {
    unsigned int score=0;

    for (unsigned int i=0; i<s1.get_size(); ++i) {
        if (s1[i]!=s2[i]) ++score
    }
    return score;
}
```

On a redéfini ici l'opérateur `[]` de manière à pouvoir accéder à chacun des caractères ADN de la séquence.

On stockera les scores de parcimonie dans une matrice et on affichera le résultat comme suit :

	S1	S2	...	Sn
S1	0	25	...	23
S2	25	0	...	12
...				
Sn	23	12	...	0