

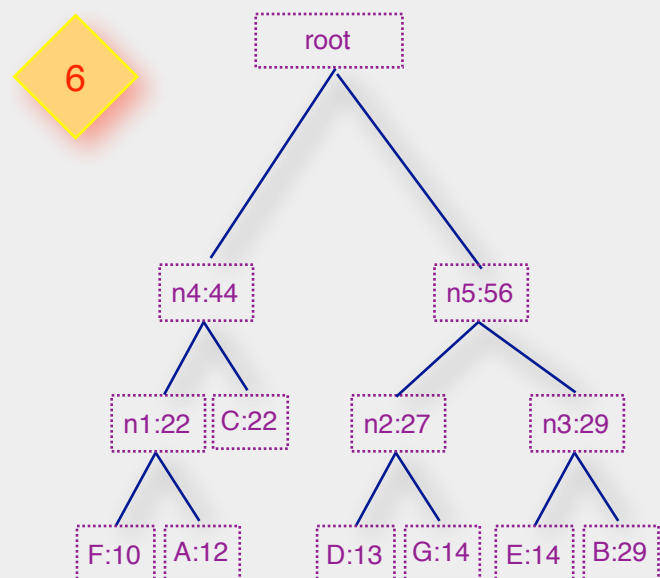
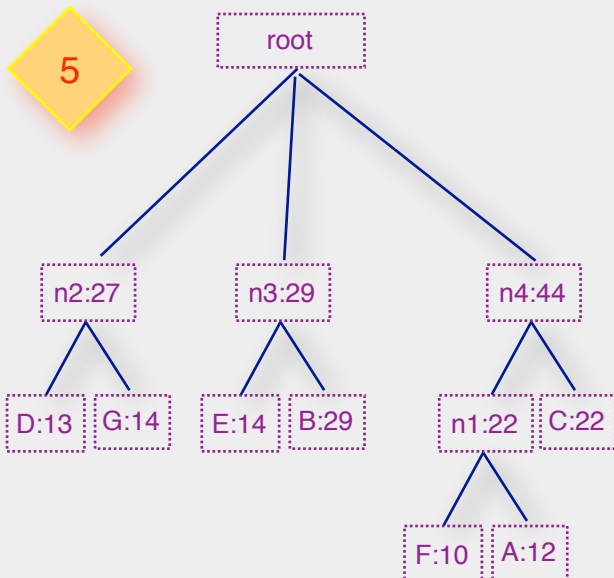
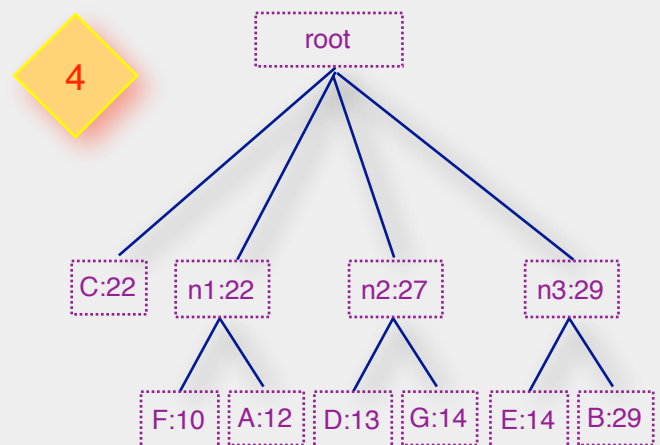
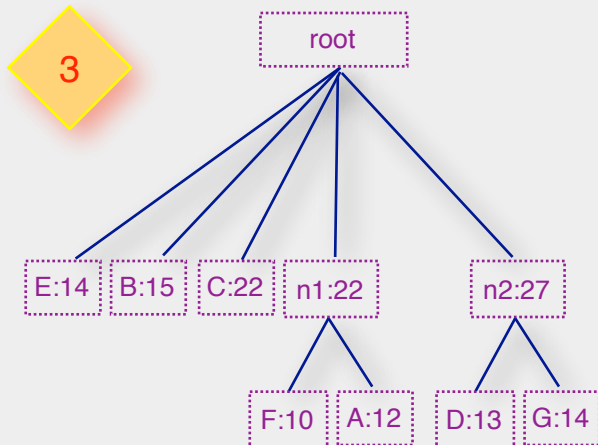
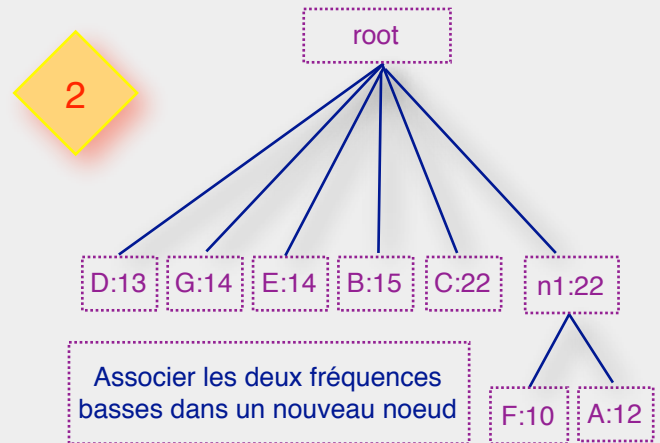
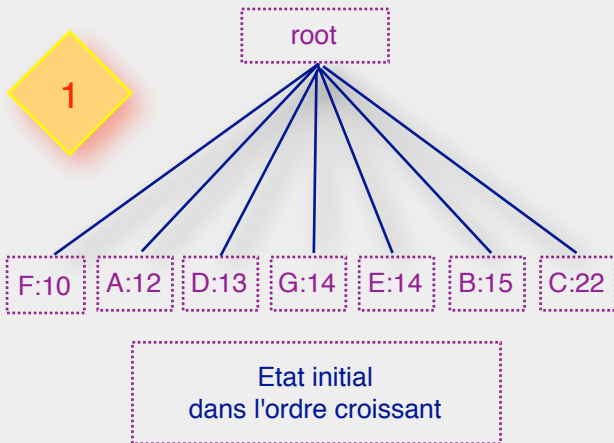
Compression Huffman

Création d'un arbre pour coder les caractères en fonction de leur fréquence.

Considérons un texte, utilisant les caractères A , B , C , D , E , F auxquels on associe une pondération (par exemple le nombre d'apparition dans le texte)

A	B	C	D	E	F	G
12	15	22	13	14	10	14

On réalise une arborescence binaire (deux fils au plus par père) par regroupements successifs .



Cet algorithme permet ainsi de considérer les lettres comme des feuilles de l'arbre, et quand on parcourt l'arbre "à main gauche", on rencontre les lettres dans l'ordre croissant des fréquences d'apparition.

Premier travail

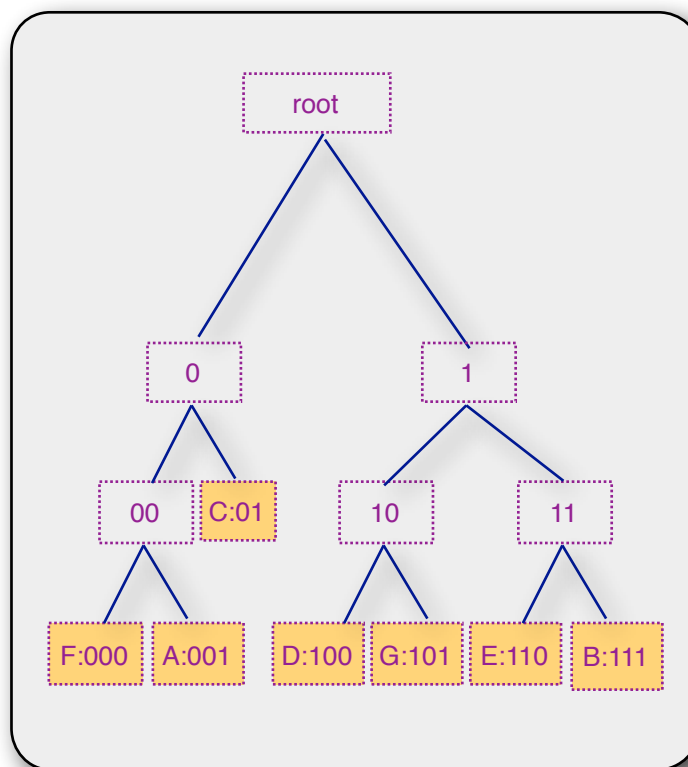
Réaliser l'arborescence à partir des fréquences suivantes, en prenant soin d'obtenir les pondérations dans l'ordre croissant :

A	B	C	D	E	F	G	H	I	J	K	L
25	12	10	7	48	5	8	0	15	3	0	15

Codage de Huffman d'un arbre.

On code les noeuds et les feuilles par des mots binaires. Ces mots codent le chemin à réaliser pour se déplacer de la racine de l'arbre au noeud ou à la feuille de l'arbre.

- ▶ le 0 signifie sous-arbre gauche,
- ▶ le 1 signifie sous-arbre droit.



Traduire la suite binaire :

01001000110100110111001101100001100

Ecrire, "en français",

- un algorithme de réalisation de l'arbre,
- un algorithme de décodage d'une suite d'octets,
- un algorithme de codage d'un texte.

Le type arbre binaire

Les types utiles.

Comment coder un arbre en Python ?

Premiers contacts.

Ecrire un programme qui essaie ces types:

- Initialisation d'un mini-arbre ,
- Présentation sur la console de ce bonsaï .

Profiter de cette première expérimentation pour mettre au point une fonction récursive:

```
def affiche_arbre(arbre, profondeur):
```

La profondeur 0 est attribuée à la racine r, 1 pour les noeuds codés 0 et 1, 2 pour 00, 01, 10, 11 ...

On pourra s'inspirer de la présentation des structures arborescentes des fichiers dans les fenêtres.

Ecrire une fonction de construction d'un arbre

Des fonctions utiles sur les arbres.

Mettre au point chacune des fonctions suivantes, en faisant une expérimentation simple pour chacune.

```
//Une fonction pour détruire proprement un arbre:
```

```
def kill(myTree);
```

```
// Une fonction pour réaliser un arbre père à partir de deux arbres fils
```

```
def joindre(leftTree, rightTree):
```

```
// Code de Huffman pour une feuille ou noeud de l'arbre
```

```
def codeHuffman(aTree, node)
```

```
// Code de Huffman pour une feuille ou noeud de l'arbre repérée par son étiquette
```

```
def codeHuffman(aTree, value)
```

Mise au point de Huffman

Niveau direct.

Il faut maintenant mettre au point les fonctions principales:

- Lecture d'un fichier texte pour la détermination des fréquences d'apparition de chaque lettre,
- Réalisation de l'arbre à partir des fréquences d'apparition,
- Codage d'un caractère,
- Codage d'une chaîne,
- Codage d'un fichier.

Utilisation maximale de Huffman

On veut minimiser l'utilisation des bits en mémoire pour obtenir la compression maximale:

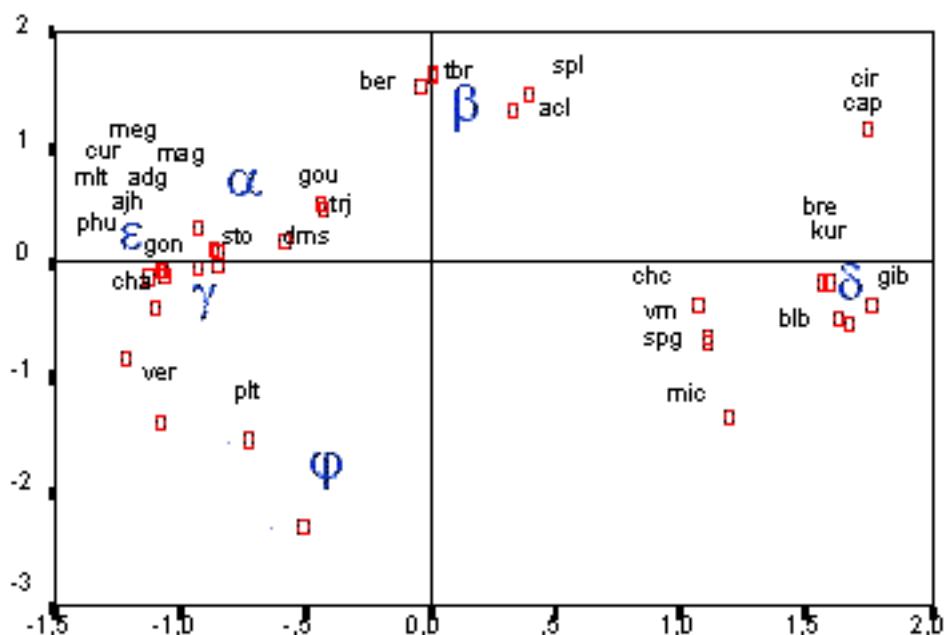
Placer le codage de Huffman dans un fichier temporaire,

Puis découper le fichier temporaire par tranches de 8 caractères que l'on réduit à un seul.

Dendogramme

Document 1: Genetical Distances Between the Main Chondriome Types α , β , γ , δ , ϵ in Potato
<http://www.flg.tum.de/pbpz/mm/mt/dendro.htm>

For subsequent selection of fusion combinations, it was necessary to get an overview about the different phylogenetic branches the main mt types in comparison to the related *Solanum* wild species. For this purpose the fragment patterns which were generated by the Mt DNA probes were used to perform a cluster analysis and to calculate the relationships of α , β , γ , δ and ϵ to *Solanum* wild species. The figure below shows a two dimensional screening, calculated with the model of euclidian distances.



According to dendrogramme (cf page suivante) and 2-dimensional screen the mt- α and mt- γ type were closely related to *S. stoloniferum* and *S. demissum* mt genomes. Mt- β is defined as *S. tuberosum* mt genome. Mt- ϵ is related to *S. andigena*, *goniocalyx*, *phureja*, *curtilobum*, *multidissectum*, *ajanhui* and *chauca* mt genomes.

Pour un nuage de points, on peut réaliser des regroupements en différentes zones selon une règle de proximité (choix d'une mesure des distances).

A chaque étape de la réalisation de l'arbre de Huffman, on a un tel partitionnement.

Le travail que l'on peut faire est le suivant:

Lire dans un fichier texte des données d'un nuage de points,

Visualiser,

Boucle(profondeur de l'arbre de Huffman)

visionner l'état de l'arbre par le nuage

fin Boucle

On pourra utiliser l'enveloppe convexe (convex hull) d'un nuage de points pour visualiser le partitionnement à chaque étape.

