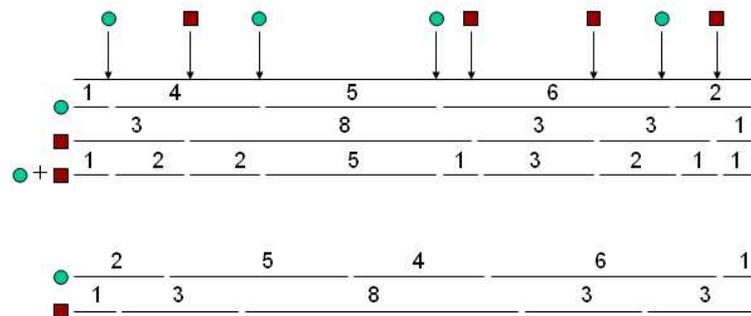


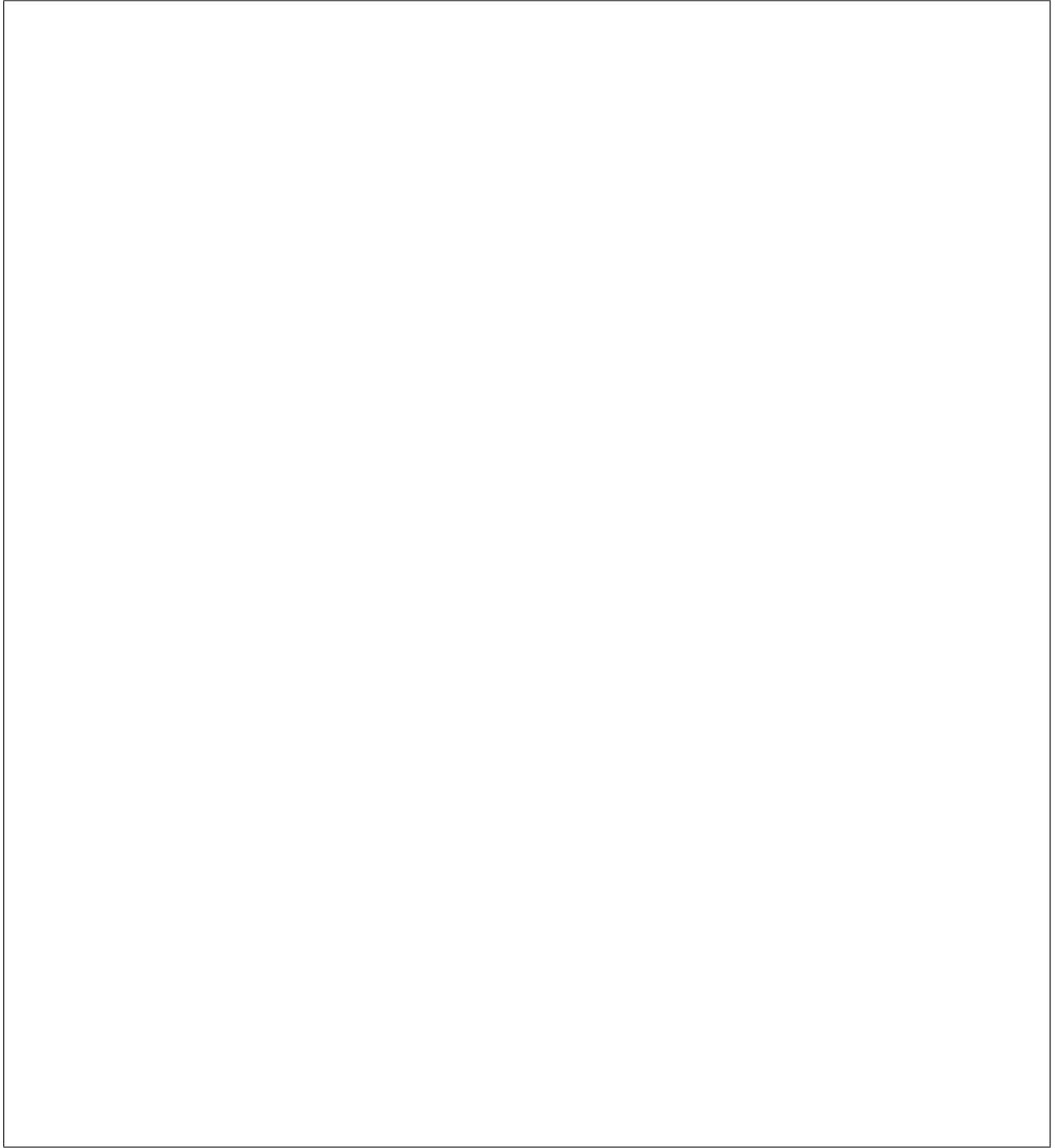
Module AAGB. Examen Mars 2006. Durée 3h.
Tous documents autorisés
Téléphones portables éteints et rangés dans vos sacs

Exercice 1 (10 points) Reconstruction de cartes physiques (5 points)

Le *Double Digest Mapping* est une technique de cartographie qui utilise deux enzymes de restriction. Dans cette approche, l'ADN est digéré de telle façon que seuls des fragments entre sites consécutifs sont formés. Une possibilité pour reconstruire la carte double digest est de mesurer la longueur des fragments (mais pas l'ordre) à partir d'une digestion complète de l'ADN par chacun des deux enzymes seuls, et en suite par les deux appliqués ensemble. Le problème de déterminer les positions des coupures à partir de la longueur des données est appelé "Double Digest Problem" (DDP). La figure suivante montre l'ADN coupé par deux enzymes, A (cercle) and B (carré).



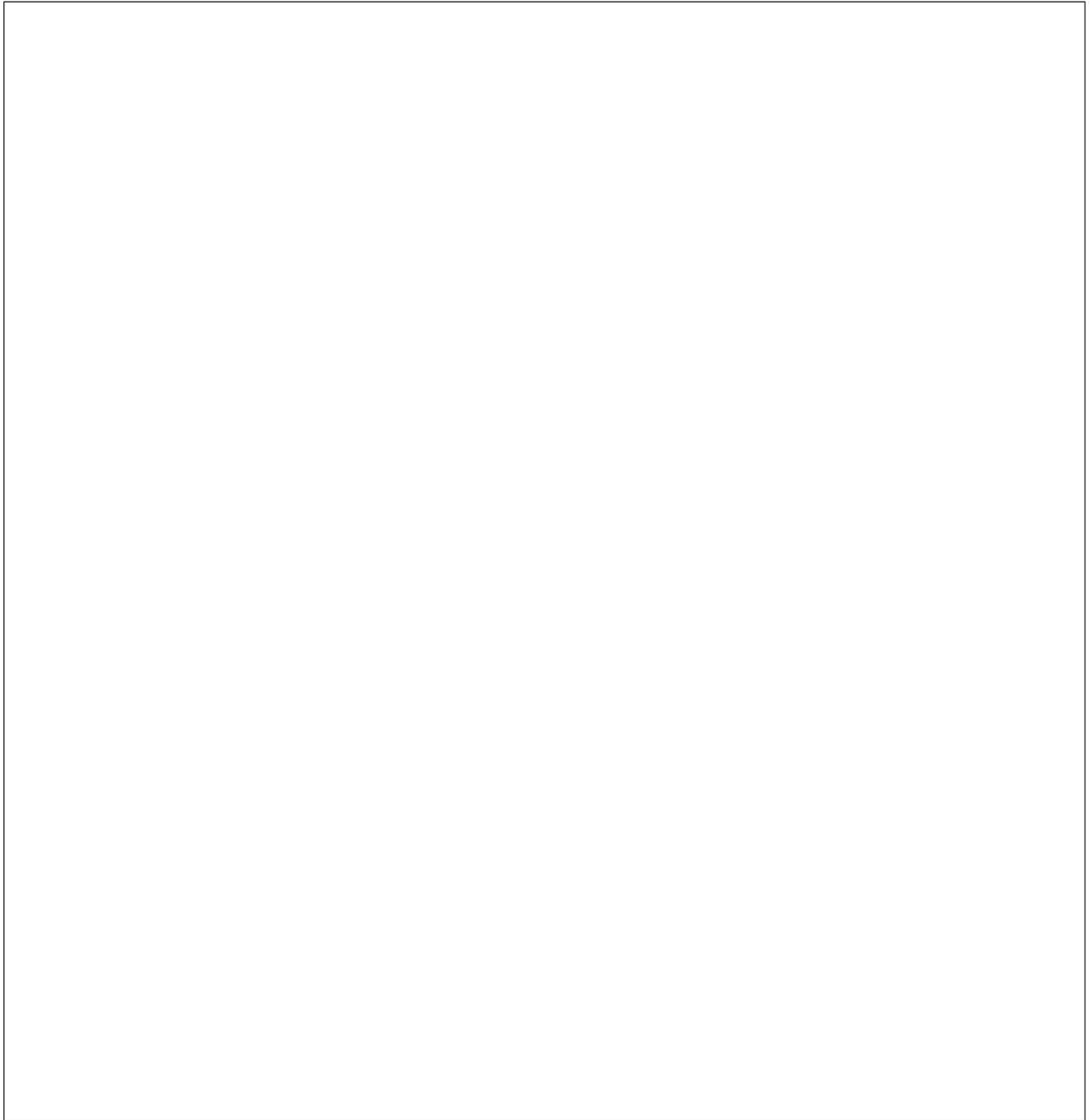
Donner un algorithme Brute-force pour DDP et suggérer une approche branch-and-bound pour améliorer sa performance.



Exercice 2. Reconstruction d'arbres phylogénétiques (5 points)

1. **(6 points)** Appliquer l'algorithme de Sankoff à la matrice pondérée M suivante et l'algorithme de Fitch à sa discretisation A ($A_{i,j} = 0$ si $M_{i,j} = 0$ et $A_{i,j} = 1$ si $M_{i,j} \neq 0$)

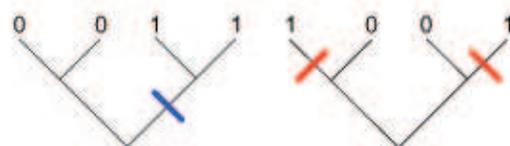
	A	C	T	G
A	3	8	2	0
C	3	1	0	1
T	0	2	4	5
G	2	2	1	1



2. (5 points) Donner les conditions d'équivalence des deux algorithmes et montrer que sous ces conditions, ils se comportent dans la même manière.



3. (6 points) On veut déterminer la meilleure phylogénie étant donnée une matrice binaire M , où M est une matrice $n \times n$ telle que $M_{i,j}$ représente la valeur du j -ème caractère de la i -ème espèce. On dit qu'un caractère binaire c est compatible avec un arbre T ssi il existe une affectation des états du caractère aux noeuds de T , tel qu'un seul arc est affecté par un changement. La figure



illustre des caractères binaires compatibles et incompatibles sur 4 espèces. Le caractère de l'arbre de gauche est compatible avec l'arbre car seulement un changement (indiqué par la ligne bleue transverse) est demandé pour évoluer de la racine vers les feuilles. Au contraire, le caractère de l'arbre de droite demande deux changements.

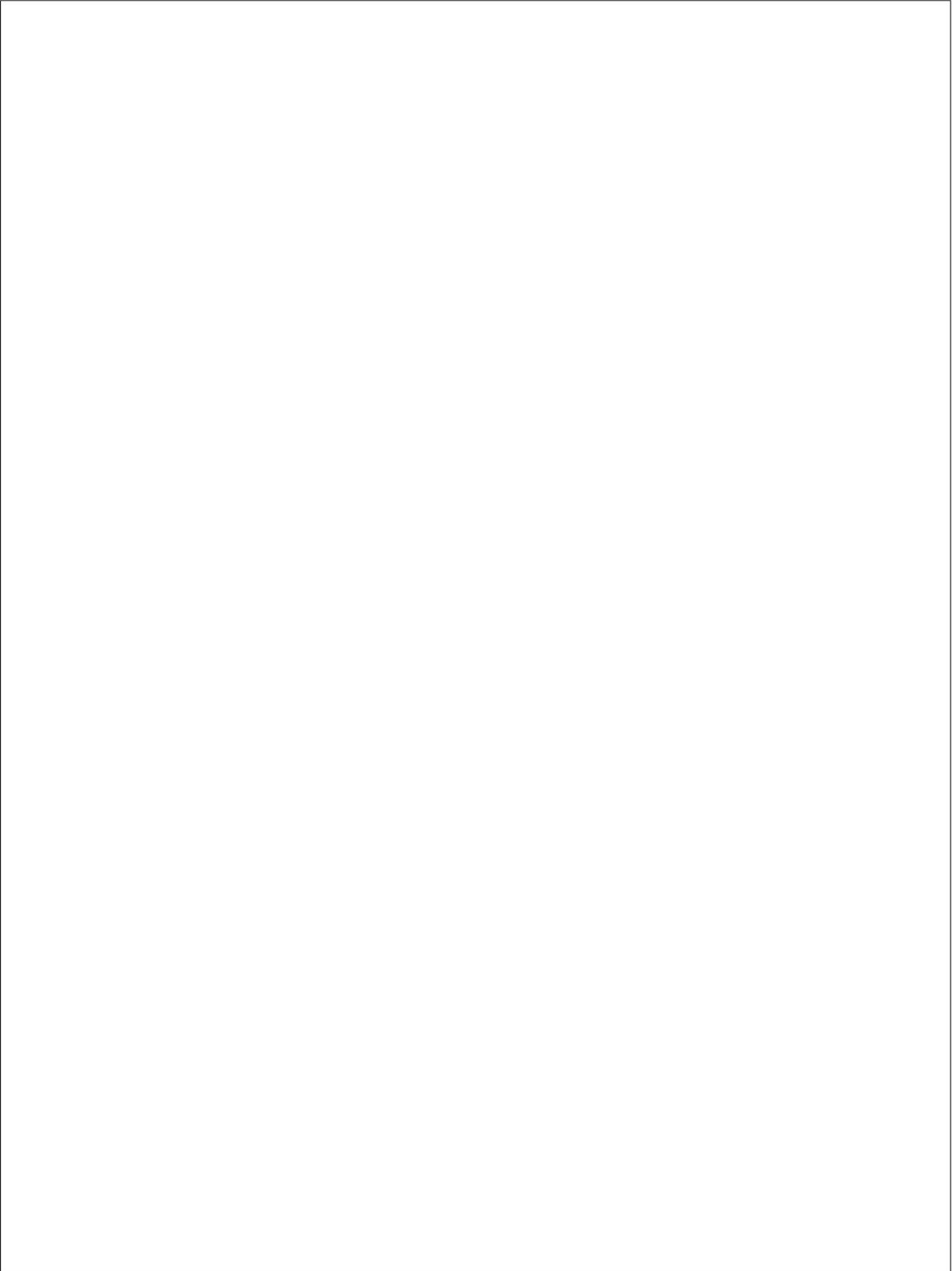
On dit que les caractères binaires c_1, c_2 sont compatibles en paire ($PC(c_1, c_2)$) s'il existe un arbre T tel que pour chaque i , c_i est compatible avec T .

Démontrer la proposition suivante:

Pour chaque caractère i, j , soit $S_{i,j} = \{(x, y) : \exists \text{ espèces } k \text{ tel que } M_{k,i} = x \text{ and } M_{k,j} = y\}$. Alors, $PC(c, c')$ ssi $S_{cc'} \neq \{0, 1\}^2$.



4. **(8 points)** Entre les différentes méthodes de parcimonie, l'algorithme "Nearest Neighbor Interchange" peut être vu comme un algorithme pour l'exploration de l'espace des arbres phylogénétique à n feuilles. Cet échantillonnage de l'espace permet-il de trouver des arbres optimaux ? Proposer deux versions de l'algorithme basées sur des opérations combinatoires alternatives de manipulation des arbres. Expliquer pourquoi ces deux versions sont intéressantes et dans quel sens elles diffèrent.



Exercice 3 (10 points) Rearrangement des génomes

Appliquer l'un des algorithmes de rearrangement des génomes vus en cours (en mettant en évidence à chaque étape le bloque renversé) à la permutation suivante

12 31 34 28 26 17 29 4 9 36 18 35 19 1 16 14 32 33 22 15 11 27 5 20 13 30 23 10 6 3 24 21 8 25 2 7

qui représente l'arrangement des gènes de l'ADN mitochondriale du ver *Ascaris Suum* pour trouver l'identité (id) représentant l'ADN mitochondriale de l'homme.



Exercice 4 (15 points) Réseaux biologiques

Donner quelques exemples d'algorithme de génération de graphes aléatoires très larges. Discuter les propriétés et les différences entre ces modèles, les raisons qui ont conduit à leur définition et leur applicabilité dans la modelisation des réseaux réels.

