

Comparaison des structures d'ARN

Claire Herrbach

Cours M2 BIBS
08 octobre 2007

Plan du cours

Introduction

Modélisation des structures d'ARN

Edition et alignement d'arbres

Edition et alignement de structures d'ARN

Fonctions de l'ARN

- ▶ Intermédiaire entre ADN et protéines (ARN messenger)
- ▶ Transcription (ARN polymérase et ARN 6S)
- ▶ Traduction (ARN de transfert et ARN ribosomiques)
- ▶ Maturation des pré-ARNm
- ▶ ...

+ 500 familles d'ARN dans la base de données RFam.

Fonctions des ARN et comparaison des structures

Conformation spatiale et fonction sont très liées.

Applications de la comparaison des structures d'ARN

- ▶ Prédiction de fonction ou de structures.
- ▶ Phylogénie.
- ▶ Classification en familles fonctionnelles.
- ▶ Identification de motifs fonctionnels.

Structures secondaires sans pseudo-nœud

Deux structures similaires peuvent avoir des séquences très différentes.

- ▶ Comparer les séquences n'est pas suffisant.

Structures tertiaires : complexité algorithmique.

Structures secondaires sans pseudo-noeuds plus faciles à modéliser.

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	G	A	T	T	A	C	A
C							
A							
T							
C							
A							
T							

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	0	G	1	A	2	T	3	T	4	A	5	C	6	A	7
C															
1	---														
A															
2	---														
T															
3	---														
C															
4	---														
A															
5	---														
T															
6	---														

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

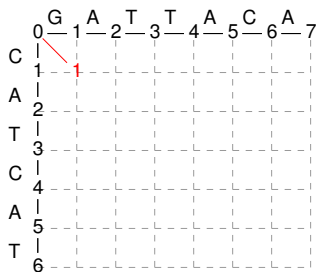
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

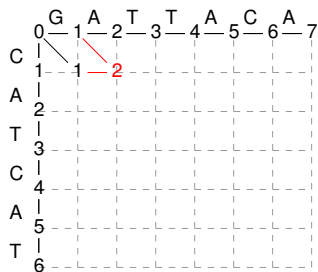
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

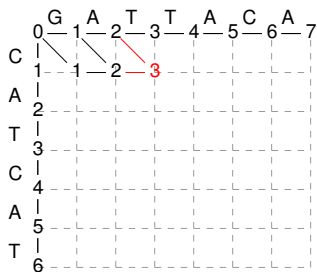
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	G	A	T	T	A	C	A	
0	1	2	3	4	5	6	7	
C	1	1	2	3	4	5	5	6
A	2	2	1	2	3	4	5	5
T	3	3	2	1	2	3	4	5
C	4	4	3	2	2	3	3	4
A	5	5	4	3	3	2	3	3
T	6	6	5	4	3	3	3	4

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	0	1	2	3	4	5	6	7
C	1	1	2	3	4	5	5	6
A	2	2	1	2	3	4	5	5
T	3	3	2	1	2	3	4	5
C	4	4	3	2	2	3	3	4
A	5	5	4	3	3	2	3	3
T	6	6	5	4	3	3	3	4

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	G	A	T	T	A	C	A	
	0	1	2	3	4	5	6	7
C		/		/		/		/
1	1	1	2	3	4	5	5	6
A		/		/		/		/
2	2	2	1	2	3	4	5	5
T		/		/		/		/
3	3	3	2	1	2	3	4	5
C		/		/		/		/
4	4	4	3	2	2	3	3	4
A		/		/		/		/
5	5	5	4	3	3	2	3	3
T		/		/		/		/
6	6	6	5	4	3	3	3	4

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$

	0	G	A	T	T	A	C	A	7
C	0	1	2	3	4	5	6	7	
A	1	1	2	3	4	5	5	6	
T	2	2	1	2	3	4	5	5	
C	3	3	2	1	2	3	4	5	
A	4	4	3	2	2	3	3	4	
T	5	5	4	3	3	2	3	3	
T	6	6	5	4	3	3	3	4	

Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

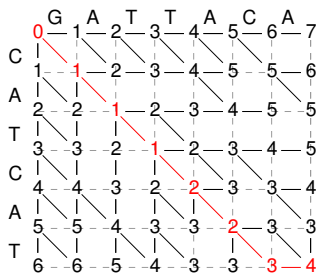
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

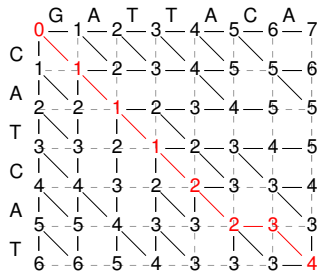
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

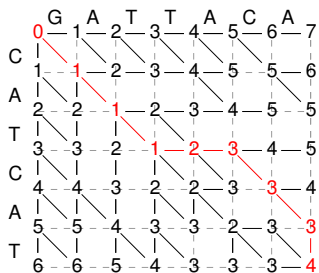
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

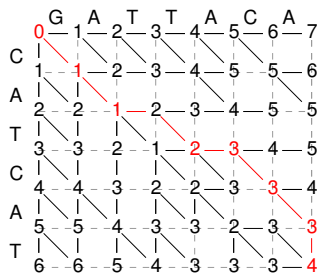
$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Alignement de séquences (ADN, protéines ou ARN)

Distance : minimisation d'un coût,

Score : maximisation.

$$\delta(i, j) = \min \begin{cases} \delta(i-1, j) + \text{del}(i) \\ \delta(i, j-1) + \text{ins}(j) \\ \delta(i-1, j-1) + \text{subst}(i, j) \end{cases}$$



Avec les coûts : $\text{del}(i) = \text{ins}(j) = 1$,

$$\text{subst}(i, j) = \begin{cases} 0 & \text{si } S[i] = S'[j] \\ 1 & \text{sinon} \end{cases}$$

Plan du cours

Introduction

Modélisation des structures d'ARN

Edition et alignement d'arbres

Edition et alignement de structures d'ARN

Séquences arc-annotées

Séquence de lettres : séquence primaire d'un ARN,

Ensemble de couples : appariements de la structure d'ARN
(représentés par des arcs).

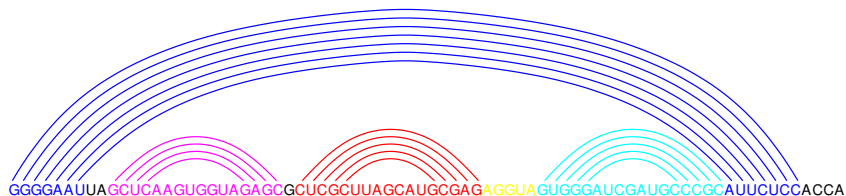
GGGGAAUUA^{GCUCAAGUGGUAGAGCGCUCGCUUAGCAUGCGAGAGGUAGUGGGAUCGAUGCCCGCAUUCUCCACCA}

PLAIN (sans arcs) : séquence sans appariement

Séquences arc-annotées

Séquence de lettres : séquence primaire d'un ARN,

Ensemble de couples : appariements de la structure d'ARN
(représentés par des arcs).

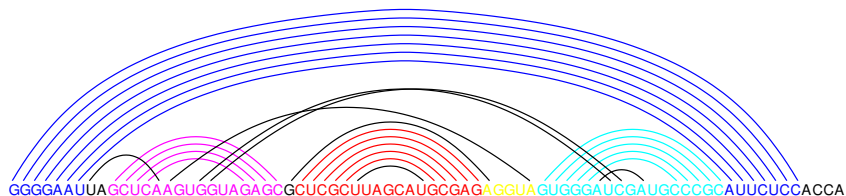


NESTED (imbriquée) : structure secondaire sans pseudo-noeud

Séquences arc-annotées

Séquence de lettres : séquence primaire d'un ARN,

Ensemble de couples : appariements de la structure d'ARN
(représentés par des arcs).

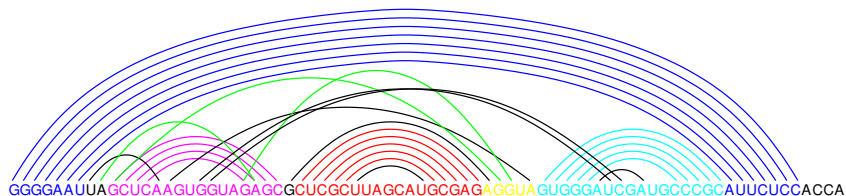


CROSSING (croisée) : structure secondaire avec pseudo-noeuds

Séquences arc-annotées

Séquence de lettres : séquence primaire d'un ARN,

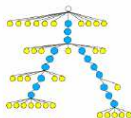
Ensemble de couples : appariements de la structure d'ARN
(représentés par des arcs).



UNLIMITED (générale) : structure tertiaire

Arbres et forêts

On peut construire des arbres avec différent degré de précision.



1



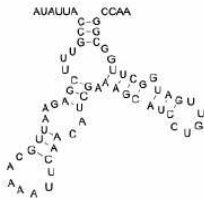
2



3



4

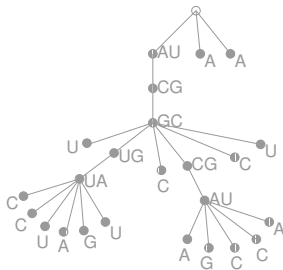
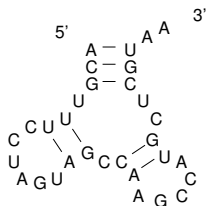


5

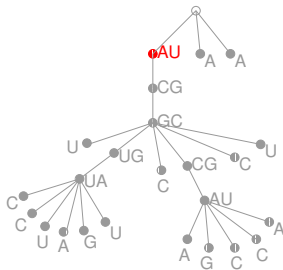
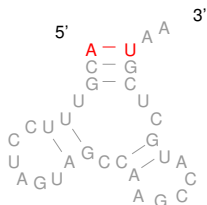


6

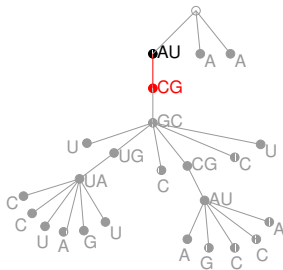
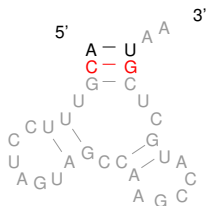
Construction de l'arbre



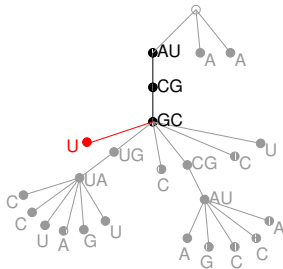
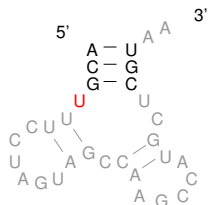
Construction de l'arbre



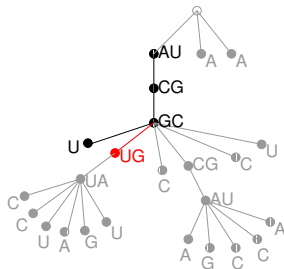
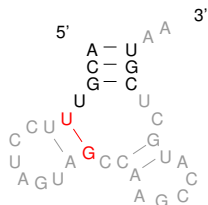
Construction de l'arbre



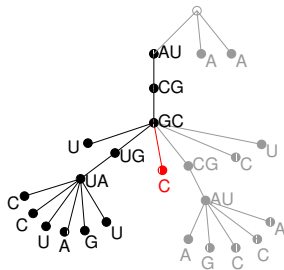
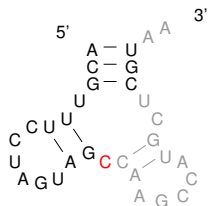
Construction de l'arbre



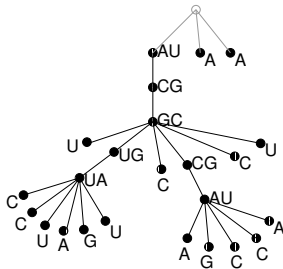
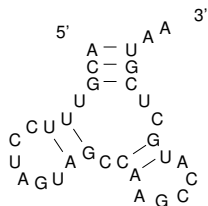
Construction de l'arbre



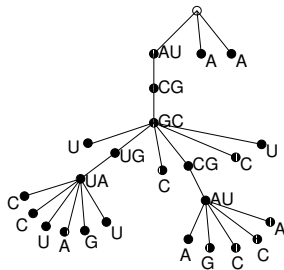
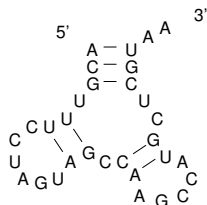
Construction de l'arbre



Construction de l'arbre



Construction de l'arbre



- ▶ Il est possible d'ajouter un racine pour obtenir un arbre.
- ▶ Les bases non appariées sont des feuilles.
- ▶ Les paires de bases sont des sommets internes.

Plan du cours

Introduction

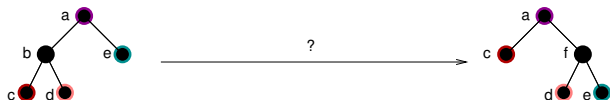
Modélisation des structures d'ARN

Edition et alignement d'arbres

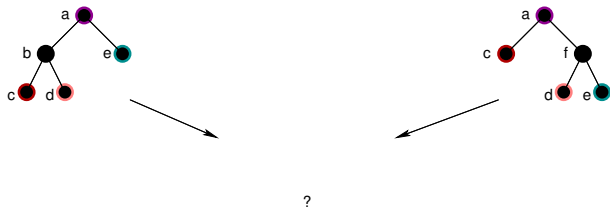
Edition et alignement de structures d'ARN

Edition et alignement d'arbres

Edition : Transformer un arbre en un autre.

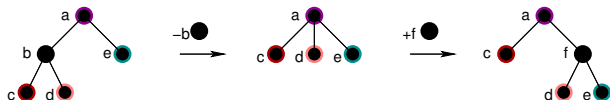


Alignement : Construire un super-arbre commun à deux arbres.

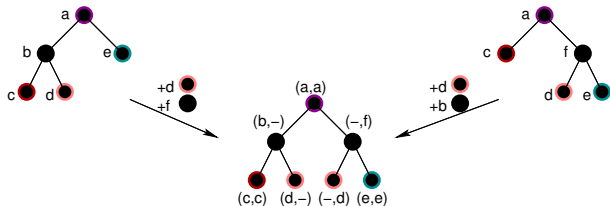


Edition et alignement d'arbres

Edition : Transformer un arbre en un autre.



Alignement : Construire un super-arbre commun à deux arbres.

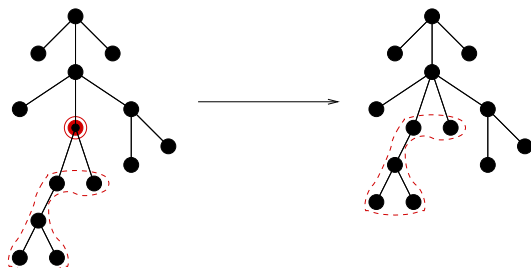


Opérations d'édition sur les sommets des arbres

Suppression du sommet v :
les fils de v deviennent

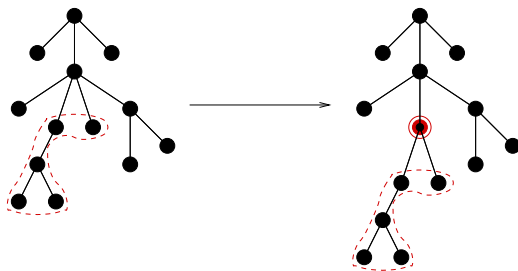
1. fils du père de v ,
2. frères droits des frères gauches de v ,
3. frères gauches des frères droits de v ,

et leur ordre est conservé.



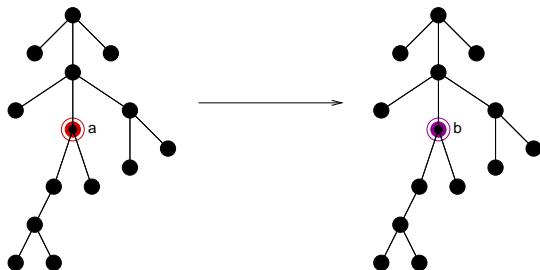
Opérations d'édition sur les sommets des arbres

Insertion du sommet v :
un sommet v est ajouté comme père d'un certain nombre de frères consécutifs (opération symétrique de la suppression).



Opérations d'édition sur les sommets des arbres

Substitution du sommet v :
l'étiquette de v est modifiée.

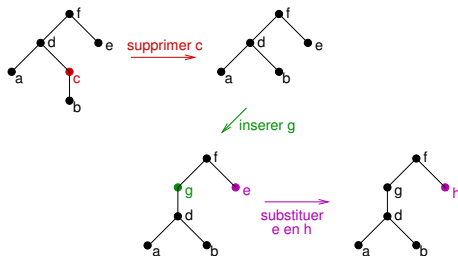


Edition d'arbres

Chaque opération d'édition a un coût.

Edition : Transformer un arbre en un autre

- ▶ par une suite d'opérations d'édition,
- ▶ en minimisant le coût.

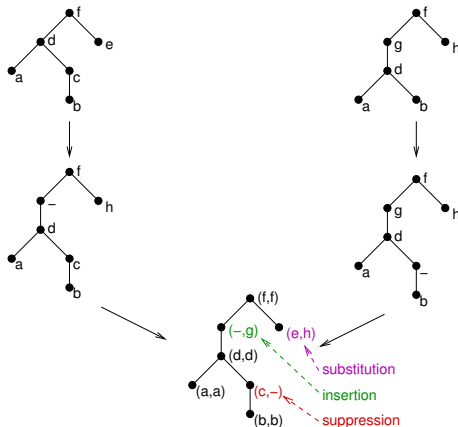


Alignement d'arbres

Chaque opération d'édition a un coût.

Alignement : Construire un super-arbre commun à deux arbres

- ▶ par une suite d'opérations d'édition,
- ▶ en minimisant le coût.



Algorithme de Zhang & Shasha, 1989

- ▶ Edition d'arbres :

$$\delta(\text{arbre}_1, \text{arbre}_2) = \min \begin{cases} \delta(\text{arbre}_1, \text{arbre}_2) + \text{del}(\bullet) \\ \delta(\text{arbre}_1, \text{arbre}_2) + \text{ins}(\bullet) \\ \delta(\text{arbre}_1, \text{arbre}_2) + \text{subst}(\bullet, \bullet) \end{cases}$$

- ▶ Edition de forêts :

$$\delta(\text{forêt}_1, \text{forêt}_2) = \min \begin{cases} \delta(\text{forêt}_1, \text{forêt}_2) + \text{del}(\bullet) \\ \delta(\text{forêt}_1, \text{forêt}_2) + \text{ins}(\bullet) \\ \delta(\text{forêt}_1, \text{forêt}_2) + \delta(\text{arbre}_1, \text{arbre}_2) \end{cases}$$

Algorithme de Jiang, Wang & Zhang, 1995

- ▶ Alignement d'arbres :

$$\delta(\text{arbre}_1, \text{arbre}_2) = \min \begin{cases} \delta(\text{arbre}_1, \text{arbre}_2) + \text{del}(\bullet) \\ \delta(\text{arbre}_1, \text{arbre}_2) + \text{ins}(\bullet) \\ \delta(\text{arbre}_1, \text{arbre}_2) + \text{subst}(\bullet, \bullet) \end{cases}$$

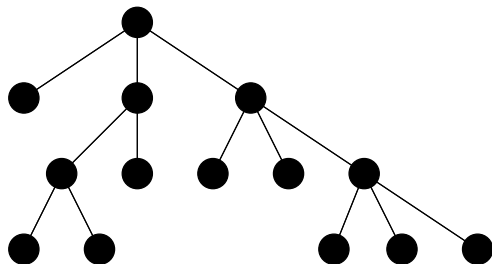
- ▶ Alignement de forêts :

$$\delta(\text{forêt}_1, \text{forêt}_2) = \min \begin{cases} \min\{\delta(\text{forêt}_1, \text{forêt}_2) + \delta(\text{forêt}_1, \text{forêt}_2) \mid \text{forêt}_1 = \text{forêt}_2\} \\ \quad + \text{del}(\bullet) \\ \min\{\delta(\text{forêt}_1, \text{forêt}_2) + \delta(\text{forêt}_1, \text{forêt}_2) \mid \text{forêt}_1 = \text{forêt}_2\} \\ \quad + \text{ins}(\bullet) \\ \delta(\text{forêt}_1, \text{forêt}_2) + \delta(\text{forêt}_1, \text{forêt}_2) \end{cases}$$

Comment décomposer les arbres ?

Exercice : Quels sont les sous-arbres et sous-forêts de l'arbre suivant nécessaires pour

- ▶ l'algorithme de Zhang & Shasha ?
- ▶ l'algorithme de Jiang, Wang & Zhang ?



Décomposition des arbres

Zhang & Shasha

- ▶ Sommets numérotés selon l'ordre suffixe de gauche à droite.
- ▶ Pour chaque sous-arbre dont la racine est numérotée i , les sous-forêts nécessaires sont composées des sommets numérotés j à k où :
 - ▶ j est la feuille la plus à gauche de i ,
 - ▶ $j \leq k \leq i$.

Jiang, Wang & Zhang

- ▶ Les sous-forêts nécessaires sont composées d'arbres frères consécutifs.

Complexité

Zhang & Shasha

- ▶ $\mathcal{O}(n^4)$ dans le pire cas
[Zhang & Shasha 1989]
- ▶ $\mathcal{O}(n^3)$ en moyenne
[Dulucq & Tichit 2003]

Jiang, Wang & Zhang

- ▶ $\mathcal{O}(n^4)$ dans le pire cas
[Jiang, Wang & Zhang 1995]
- ▶ $\mathcal{O}(n^2)$ en moyenne
[Denise, Dulucq & Herrbach 2007]

Plan du cours

Introduction

Modélisation des structures d'ARN

Edition et alignement d'arbres

Edition et alignement de structures d'ARN

Opérations d'édition réalistes pour l'ARN

<i>base-match</i>	$A \rightarrow A$
<i>base-mismatch</i>	$A \rightarrow G$
<i>base-deletion</i>	$A \rightarrow -$ $- \rightarrow A$
<i>arc-match</i>	$A \widehat{U} \rightarrow A \widehat{U}$
<i>arc-mismatch</i>	$A \widehat{U} \rightarrow G \widehat{C}$
<i>arc-removing</i>	$A \widehat{U} \rightarrow - -$ $- - \rightarrow A \widehat{U}$
<i>arc-breaking</i>	$A \widehat{U} \rightarrow A U$ $A U \rightarrow A \widehat{U}$
<i>arc-altering</i>	$A \widehat{U} \rightarrow A -$ $A \widehat{U} \rightarrow - U$ $A - \rightarrow A \widehat{U}$ $- U \rightarrow A \widehat{U}$

[Jiang, Lin, Ma & Zhang 2002]

Edition de séquences arc-annotées

Transformer une séquence arc-annotée en une autre.



Edition de séquences arc-annotées

EDIT(PLAIN,PLAIN)	nm [Needleman & Wunsch 1970]
EDIT(NESTED,PLAIN)	nm^3 [Jiang, Lin, Ma & Zhang 2002]
EDIT(NESTED,NESTED)	NP-complet [Blin, Fertin, Rusu & Sinoquet 2003]
EDIT(CROSSING,PLAIN)	MAX-SNP-dur [Jiang, Lin, Ma & Zhang 2002]
EDIT(CROSSING,NESTED)	MAX-SNP-dur [Jiang, Lin, Ma & Zhang 2002]
EDIT(CROSSING,CROSSING)	MAX-SNP-dur [Jiang, Lin, Ma & Zhang 2002]

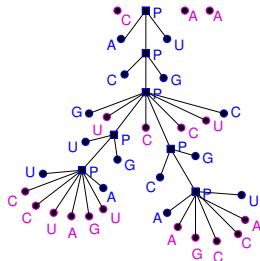
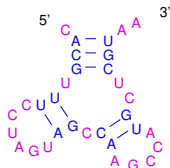
Alignement de structures d'ARN

- ▶ Construire une super-structure d'ARN commune à deux structures.
- ▶ Modélisation par des séquences arc-annotées ou par des arbres.
- ▶ Jeu d'opérations d'édition réalistes pour l'ARN.

RNAforester [Höchsmann 2005]

Alignement de structures secondaires d'ARN sans pseudo-nœud.

Modélisation



Méthode

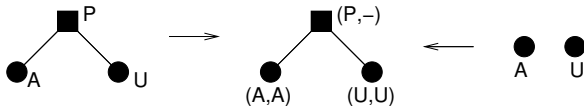
Alignement de forêts avec l'algorithme de Jiang, Wang & Zhang.

La substitution d'un sommet interne (P) et d'une feuille (base) est interdite.

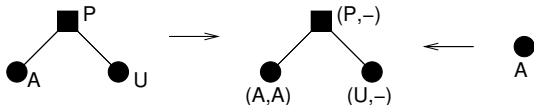
Opérations d'édition réaliste dans le modèle de RNAforester

Les opérations réalistes pour l'ARN sont simulées par composition des trois opérations simples.

arc-breaking

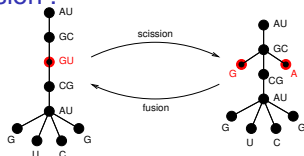


arc-altering

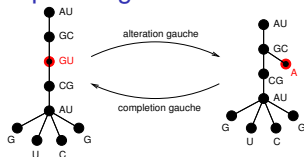


Opérations réalistes pour l'ARN, dans les forêts

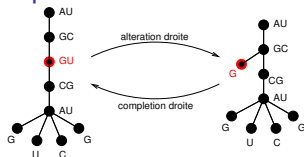
Scission / Fusion :



Altération / Complétion gauche :

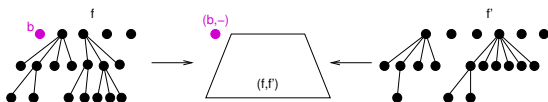


Altération / Complétion droite :

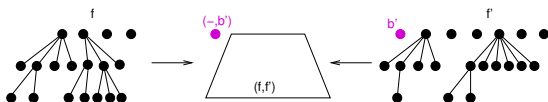


Opérations sur les bases (feuilles)

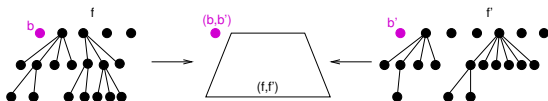
Suppression : $\text{BDel}(b) + \delta(f, f')$



Insertion : $\text{BIns}(b') + \delta(f, f')$

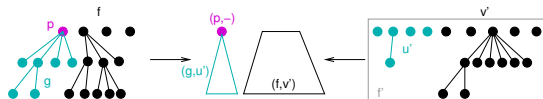


Substitution : $\text{BSub}(b, b') + \delta(f, f')$

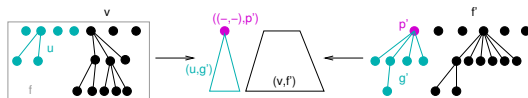


Opérations sur les paires (nœuds internes)

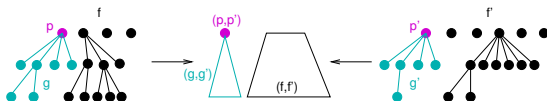
Suppression : $PDel(p) + \delta(g, u') + \delta(f, v')$



Insertion : $PIns(p') + \delta(u, g') + \delta(v, f')$

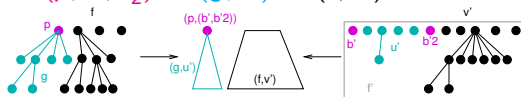


Substitution : $PSub(p, p') + \delta(g, g') + \delta(f, f')$

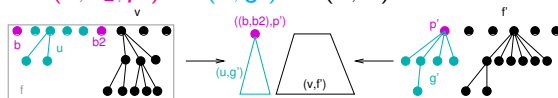


Opérations sur les paires et les bases

Scission : $\text{Sci}(p, b', b'_2) + \delta(g, u') + \delta(f, v')$

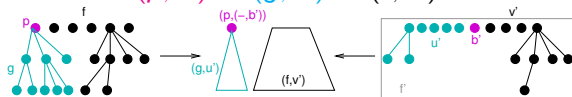


Fusion : $\text{Fus}(b, b_2, p') + \delta(u, g') + \delta(v, f')$



Opérations sur les paires et les bases

Altération gauche : $GAlt(p, b') + \delta(g, u') + \delta(f, v')$



Complétion gauche : $GComp(b, p') + \delta(u, g') + \delta(v, f')$

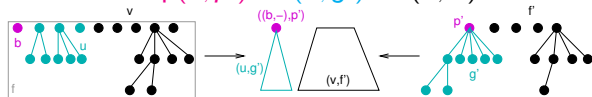


Opérations sur les paires et les bases

Altération droite : $DAlt(p, b') + \delta(g, u') + \delta(f, v')$



Complétion droite : $DComp(b, p') + \delta(u, g') + \delta(v, f')$



Alignement de structures secondaires d'ARN sans pseudo-nœud

[Herrbach, Denise, Dulucq et Touzet 2006]

$$\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) =$$

$$\min \left\{ \begin{array}{l} \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \text{BDel}(\bullet) \\ \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \text{BIns}(\bullet) \\ \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \text{BSub}(\bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{PDel}(\bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{PIns}(\bullet) \\ \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \text{PSub}(\bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{Fus}(\bullet, \bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{Sci}(\bullet, \bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{GAlt}(\bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{DAlt}(\bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{GComp}(\bullet, \bullet) \\ \min\{\delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) + \delta(\text{▲▲▲▲▲}, \text{▲▲▲▲▲}) : \text{▲▲▲▲▲} = \text{▲▲▲▲▲}\} + \text{DComp}(\bullet, \bullet) \end{array} \right.$$

si \bullet base

si \bullet base

si \bullet et \bullet bases

si \bullet paire

si \bullet paire

si \bullet et \bullet paires

si \bullet paire et \bullet base

si \bullet paire et \bullet base

si \bullet paire et \bullet base

si \bullet paire

si \bullet paire et \bullet base

si \bullet paire

Décomposition

Même décomposition que pour l'algorithme de Jiang, Wang & Zhang.

Complexité

- ▶ $\mathcal{O}(n^4)$ dans le pire cas
- ▶ $\mathcal{O}(n^2)$ en moyenne

D'autres méthodes

- ▶ Edition de structures secondaires d'ARN sans pseudo-nœud par tiges-boucles [Guignon, Chauve & Hamel 2005]
- ▶ RNA-MiGaL : édition d'arbres multi-échelles [Allali & Sagot 2004]
- ▶ Edition d'une séquence croisée et d'une séquence imbriquée avec contrainte sur les coûts [Jiang, Lin, Ma & Zhang 2002]

Bibliographie I



K. Zhang and D. Shasha

Simple fast algorithms for the editing distance between trees and related problems

SIAM J. COMPUT., vol. 18, no 6, pp 1245-1262, 1989.

Klein-Verlag, 1990.



S. Dulucq and L. Tichit

RNA secondary structure comparison : exact analysis of the Zhang-Shasha tree edit algorithm

Theoretical Computer Science, 306, pp 471-484, 2003.



P. Klein

Computing the edit-distance between unrooted ordered trees

Proceedings of 6th European Symposium on Algorithms, pp 91-102, 1998.

Bibliographie II



S. Dulucq and H. Touzet

Analysis of tree edit distance algorithms

Combinatorial Pattern Matching, 2676, pp 83-95, 2003.



T. Jiang, L. Wang and K. Zhang

Alignment of trees - an alternative to tree edit

Theoretical Computer Science, 143, pp 137-148, 1995.



T. Jiang, G. Lin, B. Ma and K. Zhang

A general edit distance between RNA structures

RECOMB'01.



G. Blin, G. Fertin, I. Rusu and C. Sinoquet

RNA Sequences and the EDIT(NESTED,NESTED) Problem

Research Report RR-IRIN-03.07, IRIN, 2003.

Bibliographie III



M. Höchsmann, T. Töller, R. Giegerich and S. Kurtz

Local similarity in RNA secondary structures

CSB 2003.



M. Höchsmann

The tree alignment model : algorithms, implementations and applications for the analysis of RNA secondary structures.

PhD Thesis, 2005.



C. Herrbach, A. Denise, S. Dulucq and H. Touzet

Alignment of RNA secondary structures using a full set of operations

Research Report LRI-1451, 2006.

Bibliographie IV



C. Herrbach

Etude algorithmique et statistique de la comparaison des structures secondaires d'ARN

PhD Thesis, 2007.



V. Guignon, C. Chauve and S. Hamel

An edit distance between RNA stem-loops

SPIRE 2005.



J. Allali and M.-F. Sagot

Novel tree edit operations for RNA secondary structure comparison

Proceedings of WABI, 2006.



J. Allali

Comparaison de structures secondaires d'ARN

PhD thesis, 2004.