

Origine des locus microsatellites : mécanismes et hypothèses

Erick Desmarais

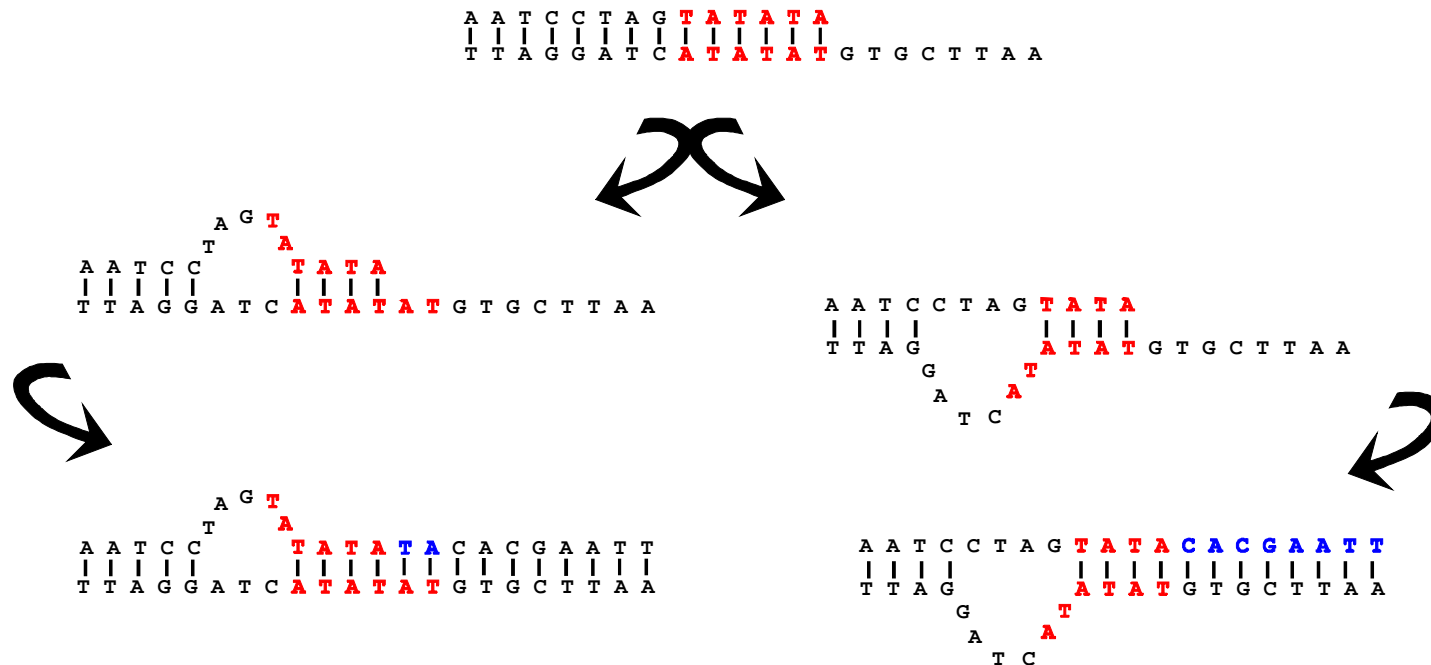
Labo. Génome Populations et Interactions

UMR5000 CNRS Univ. Montpellier II

Locus très fréquents

- Chez l'homme : Chromosome 22 séquencé (33,4Mb)
 - dinucleotide : 1775 locus, 1 /19000pb
 - tri : 166 locus, 1/201000pb
 - tetra : 404 locus, 1/83000pb
 - penta : 16 locus, 1/2087000pb
 - autres : 305 locus, 1/109000pb
 - Total VNTR : 2666 locus, 1/12500pb
 - Longueur totale : 303723 pb

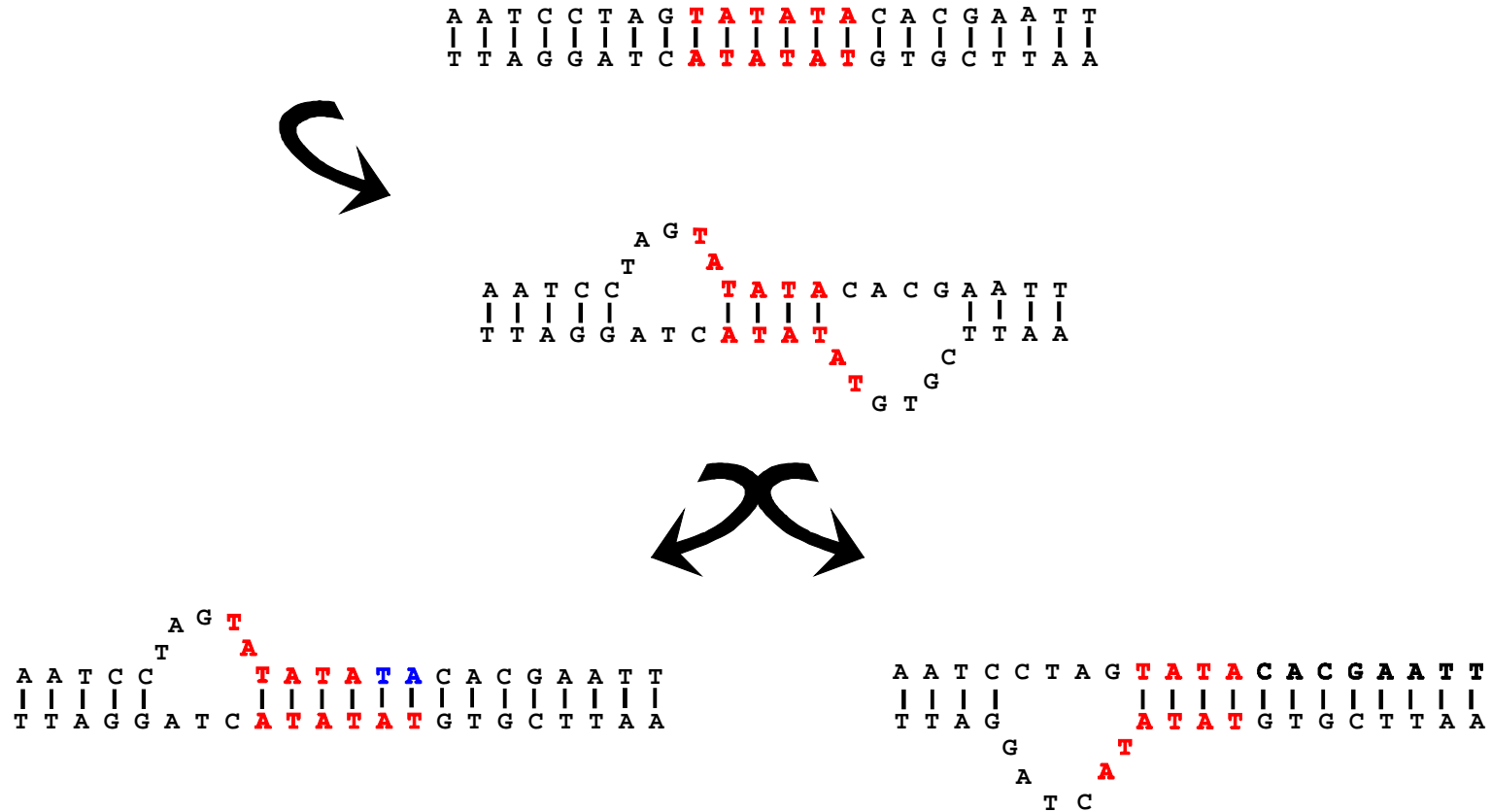
Slipped-strand Mispairing (SSM)



Boucle sur brin néo-synthétisé:
Addition d'un motif

Boucle sur matrice :
Déletion d'un motif

En dehors de replication



Suivant le sens de la réparation, insertion ou délétion sur le brin du dessus

Comment peut-on arriver à des séquences susceptibles de subir le SSM ?

- Apparition au hasard de duplications de motifs
 - Proba. d'avoir un polynucleotide de 3 répétitions d'un motif dinucleotidique = $1/256$
- Motifs polynucléotidiques associés aux retrotransposons
Noyaux pour l'émergence d'un plus grand nombre de répétitions car substrat favorable au slippage.
Effet auto-accélérateur de l'expansibilité des microsatellites.

2 phénomènes séparés :

- Apparition de séquences susceptibles de donner prise aux erreurs par SSM (Substitutions, Mutations frameshift ou retrotransposition) : erreurs pouvant créer des polynucleotides
- Slippage mispairing : erreurs de glissement quand des répétitions ont commencé à se former.
- Et puis, ces 2 phénomènes impliquent une réparation.

MAIS...

Apparition de dinucléotides au hasard :

- Chez coli : biais quasi-systématique vers des suppressions de nucléotides/répétitions dans le cas de mutations de frameshift
 - Difficile pour des microrépétitions d'émerger
 - Difficile de se maintenir
- Quasiment pas de mutations quand les polymerases utilisées ont une activité correctrice (cas de la Pol. delta)

Origine à partir des queues polyA des Alu

```

A A T C C T A G A A A A A A A A A
| | | | | | | | | | | | | | | | |
T T A G G A T C T T T T T T T T T T
    
```



Substitution

```

A A T C C T A C A G A A A A A A A A
| | | | | | | | | | | | | | | | |
T T A G G A T G T C T T T T T T T T T
    
```



Misappariement

```

      A C A
      |   |
      T   G
A A T C C T A G A A A A A A A A A
| | | | | | | | | | | | | | | | |
T T A G G A T G T C T T T T T T T T T
    
```



Remplacement du A par G mais
Mauvaise réparation de la boucle

```

      A C A
      |   |
      T   G
A A T C C T A G A G A A A A A A A A
| | | | | | | | | | | | | | | | |
T T A G G A T G T C T T T T T T T T T
    
```

Association entre Eléments répétés dispersés et microsats

- Association des microsats riches en AT avec Alu et L1
- Association d'autres motifs avec d'autres SINES (B1, MT, LTRs)

Zuliani G, Hobbs HH (1990). *Am J Hum Genet* 46:963-9

Jurka J, Pethiyagoda C (1995). *J Mol Evol* 40:120-6

Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996). *Proc Natl Acad Sci U S A* 93:6470-5

Kaukinen J, Varvio SL (1992). *Nucleic Acids Res* 20:2955-8

Kelly R, Bulfield G, Collick A, Gibbs M, Jeffreys AJ (1989). *Genomics* 5:844-56

Bois P, Williamson J, Brown J, Dubrova YE, Jeffreys AJ (1998). *Genomics* 49:122-8

Quelques chiffres :

- Chro.22 Dunham 1999 (33,4 Mb) :
 - Alu : 20188 locus, 1/1655pb
 - L1 : 8043 locus, 1/4153pb
 - Total : 28231 locus, 1/1183pb

(Rappel : microsat 1/12500pb)

2 hypothèses sur l'origine de cette association :

- Les Sines et Lines apportent les microsats.
Queue polyA ou microsats riches en AT
- Les microsatellites préexistent et les éléments répétés dispersés s'y intègrent préférentiellement

Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995). *Genomics* 29:136-44

Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996). *Proc Natl Acad Sci U S A* 93:6470-5 11

Analyse comparative

- Comparaison de locus orthologues de primates : Insertion dans des régions non répétées
- Site d'insertion non fixés dans l'espèce humaine : queue polyA courte et pas de microsatellite

MAIS...

- Si dégénérescence progressive : alors les familles jeunes d'Alu associées aux motifs les plus riches en A. [(ANA)_n= 1 pas, (GATA)_n= 2 pas]. Ce n'est pas le cas
- Si présence de polynucléotides, biais de mutations en faveur de la délétion.
- Cela n'explique pas tous les types de motifs
Le modèle simple de SSM n'est pas si simple

Yandava CN, Gastier JM, Pulido JC, Brody T, Sheffield V, Murray J, Buetow K, et al (1997). Genome Res 7:716-24

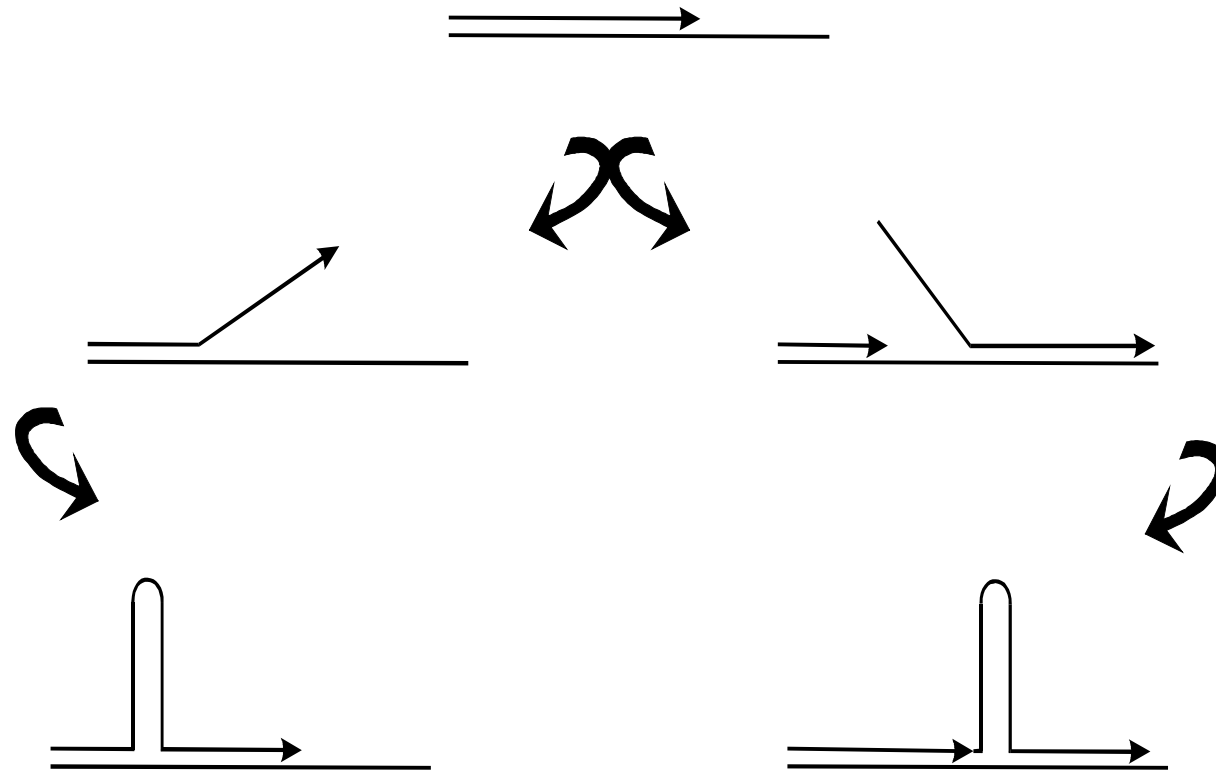
Qu'est-ce qui provoque les glissements ?

- Frameshift : Lien avec la processivité de la polymérase
- SSM : Lien avec le sens de la replication

Instabilité plus forte au moment de la synthèse du lagging strand, car synthèse discontinue.

Modèle de déplacement de l'extrémité du fragment d'Okazaki.

Déplacement du fragment d'Okazaki



Slippage sur blocs de motifs répétés : Biais en faveur des délétions

- Implication du système de mismatch-repair et NER : si déficience alors instabilité (Strand 1993 Parnieski 1999)
- Instabilité renforcée par la transcription (Wierdl 1996, Parnieski 1999)
- Biais global vers les délétions chez la levure (Strand 1995, Moore 1999, Schweitzer 1999) et chez coli (Bichara 2000)

Strand M, Prolla TA, Liskay RM, Petes TD (1993). *Nature* 365:274-6

Kang S, Jaworski A, Ohshima K, Wells RD (1995). *Nature Genetics* 10:213-218

Strand M, Earley MC, Crouse GF, Petes TD (1995). *Proc Natl Acad Sci U S A* 92:10418-21

Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD (1996). *Genetics* 143:713-21 Parniewski P,

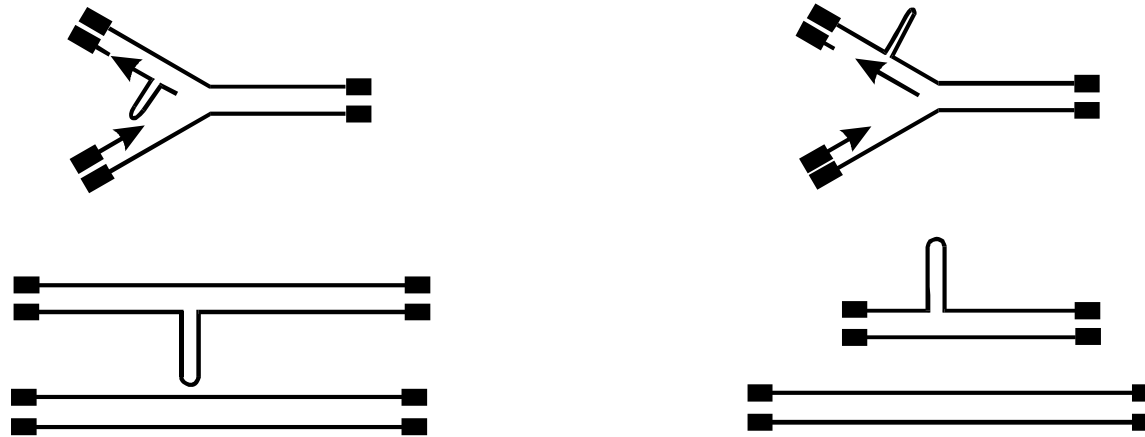
Bacolla A, Jaworski A, Wells RD (1999). *Nucleic Acids Res* 27:616-23

Moore H, Greenwell PW, Liu CP, Arnheim N, Petes TD (1999). *Proc Natl Acad Sci U S A* 96:1504-9

Schweitzer JK, Livingston DM (1999). *Genetics* 152:953-63

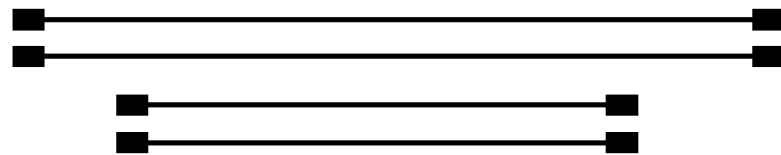
Bichara M, Pinet I, Schumacher S, Fuchs RP (2000). *Genetics* 154:533-542

Erreur de réplication

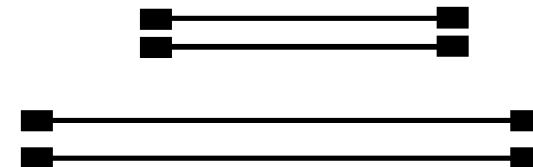


Structures hétéroduplexes : substrats pour mécanismes de réparation
Si réparation efficace alors élimination du brin néo-synthétisé erroné
et retour à la normale.

Il faut à la fois un problème au niveau de la réplication et une
erreur de réparation



Expansion



Délétion

Slippage : étape d'un processus plus complexe

- Apparition de séquences favorables à des glissements de polymérase : Pourquoi et comment évoluent ces « graines » de microsatellite ?
- Slippage mispairing : erreurs de glissement quand des répétitions ont commencé à se former. Qu'est-ce qui fait que les 2 brins peuvent glisser ?
- Les structures anormales créées par le glissement sont mal réparées. Pourquoi autant d'erreurs ?

Les éléments répétés dispersés peuvent induire ces erreurs

- Association de microsatellites avec des éléments mobiles sans queue polyA.
- Association avec des motifs pauvres en AT.

Apport d'éléments promoteurs de transcription

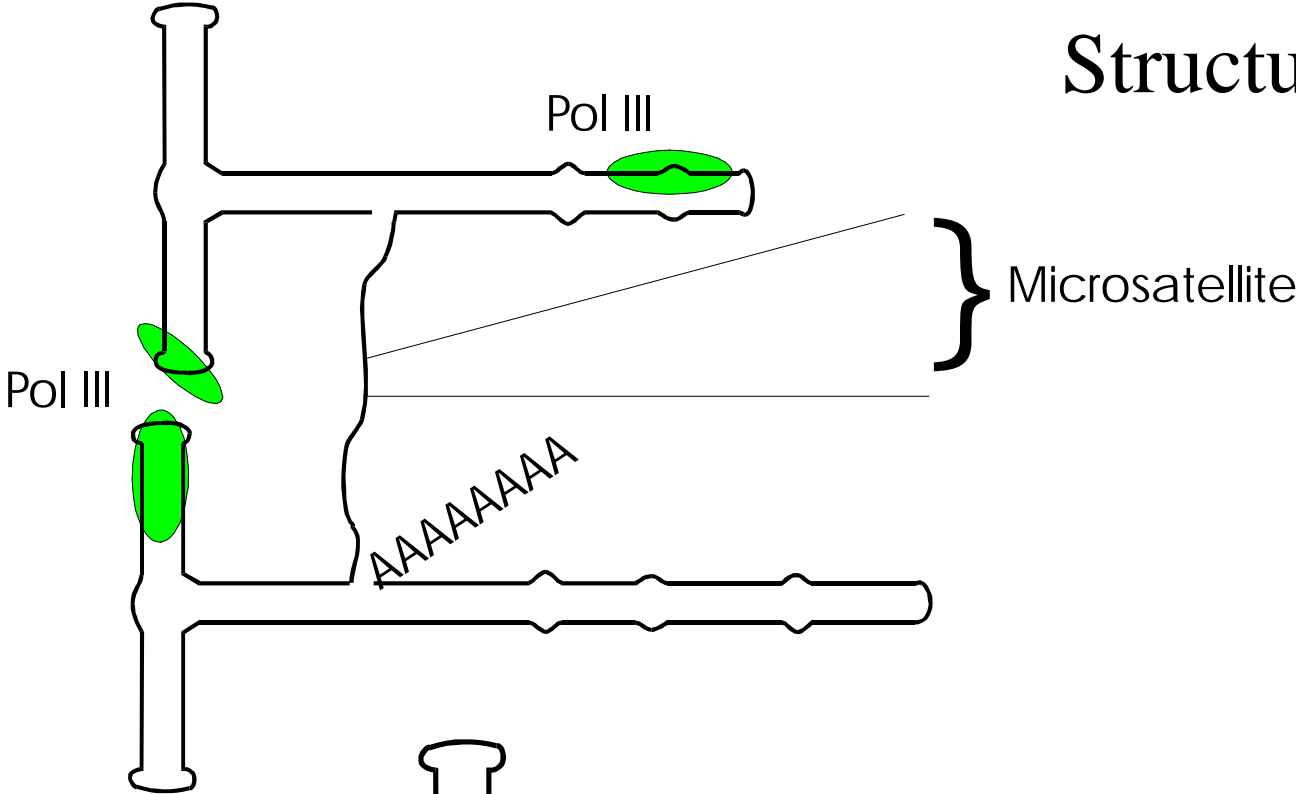
- Les Sines et LTR apportent avec eux leur promoteur de transcription.
- La transcription est un élément destabilisant la double-hélice.
- La seule présence de signaux de fixation de facteurs de transcription suffit.

Apport de séquences rompant la régularité de la double-hélice.

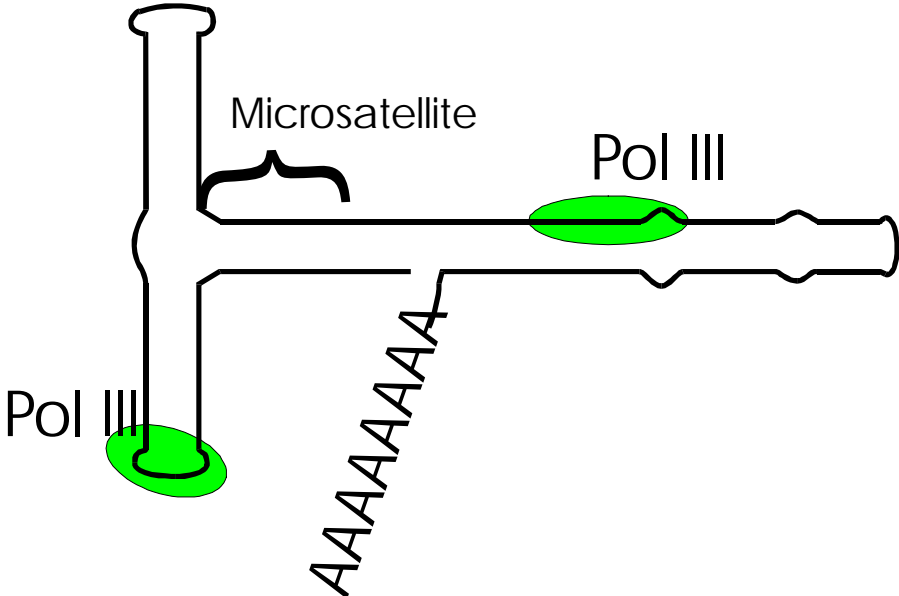
- La présence d'un polyA à elle seule entraîne une torsion de la double-hélice
- Les éléments mobiles ont des séquences potentiellement capables d'adopter des structures secondaires qui reflètent celles adoptées par leur ARN.

Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F (1988). *Faseb J* 2:2939-49

Structures secondaires



Alu

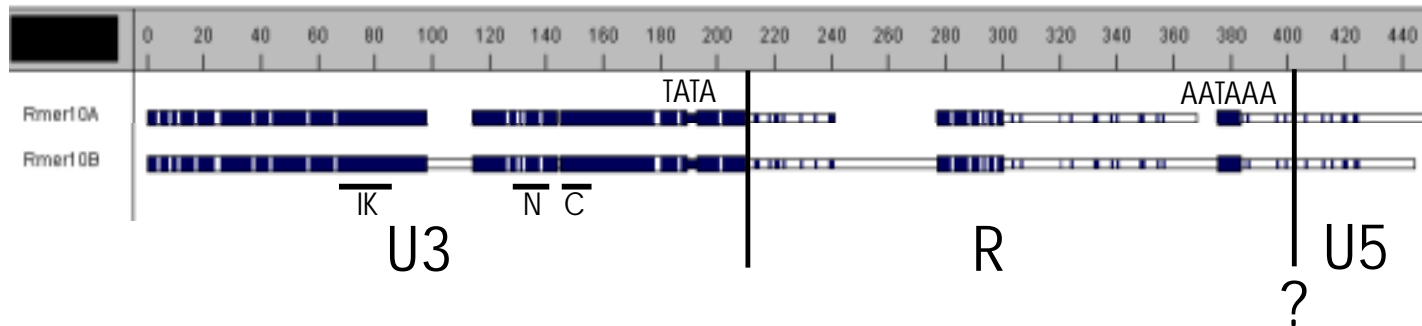


B1

Combinaison des deux : problèmes

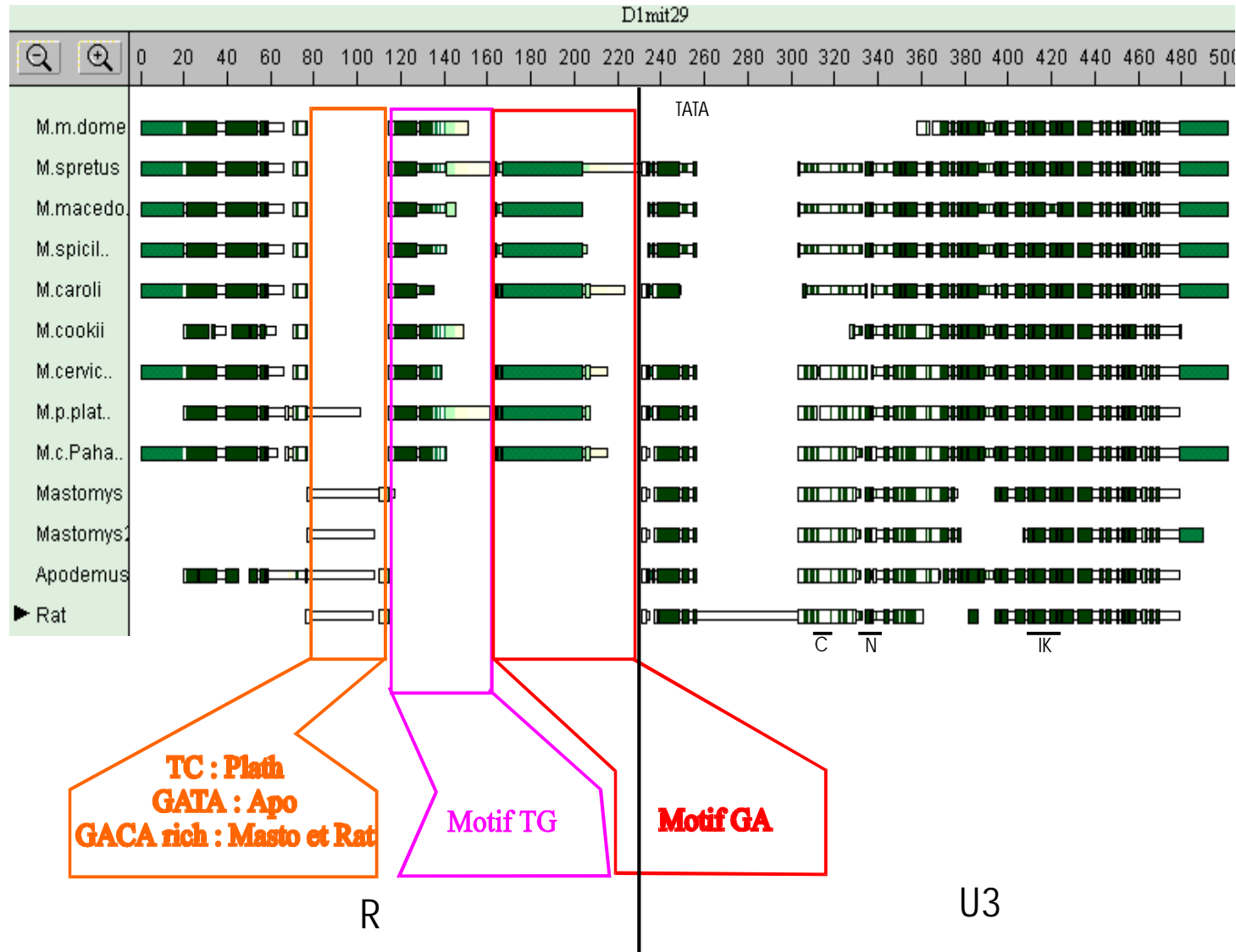
- Fixation de facteurs de transcription et/ou d'ARN polymérase
 - Déstabilisation de la double-hélice Les structures secondaires se forment sur les simple-brins dégagés.
- Télésopage avec la réplication ou mise en jeu de réparation.
- Instabilité des locus.

Les Rmer10 (LTR solo): Schema général et structure

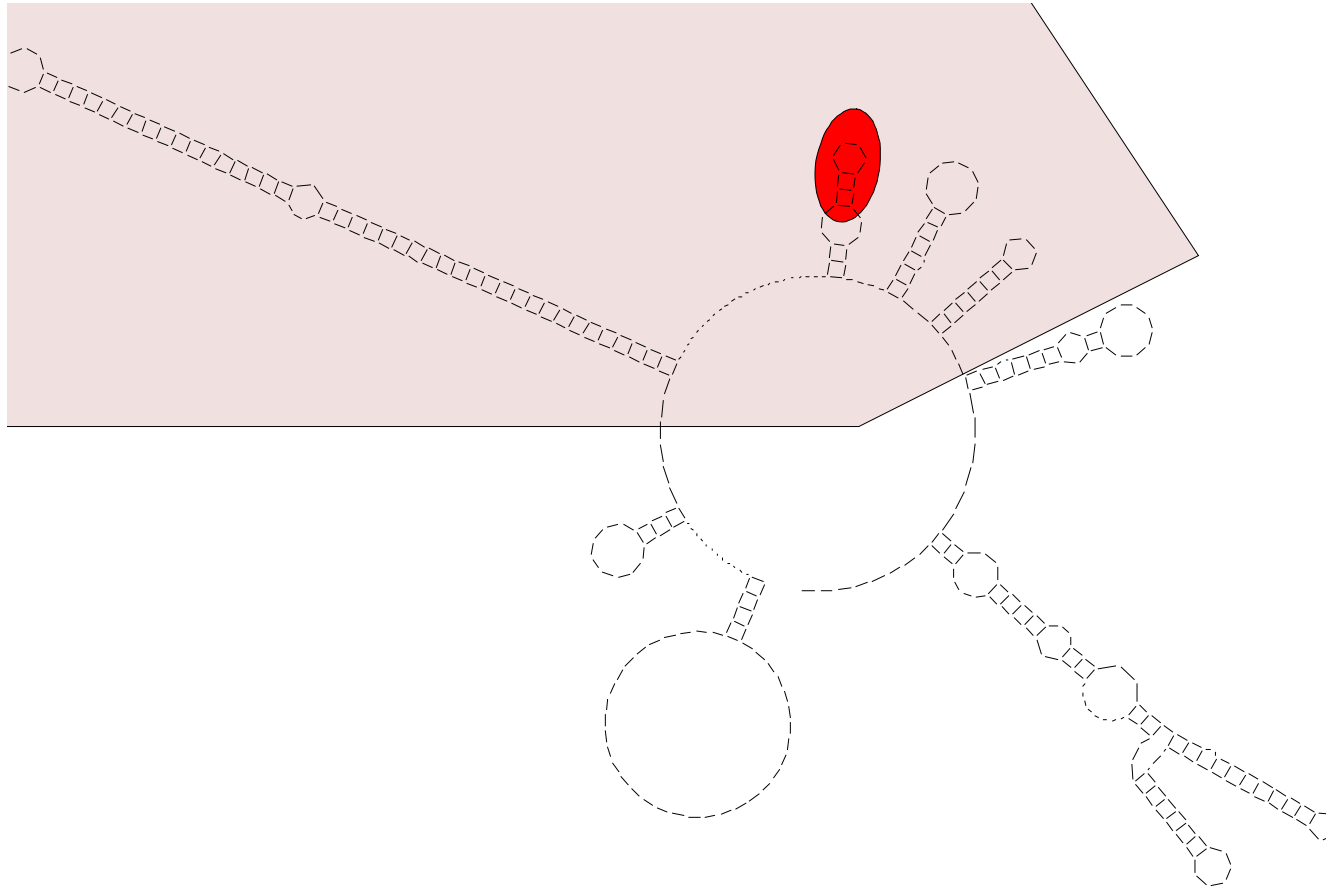


IK, N et C : facteurs d'activation de la transcription dans les cellules T

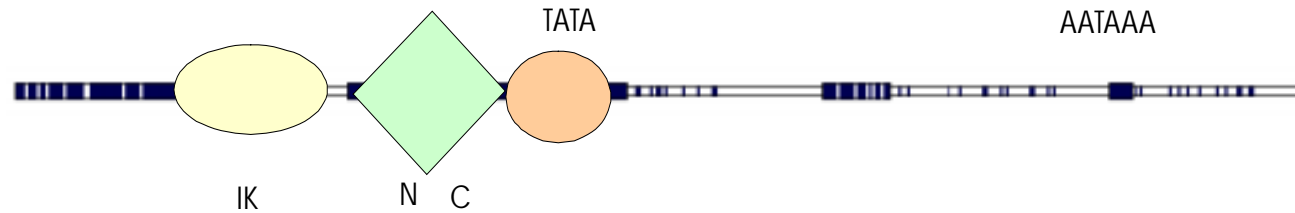
Insertion au site D1Mit29



Structure ancestrale avec séquence recomposée.

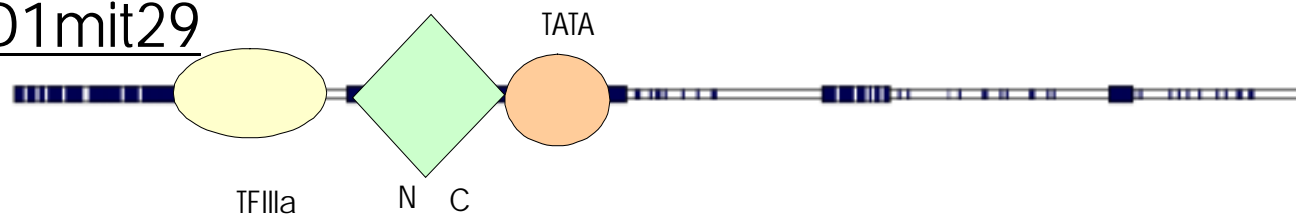


Rmer10B



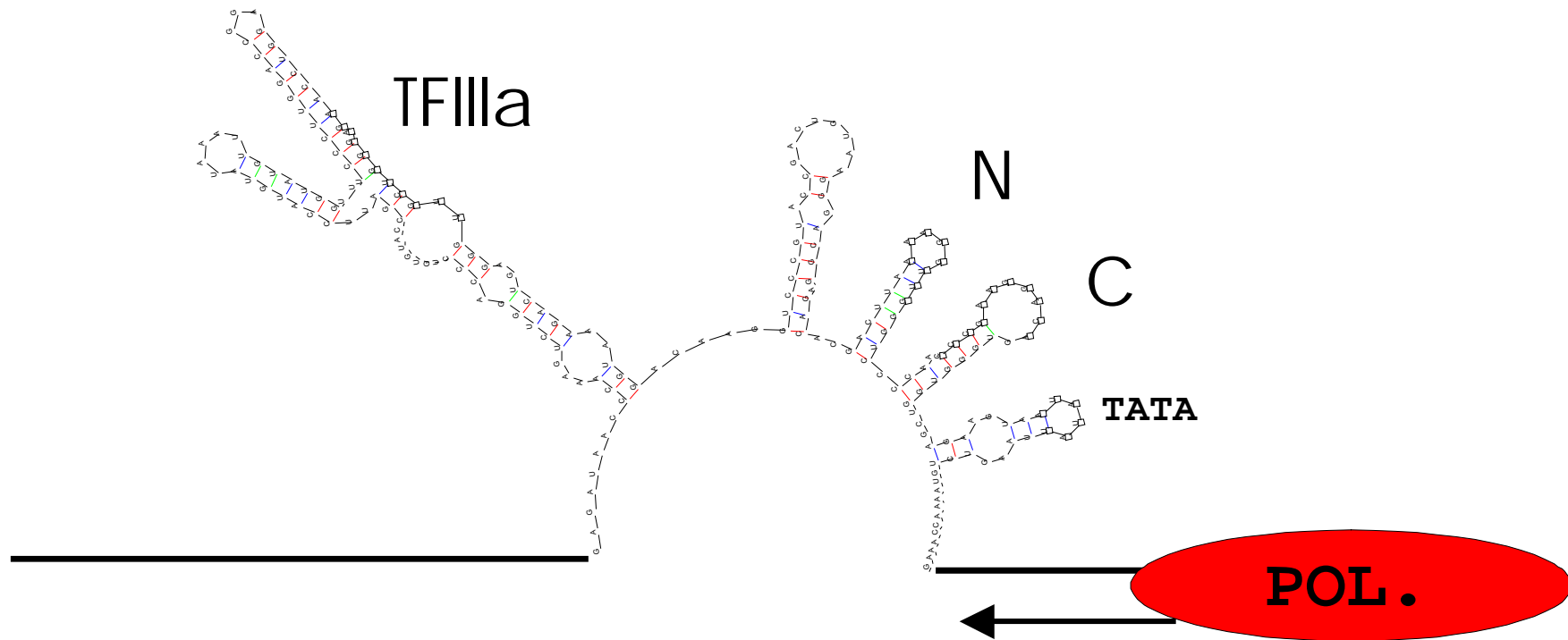
Modification de la région U3 : transformation du signal IK2 en TFIIIA et acquisition d'un palindrome très stable.

D1mit29



Production de microsatellites

Pause pendant la réplication : Slipped Strand Mispairing



Généralisation?

- D'autres cas décrits de LTR ou Sines contenant des répétitions en tandem.
- Exemple de MS32 : Répétitions de (GACTCAGAATGGAGCAGGCGGCCA GGGGT) au milieu d'un LTR.

Conclusion

- L'émergence de nouveaux locus microsatellites n'est peut-être pas due au hasard.
- Ils apparaissent peut-être préférentiellement en des endroits où la structure double-hélice de l'ADN est rendue instable (comme en cas d'insertion d'éléments répétés)