

**Quelques caractéristiques
des microsattellites :
définition, densité et structure**

P. Jarne

CEFE - Montpellier

Buts

*** Définitions**

*** Méthodes d'études de la distribution**

*** Quelques résultats : distribution dans les génomes**

*** Conclusion : modèles d'évolution des microsatellites**

Définition et historique

Déf. : les microsatellites sont des séquences d'ADN répétées en tandem dont l'unité de répétitions fait de une à six pb

Noms divers : SSR, STR, short tandem repeats, VNTR, maladies à triplets...

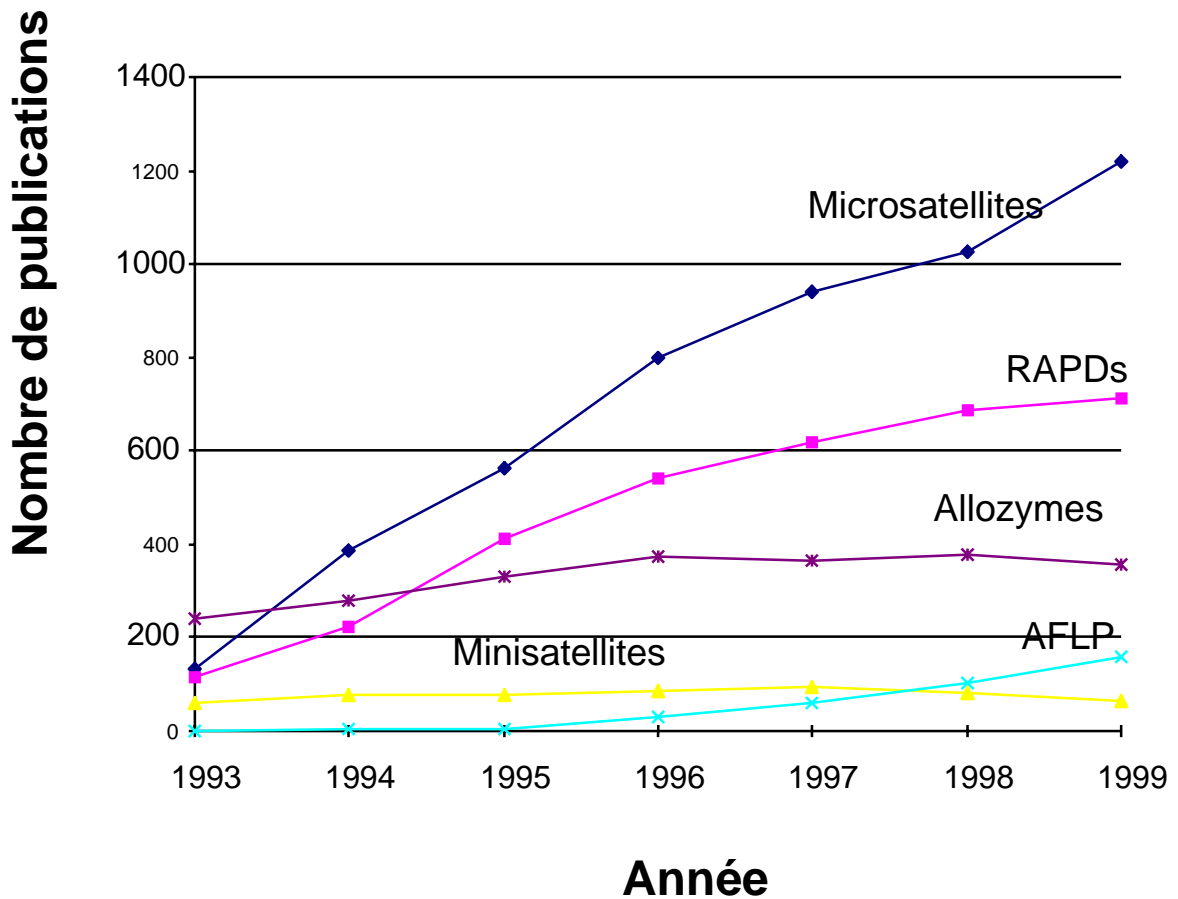
Une grande famille : satellites, minisatellites et microsatellites

Quelques dates :

- 1960 -1965 : premières descriptions
- 1980 -1985 : présence dans de nombreux génomes
- 1989 : utilisation en tant que marqueurs génétiques

Définition et historique

Nombre de publications - BIOSIS



Définition et historique

Trois familles (Weber 1990) :

- pur : $(CA)_n$ avec $n > 4$

- composé : $(CA)_n(GA)_m$

avec n et $m > 3$

- interrompu : $(CA)_n N_l (GA)_m$

avec $l < 4$ et n et $m > 3$

+ motifs juxtaposés

Type : longueur de l'unité de répétition

- mononucléotide : A_n ou C_n

- dinucléotide : $(CA)_n$, $(GA)_n$, $(TA)_n$
et $(CG)_n$

- trinucleotide : 10 motifs

Définition et historique

Exemples :

- *D. simulans* : (CA)₁₅

(CA)₄(AC)₂ CT(CA)₁₂(TG)₁₀TATG

- homme :

(CGG)₆N₅₇ (G)₁₀ (CGG)₁₀(NGG)₇

(gène de protéase)

Locus : lieu du génome contenant un microsatellite (parfois variable), caractérisé par des régions flanquantes

Allèles : généralement définis par leur taille (en pb), et par inférence par leur nombre de répétitions

Méthodes d'analyse

But : analyser la distribution de distances entre locus microsatellites (présence / absence, distance moyenne, agrégation)

+ données sur la taille (moyenne, allèle cloné)

*** Hybridation d'ADN total (dot blot) : densité moyenne**

Voir Primmet et al., 1997, Genome Research 7: 471-482.

Problèmes : calibration, connaissance de la taille du génome

Méthodes d'analyse

* **Clonage : $n / (x N)$**

n = nombre clones positifs, x = taille clones, N = nombre clones

**Problèmes : coupure irrégulière par enzymes,
taille pas toujours connue précisément,
conditions de stringence**

=> peu précis, mais large spectre

\times

* **Hybridation *in situ* sur chromosomes :**

- visualisation directe et globale

Problème : sensibilité limitée

Méthodes d'analyse

- * Recherche en banques de séquences
 - ex. : Genbank ou TIGR
 - procédures de type BLAST
 - génomes entiers séquencés (*S. cerevisiae*)

Problèmes : représentativité des banques (cDNA, ADN codant, restreinte à quelques espèces) et procédures de recherches utilisées par différents auteurs

Présence

Echelle phylogénétique

Recherche dans BIOSIS (1993-1999)

- bactéries : 46
- champignons : 101
- protistes : 24
- animaux : 4600
- plantes : 420

NB : absence dans certains génomes bactériens

Fraction du génome représentée par les MS

* Tous motifs confondus :

- bactéries : $< 10^{-4}$
- *S. cerevisiae* : $6 \cdot 10^{-4}$
- homme : $5 \cdot 10^{-3}$

* Dinucléotides :

- poulet : 10^{-4}
- rongeurs : $3 \cdot 10^{-3}$
- homme : 6 à $9 \cdot 10^{-4}$

=> fraction assez faible

Densité

Espèce	Tous motifs	Dinucléotides
Bactéries	160	300
<i>S. cerevisiae</i>	20	30
<i>D. melanogaster</i>	40	60
Poulet	30	100
Homme	6	30

Densité

Comparaison entre types

Espèce	Mono	Tri	Tétra
<i>S. cerevisiae</i>	113	35	< 0.1
<i>D. melanogaster</i>	-	45	6
Poulet	175	260	190
Homme	340	45	140
<i>A. thaliana</i>	75	370	15

Nb : valeur des dinucléotides fixées à 100

Densité

Comparaison entre motifs dinucléotidiques

Espèce	GA	TA	CG
<i>S. cerevisiae</i>	33	840	0
<i>D. melan.</i>	33	67	2
Poulet	23	35	<0.1
Homme	30	20	<0.5
<i>A. thaliana</i>	75	370	15

Nb : valeur des CA fixées à 100

Densité

*** Comparaison entre familles (Weber 1990) :**

- pur : 0,65

- interrompus : 0,25

- composés : 0,10

*** Données plus complètes (rongeurs) :**

- 1/3 des $(GA)_n = (GA)_m (CA)_l$

- 1/10 des $(TA)_n = (TA)_m (CA)_l$

Distribution génomique

- * **Données anecdotiques**
- * **Cartes génétiques (homme) : distribution régulière, et moindre densité sur le X**
- * **Mono- à trinuécléotides chez la levure**

Voir Field & Wills, 1998, PNAS 95: 1647-1652.

Association avec autres éléments

Analyse de 3 Mbp de génome humain
(Nadir et al. 1996)

Element	N	3' / <i>Alu</i>	Autres (ET)	Pas de liens
A	393	77	11	12
CA	78	22	13	65
NA₂₋₅	312	78	11.5	11.5

* Lien fonctionnel entre microsat. et *Alu*

* Résultat non confirmé chez le poulet

Association avec régions codantes

**Analyse de 3 Mbp de génome du poulet
(Primmer et al. 1997)**

Zone	Mono, di et tétranucl.	Trinucl.
Non codant	31	4
UTR	23	3
Codant	0	28
?	16	7
Total	75	42

*** Distributions différentes**

***NB : biais en faveur des zones codantes**

Taille

*** Taille max. = qqes dizaines de répétitions**

*** Distributions de tailles en banques : dinucléotides chez rongeurs**

n	CA_n	GA_n	TA_n	CG_n
2 < n < 5	63305	79545	18956	7740
5 < n < 10	277	151	58	9
n > 10	335	142	29	1

Taille

Comparaison avec modèles (Bell & Jurka 1997)

- Bernoulli
- marche aléatoire

Voir Bell & Jurka, 1997, J. Mol. Evol. 44:
414-421.

Conclusion

But : proposer un modèle d'évolution (vie-mort)

*** Approche comparative :**

- génomes entiers**
- paramètres : taille du génome, richesse en AT, niveau de ploïdie, rapport codant / non codant**

*** Approche par modélisation**