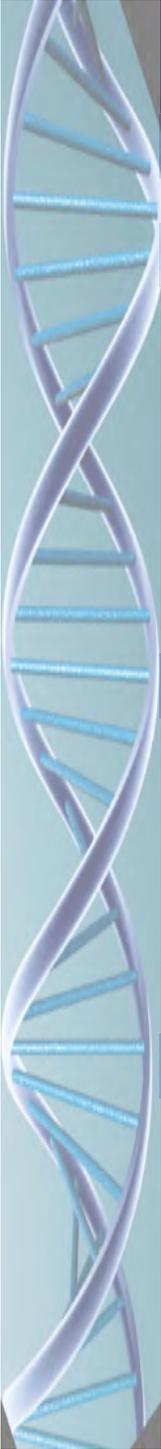


Nouvelles technologies de séquençage : quelques applications en écologie moléculaire

Jawad Abdelkrim

Muséum National d'Histoire Naturelle
jawad@mnhn.fr



Plan

•1 - Introduction à l'écologie moléculaire

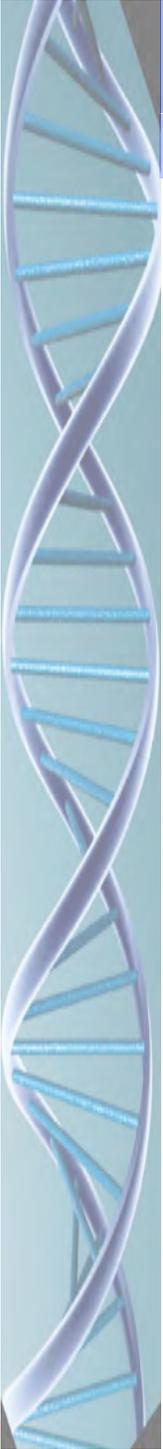
- But et histoire
- Généralités et rappels de génétique moléculaire
 - Clonage
 - PCR
 - Séquençage traditionnel

•2 - Introduction aux techniques de séquençage de nouvelle génération

- Généralités
- Plateformes principales
- Principe du pyroséquençage
- Détails de la plateforme Roche/454

•3 - Quelques applications en écologie moléculaire

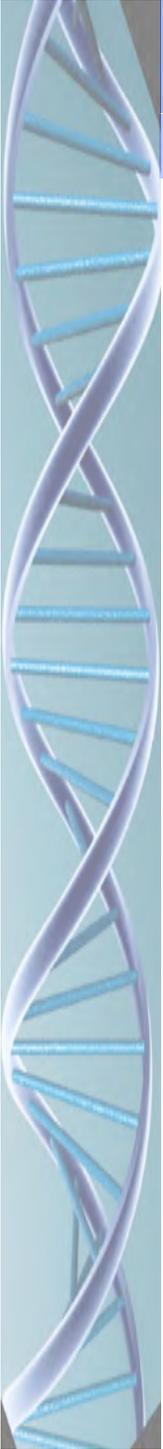
- Développement de marqueurs microsatellites
- Caractérisation de régimes alimentaires
- Détermination de la biodiversité à partir d'échantillonnages environnementaux
- Détermination de la biodiversité présente dans un environnement passé



1 - Introduction à l'écologie moléculaire

But et histoire

- Ecologie moléculaire = étude de problématiques écologiques à l'aide de techniques de biologie moléculaire
- 'Problématique écologiques'
 - Ecologie comportementale
 - Structure de population, flux de gènes
 - Spéciation
 - Conservation
 - Ecologie des communautés
 - Adaptation



1 - Introduction à l'écologie moléculaire

But et histoire

- C'est une discipline relativement 'jeune' qui suit l'évolution des techniques moléculaires et de la compréhension du fonctionnement de la machinerie génétique.
 - 1953 La structure en double hélice de l'ADN est établie par James D. Watson and Francis Crick
 - 1958 L'expérience de Meselson-Stahl démontre que l'ADN est répliqué de façon semi-conservative
 - 1970 Découverte des enzymes de restriction qui vont permettre de couper et coller des fragments d'ADN



1 - Introduction à l'écologie moléculaire

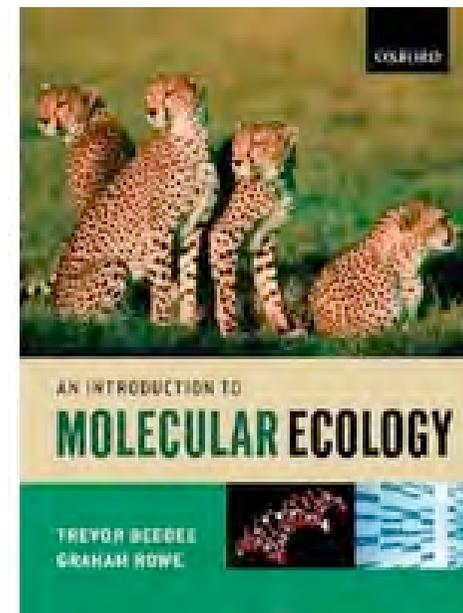
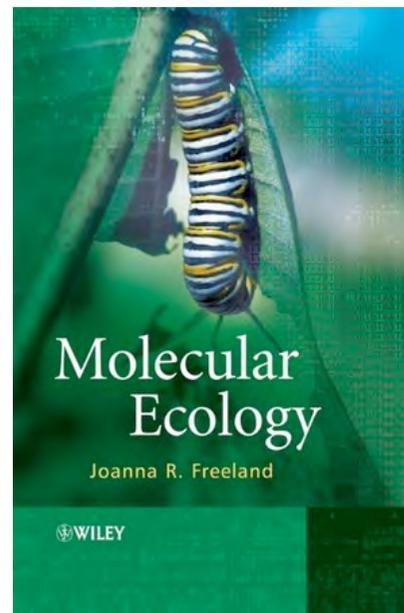
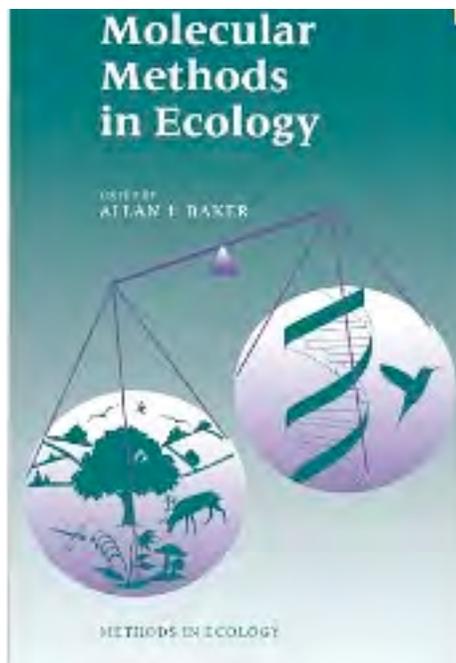
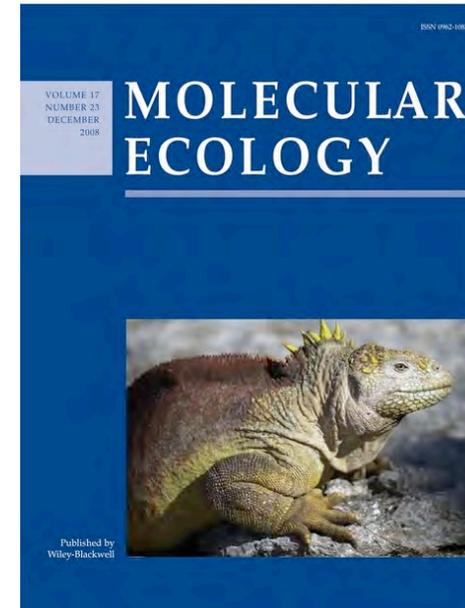
But et histoire

- Certaines techniques vont permettre de véritables bonds dans les capacités d'étude de l'ADN. C'est le cas par exemple de la PCR (Polymerase Chain Reaction).
 - 1986 Première publication publique sur la PCR par Kary Mullis (prix Nobel de chimie en 1993)
- Au cours de ces années, différents marqueurs moléculaires se succèdent, offrant toujours plus de précision et de pouvoir de résolution
- Petit à petit, ces découvertes et techniques vont sortir du cadre de l'étude des espèces modèles et intéresser de plus en plus les écologistes et de façon général, les chercheurs qui travaillent sur les populations naturelles.

1 - Introduction à l'écologie moléculaire

But et histoire

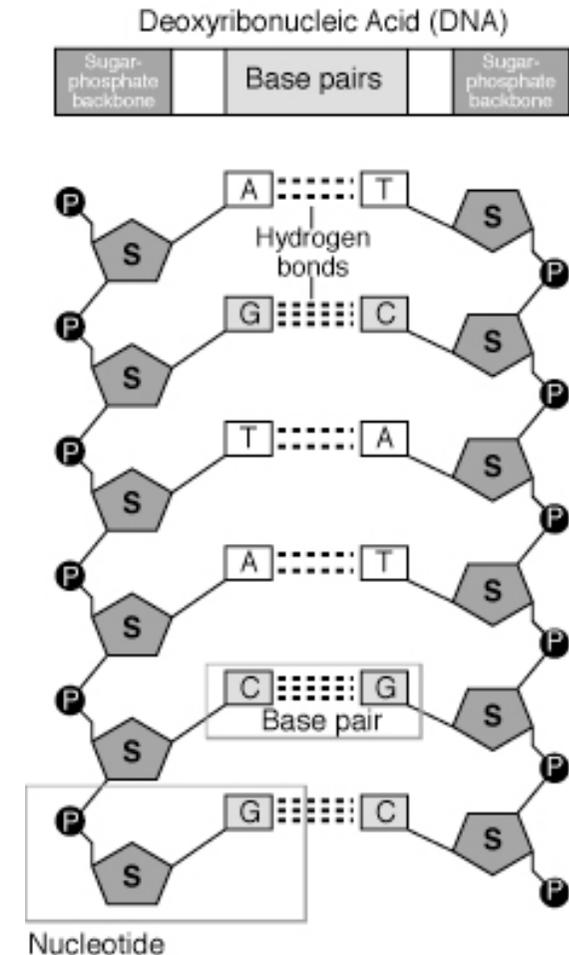
- En 1992, Publication du premier journal consacré à l'écologie moléculaire
- Suivront plusieurs ouvrages sur le sujet

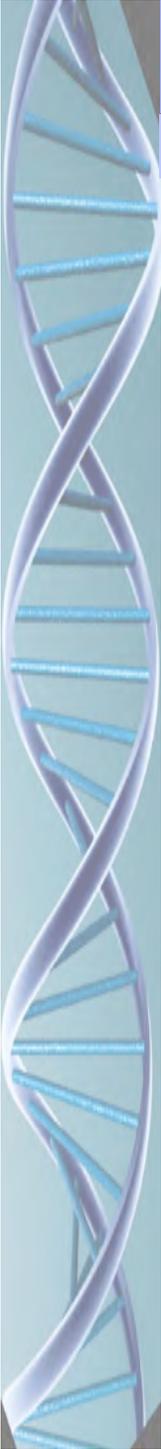


1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

- L'ADN est le support de l'information génétique
- Constitué d'un enchaînement de nucléotides, eux même composés de 3 parties:
 - 1. un groupement phosphate (ou acide phosphorique)
 - 2. un sucre à cinq atomes de carbone (désoxyribose)
 - 3. une base azotée variable en fonction du type de nucléotide
- Il y a 4 bases azotées qui peuvent s'unir 2 à 2 par complémentarité
 - Adénine <-> Thyminine
 - Guanine <-> Cytosine
- La séquence d'ADN va être transcrite en ARN
- Ces ARN vont être traduits en protéines





1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

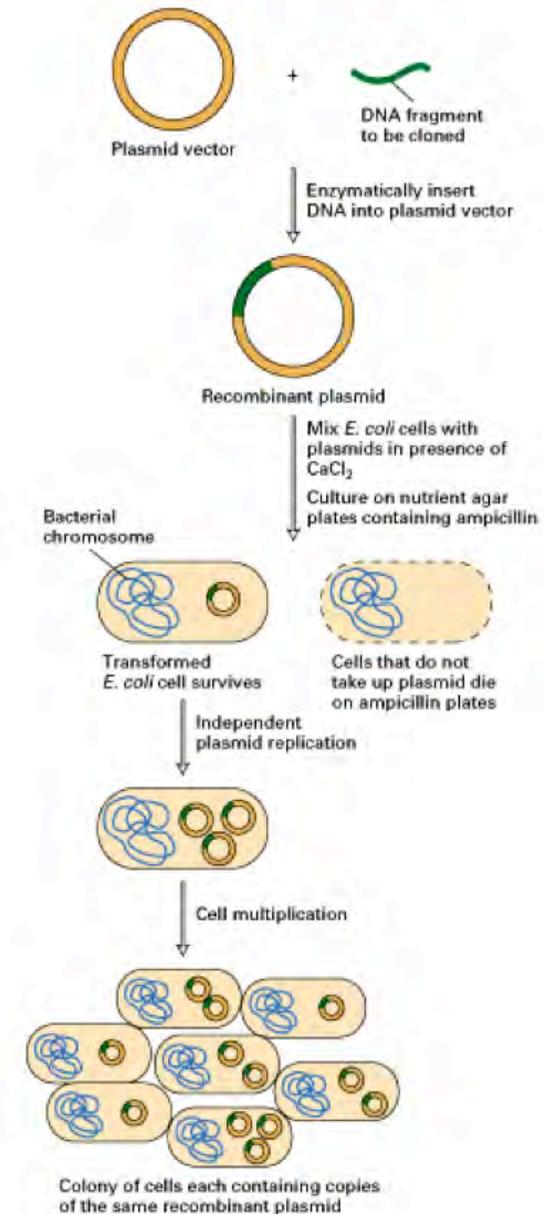
- Décrypter cette séquence d'ADN permet d'obtenir de nombreuses informations en biologie évolutive et en écologie
- Pour travailler sur un gène ou un marqueur génétique particulier, il est nécessaire d'en obtenir une quantité importante par rapport au reste du génome (bruit)
- 2 techniques importantes pour cela sont
 - Le clonage
 - La PCR

1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

Le clonage

- Utiliser un organisme capable d'autoréplication pour isoler et multiplier un fragment d'ADN particulier
- On insère le fragment d'ADN dans un vecteur (e.g. plasmide = petit ADN circulaire d'origine bactérienne)
- On transforme une bactérie avec ce plasmide recombinant
- On détecte les bactéries transformées et on attend qu'elles aient multiplié de façon clonale le fragment d'origine
- On extrait ce fragment de la colonie bactérienne





1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

La PCR

- Polymerase Chain Reaction ou amplification en chaîne par polymérase
- Véritable révolution pour l'étude de l'ADN, surtout pour l'étude de populations naturelles ou il n'est pas souvent possible d'obtenir de grandes quantités de tissus
- A l'aide d'amorces (ou primers) spécifiques, un fragment d'ADN va être multiplié grâce à une polymérase lors de cycles de variations de température
- Chaque cycle va multiplier par 2 (à peu près) la quantité de copies du gène ciblé (produit de PCR)

1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

La PCR

• A chaque cycle, l'amplification se fait en trois étapes:

- 1 - Dénaturation (95°)
- 2 - Hybridation (annealing) (+/-60°)
- 3 - Elongation (72°)

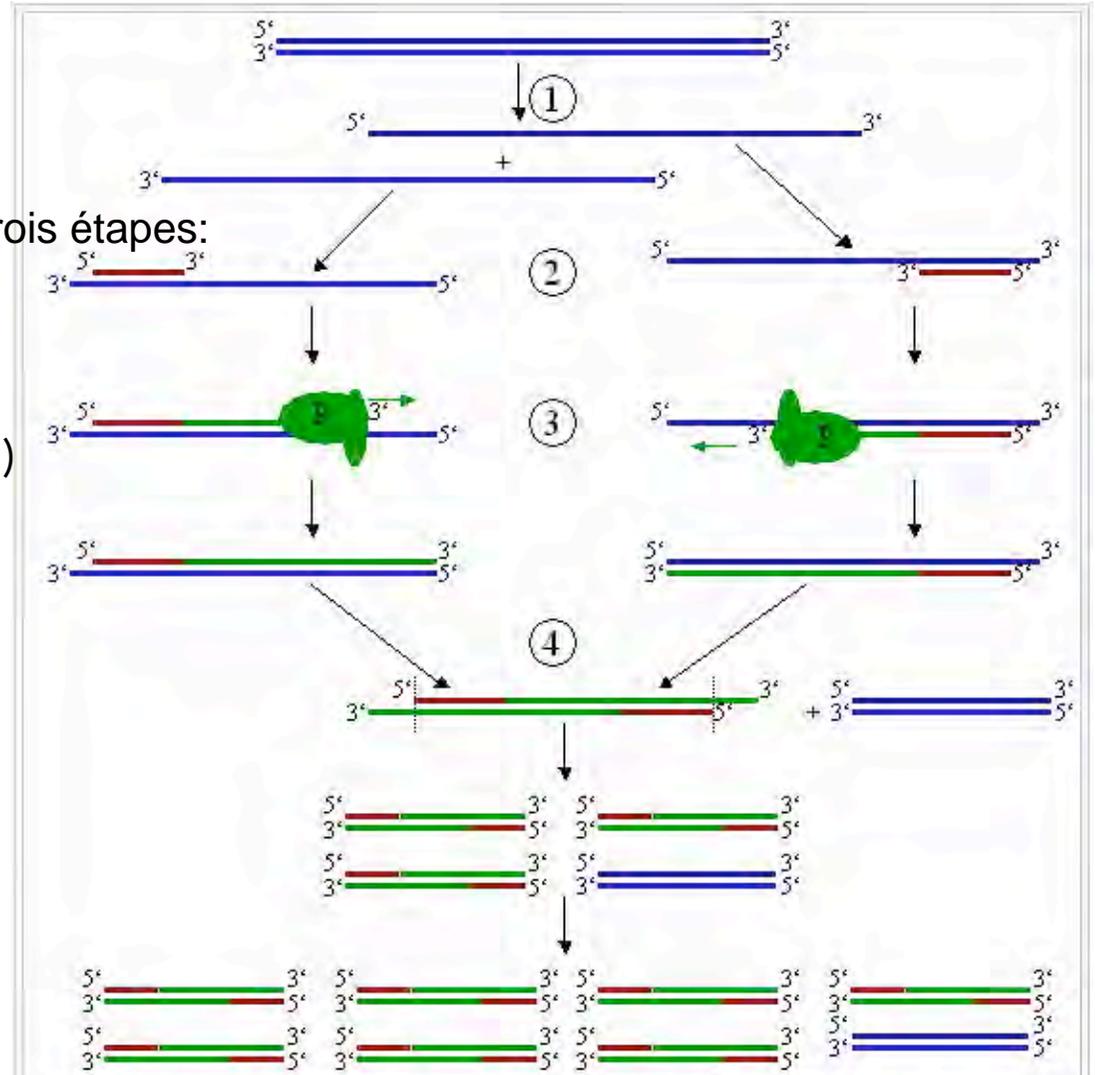


Figure 2: Schematic drawing of the PCR cycle. (1) Denaturing at 94-96°C. (2) Annealing at (eg) 68°C. (3) Elongation at 72°C (P=Polymerase). (4) The first cycle is complete. The two resulting DNA strands make up the template DNA for the next cycle, thus doubling the amount of DNA duplicated for each new cycle.

1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

La PCR

• A chaque cycle, l'amplification se fait en trois étapes:

- 1 - Dénaturation (95°)
- 2 - Hybridation (annealing) (+/-60°)
- 3 - Elongation (72°)



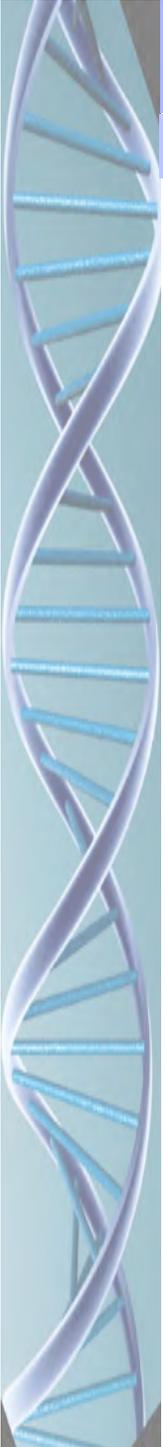
1 - Introduction à l'écologie moléculaire

Généralités et rappels de génétique moléculaire

La PCR

- A chaque cycle, l'amplification se fait en trois étapes:
 - 1 - Dénaturation (95°)
 - 2 - Hybridation (annealing) (+/-60°)
 - 3 - Elongation (72°)



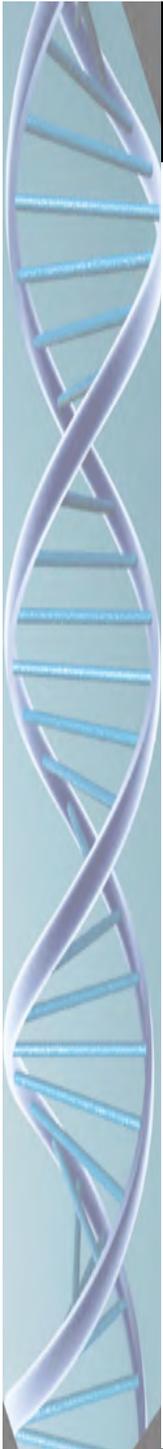


1 - Introduction à l'écologie moléculaire

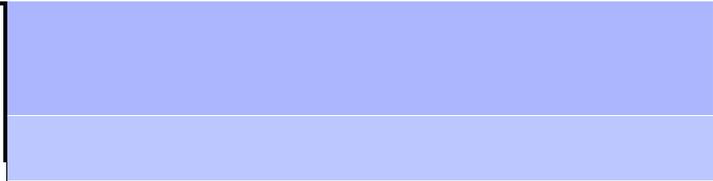
Généralités et rappels de génétique moléculaire

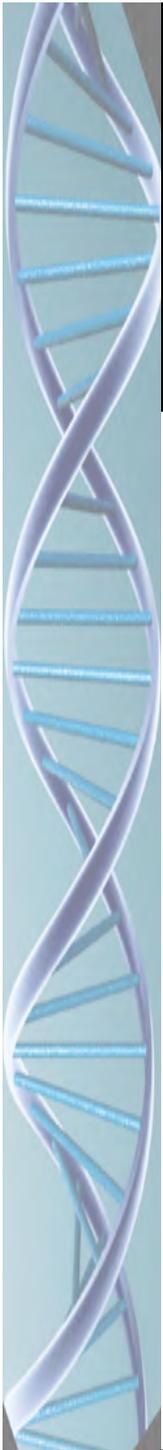
Séquençage traditionnel

- Maintenant que le fragment d'ADN d'intérêt est disponible en quantité suffisante, il va être possible d'en obtenir la séquence
- Méthode classique = séquençage 'SANGER'
- Développé dans les années 1970 par Walter Gilbert, aux États-Unis, par Frederick Sanger, en Grande-Bretagne (prix Nobel en 1980)
- Basé sur l'utilisation de didésoxyribonucléotide (ddNTP) lors d'une réaction de séquence
- Du fait de la méthode utilisée, la longueur de séquence maximale est de l'ordre de 1000 nucléotides



Template (original) DNA strand



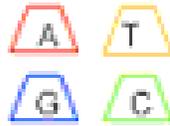


Template (original) DNA strand

5' ————— 3'
5'



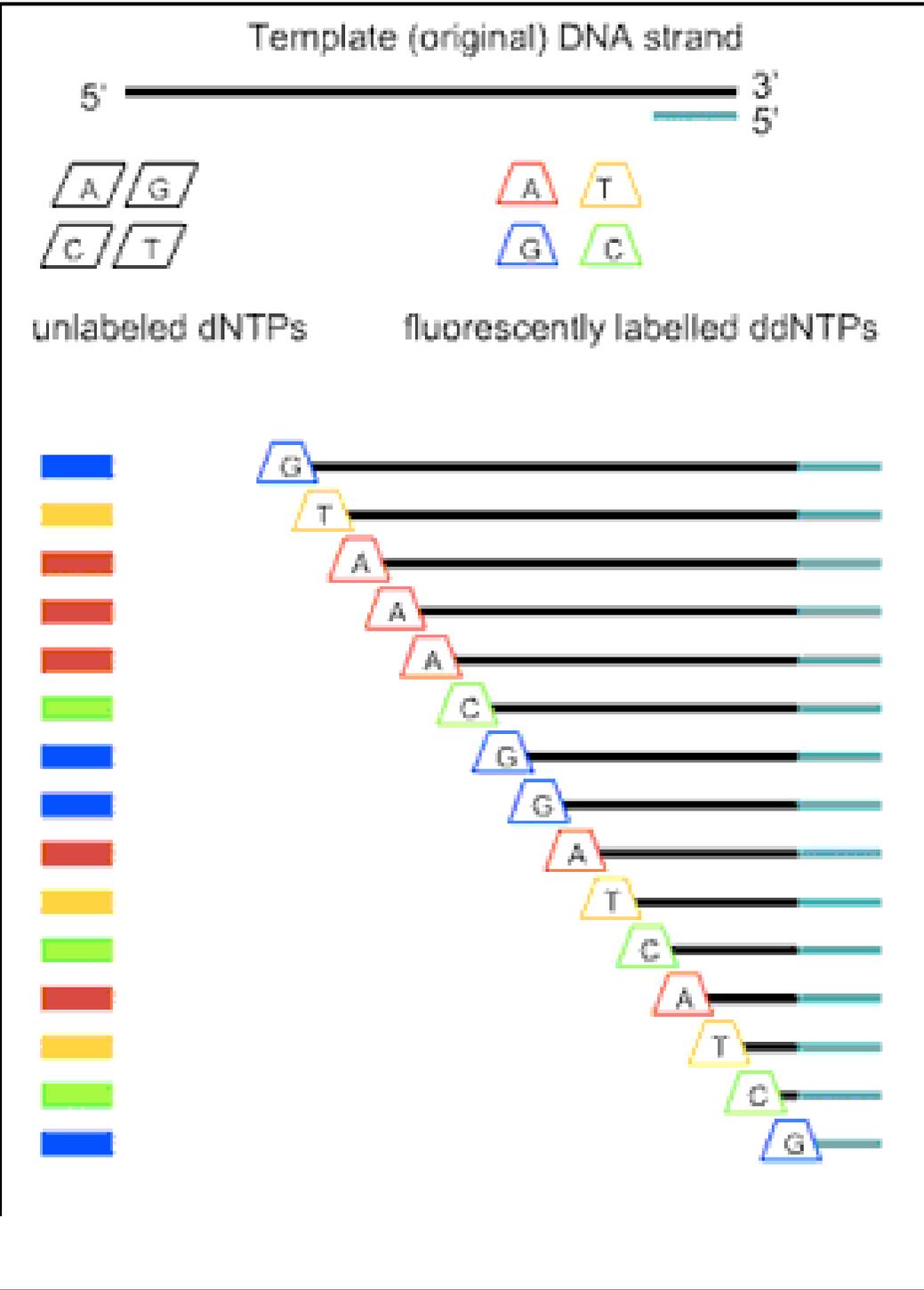
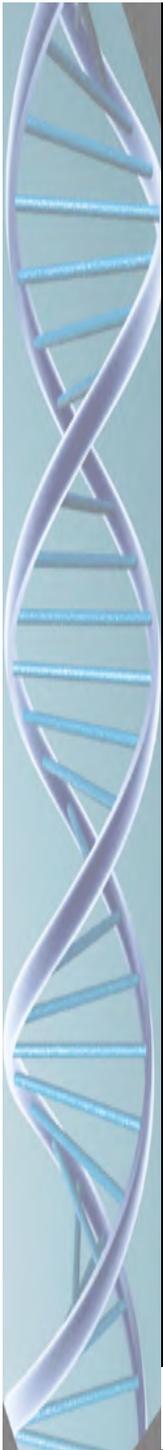
unlabeled dNTPs



fluorescently labelled ddNTPs

+ un tampon de réaction et une polymérase

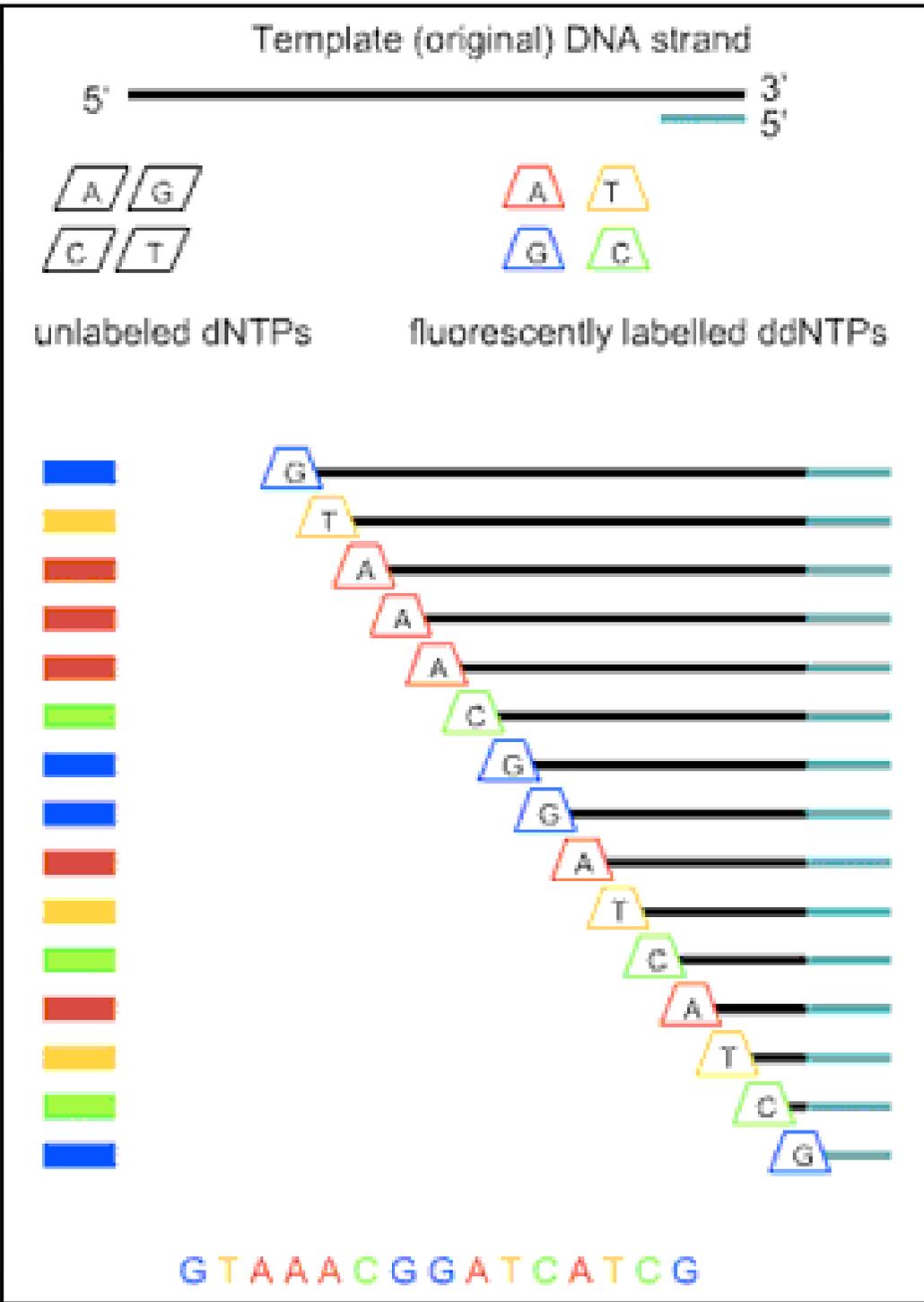
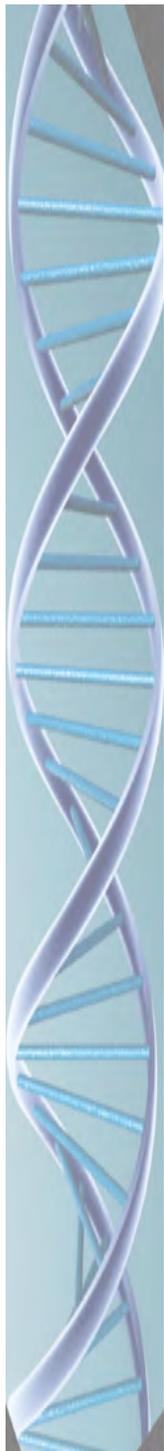
Les ddNTP sont en sous-nombre par rapport aux dNTP



+ un tampon de réaction et une polymérase

Les ddNTP sont en sous-nombre par rapport aux dNTP

On obtient un mélange de fragments aléatoirement terminés plus ou moins tôt

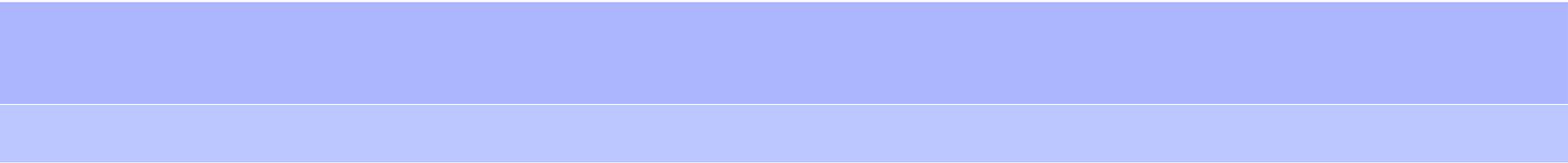
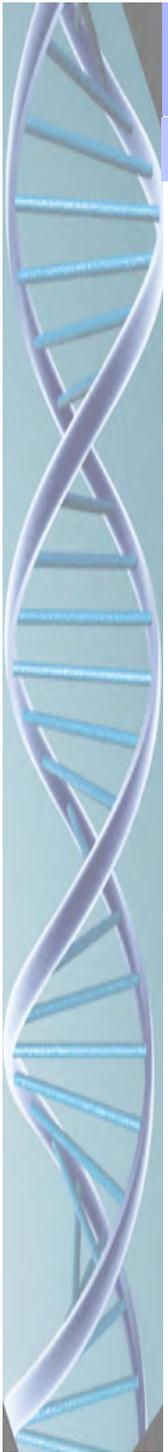


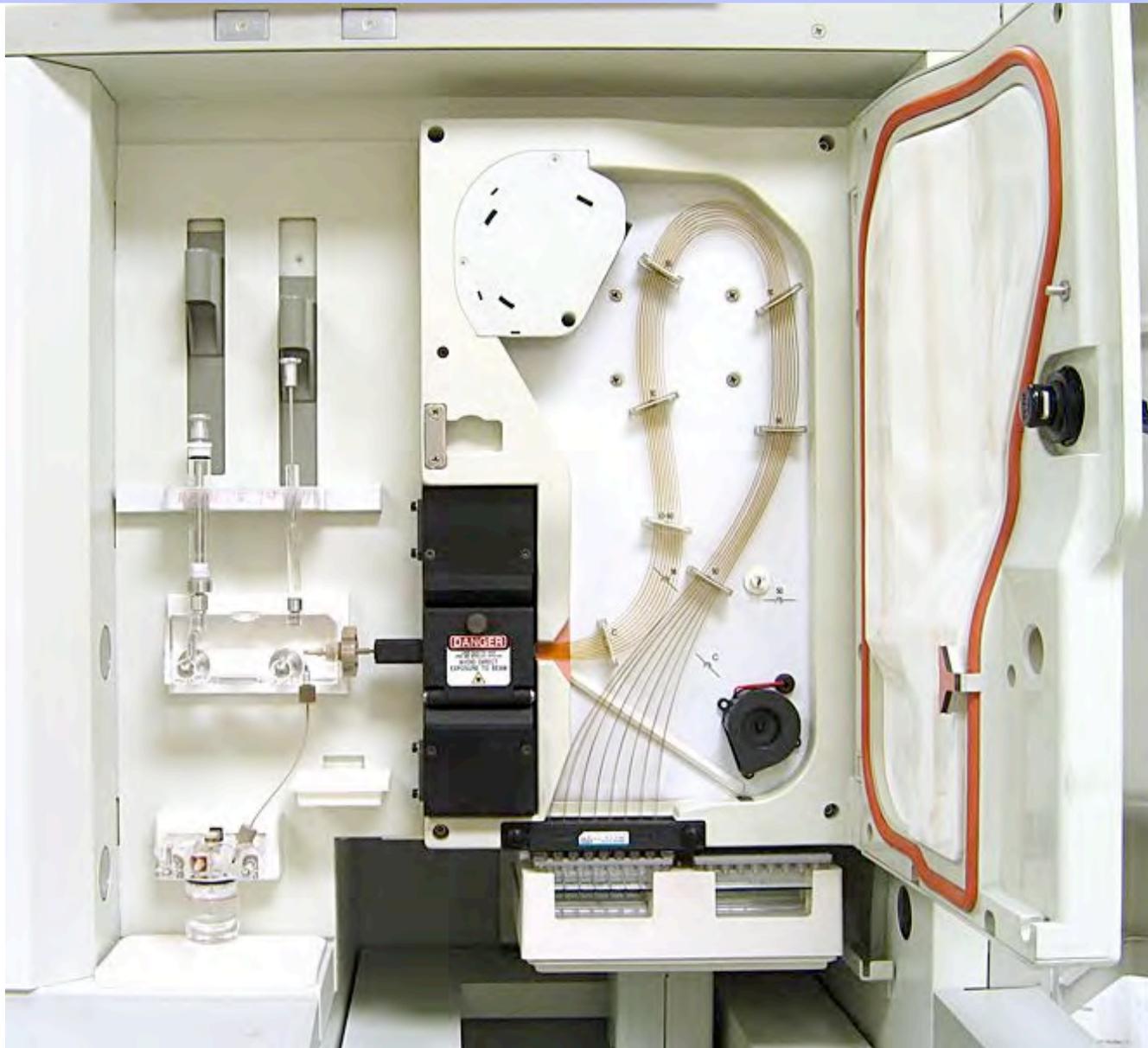
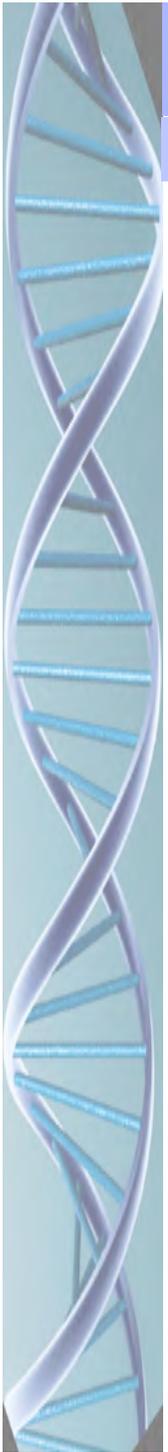
+ un tampon de réaction et une polymérase

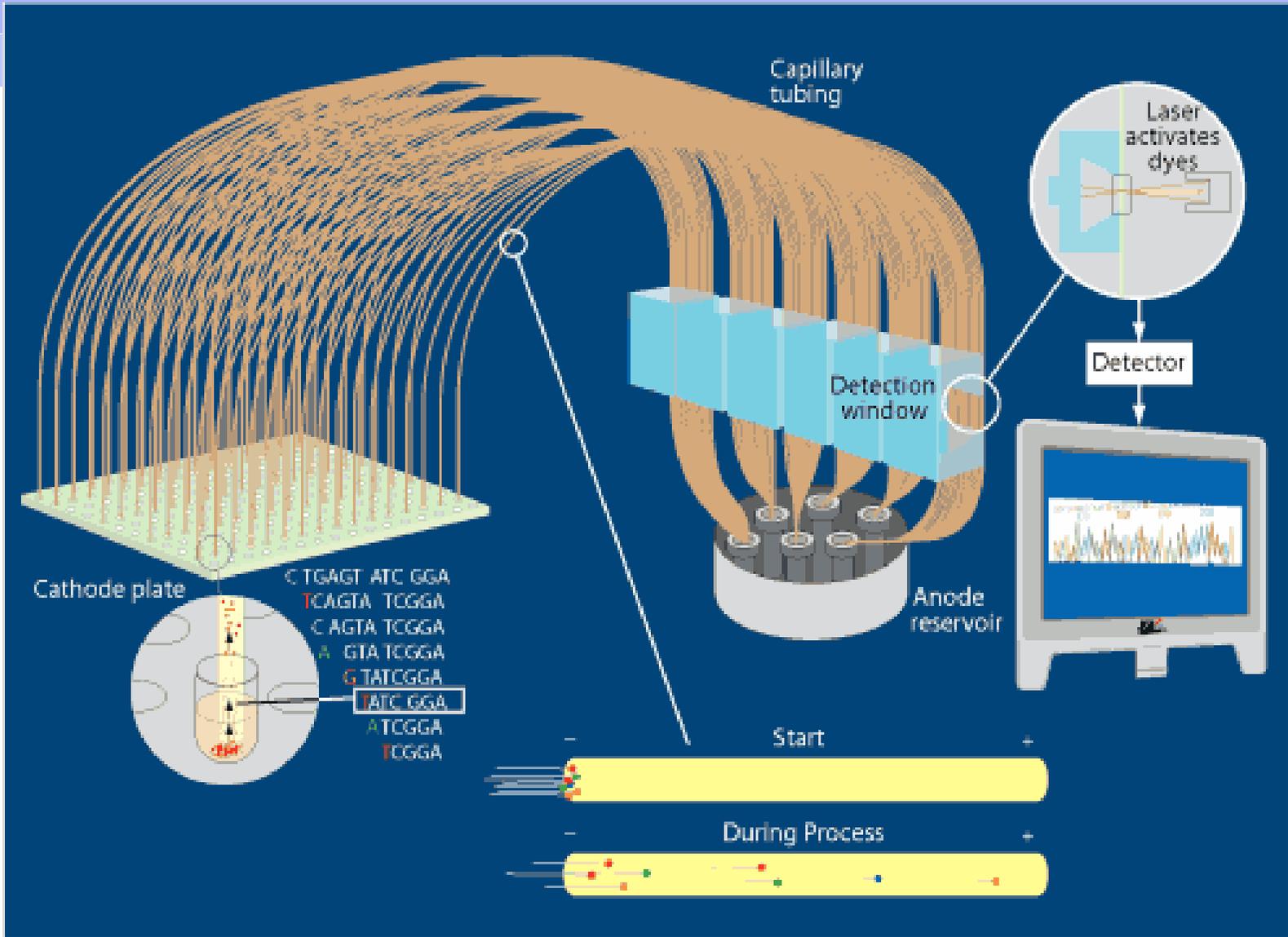
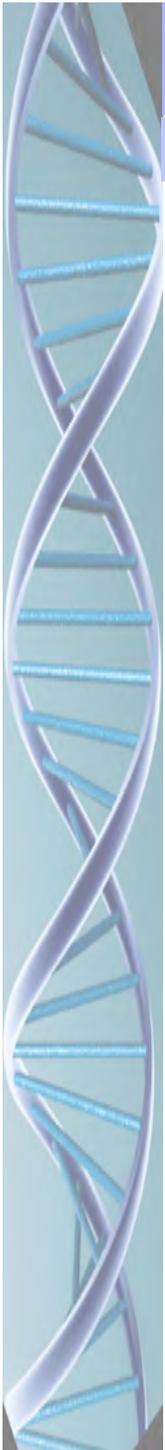
Les ddNTP sont en sous-nombre par rapport aux dNTP

On obtient un mélange de fragments aléatoirement terminés plus ou moins tôt

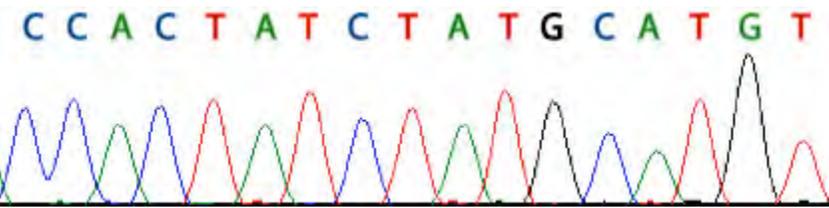
On fait migrer ce mélange par électrophorèse et on identifie l'enchaînement de signaux fluorescents

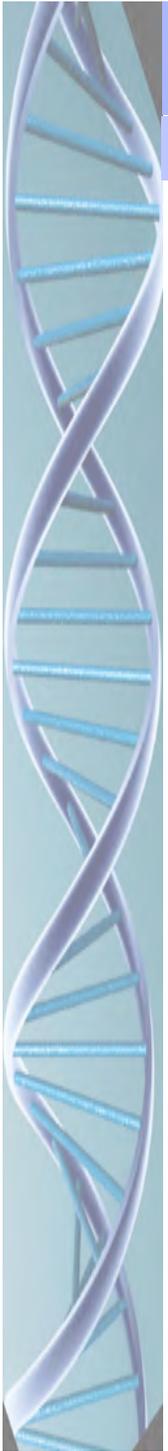






C TGAGT ATC GGA
TCAGTA TCGGA
C AGTA TCGGA
A GTA TCGGA
GTATCGGA
TATC GGA
ATCGGA
TCGGA



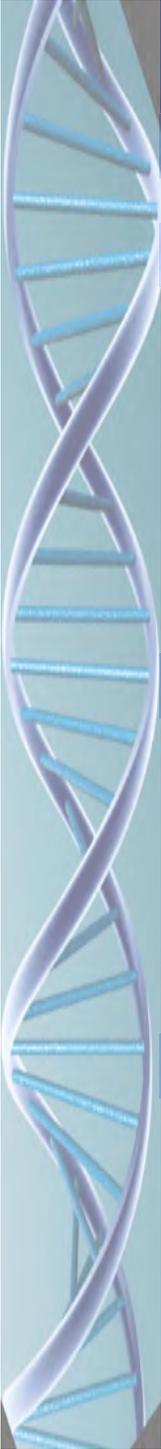


Questions

Que se passe-t-il si il y a plusieurs séquences d'ADN différentes dans le mélange?

Dans quelles situations cela peut-il se produire?

Que peut-on faire dans cette situation?



Plan

•1 - Introduction à l'écologie moléculaire

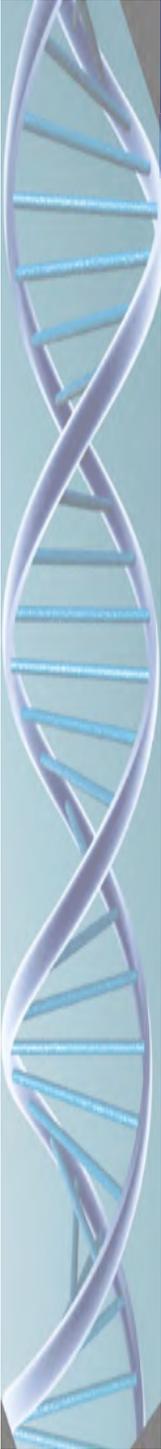
- But et histoire
- Généralités et rappels de génétique moléculaire
 - Clonage
 - PCR
 - Séquençage traditionnel

•2 - Introduction aux techniques de séquençage de nouvelle génération

- Généralités
- Plateformes principales
- Principe du pyroséquençage
- Détails de la plateforme Roche/454

•3 - Quelques applications en écologie moléculaire

- Développement de marqueurs microsatellites
- Caractérisation de régimes alimentaires
- Détermination de la biodiversité à partir d'échantillonnages environnementaux
- Détermination de la biodiversité présente dans un environnement passé



2 - Introduction au séquençage de nouvelle génération

Généralités

- Séquençage haut-débit = toute technologie capable de générer un 'grand nombre' de séquence en un temps réduit. Cela englobe aussi bien les plateformes de séquençage (Sanger) à capillaires les plus performantes que les nouvelles technologies de séquençage
- Nouvelles technologies de séquençage = next generation sequencing (NGS)
- Ces nouvelles plateformes de séquençage utilisent des techniques différentes du séquençage traditionnel (Sanger) et permettent de générer des quantités de données impossibles à égaler par des techniques traditionnelles.

2 - Introduction au séquençage de nouvelle génération

Généralités

Molecular Ecology Resources (2008) 8, 3–17

doi: 10.1111/j.1471-8286.2007.02019.x

TECHNICAL REVIEW

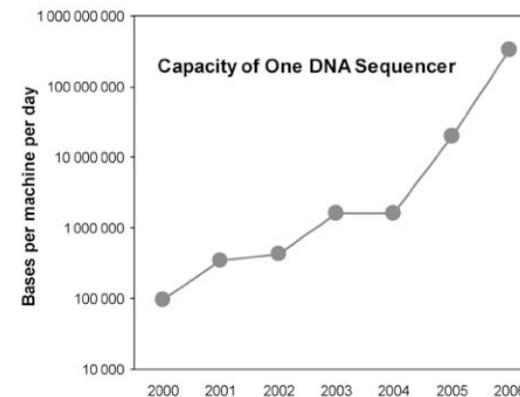
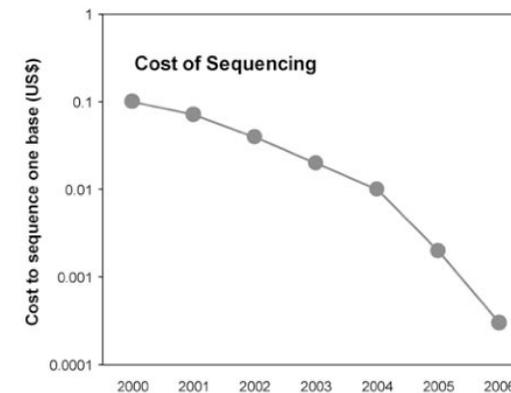
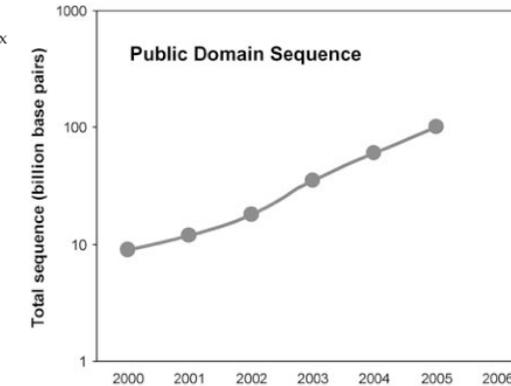
Sequencing breakthroughs for genomic ecology and evolutionary biology

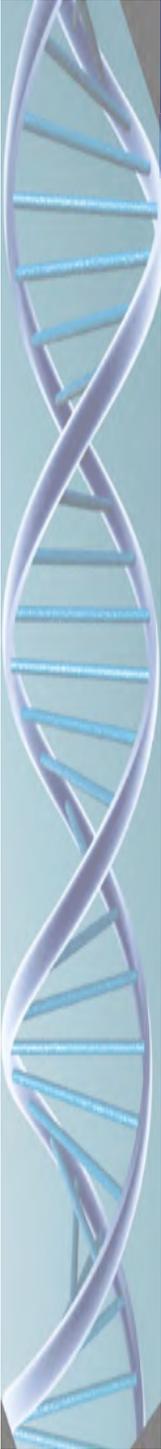
MATTHEW E. HUDSON

Department of Crop Sciences, University of Illinois, Urbana, 334 NSRC, 1101 W. Peabody Blvd., IL 61801, USA

Fig. 1 Advances in DNA sequencing technology and their effect. Upper Graph: the increasing amount of DNA sequence data in the public domain (total deposited in International Nucleotide Sequence Databases) from 2000 to 2005. Centre Graph: the cost per base of DNA sequenced for genome sequencing projects from 2000 to 2006. Lower Graph: the capacity of the highest-throughput commercially available DNA sequencer from 2000 to 2006. Note that the 454/Roche sequencing technology was introduced in 2005 and the Solexa/Illumina 1G sequencing technology in 2006, and that the 2005 publicly available sequence was essentially all generated using Sanger technology. Y axis scales are logarithmic.

Explosion de la quantité de séquences générées ces 10 dernières années!





2 - Introduction au séquençage de nouvelle génération

Plateformes principales

- Il existe 3 plateformes commerciales principales
 - GS FLX (Roche/454): 2004
 - Illumina Solexa Genome Analyzer: 2006
 - Applied Biosystems SOLiD: 2007

2 - Introduction au séquençage de nouvelle génération

Plateformes principales



GS FLX titanium (Roche/454)

- 400 millions de bases par run
- Séquence longue, 200 - 500 bases
- Jusque 1.3 million de séquences par run

2 - Introduction au séquençage de nouvelle génération

Plateformes principales



SOLID 4hq (ABI)

- 300 milliards de bases par run
- Séquences courtes, 75 bases

2 - Introduction au séquençage de nouvelle génération

Plateformes principales



Genome analyzer IIx (illumina)

- 100 milliards de bases par run
- Séquences moyennes, 2x100 bases
- Jusque 500 millions de séquences par run

2 - Introduction au séquençage de nouvelle génération

Principe du pyroséquençage

- Basé sur un principe de séquençage par synthèse, par opposition au séquençage par 'termination' (méthode Sanger)
- Cette méthode consiste à séquencer un ADN monobrin par synthèse du brin complémentaire, base par base, en détectant à chaque étape le nucléotide qui a été ajouté



2 - Introduction au séquençage de nouvelle génération

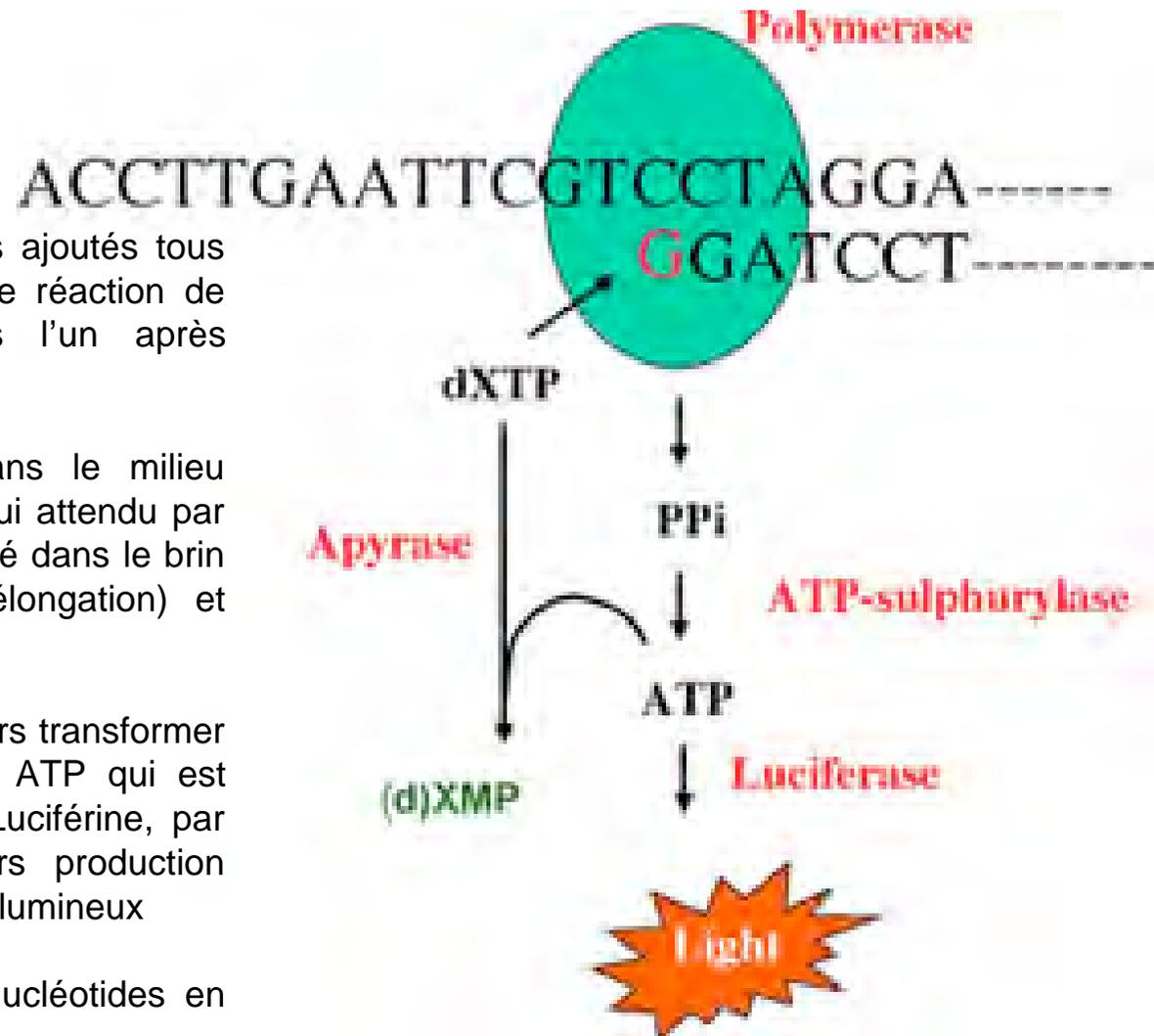
Principe du pyroséquençage

Les nucléotides ne sont pas ajoutés tous ensemble comme dans une réaction de séquençage normale mais l'un après l'autre

Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par la polymérase, il est incorporé dans le brin en cours de synthèse (d'élongation) et libère un Pyrophosphate

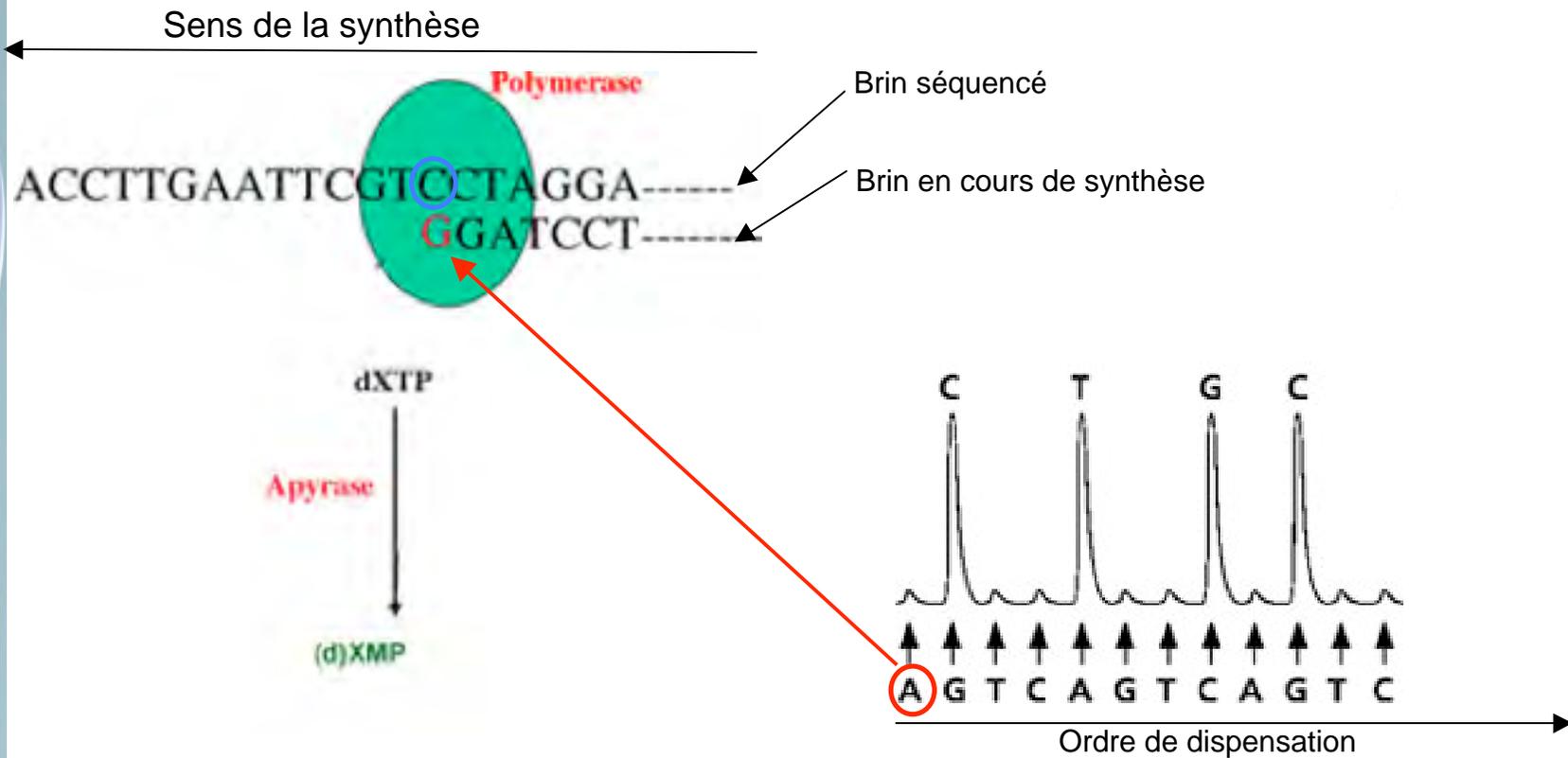
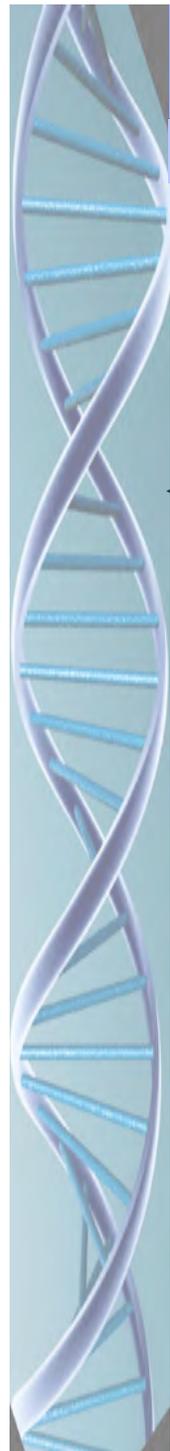
Une ATPsulfurylase vient alors transformer ce Pyrophosphate (PPi) en ATP qui est alors utilisé, couplé à une Luciférine, par une Luciférase. On a alors production d'Oxyluciférine et d'un signal lumineux

Une Apyrase dégrade les nucléotides en surplus.



2 - Introduction au séquençage de nouvelle génération

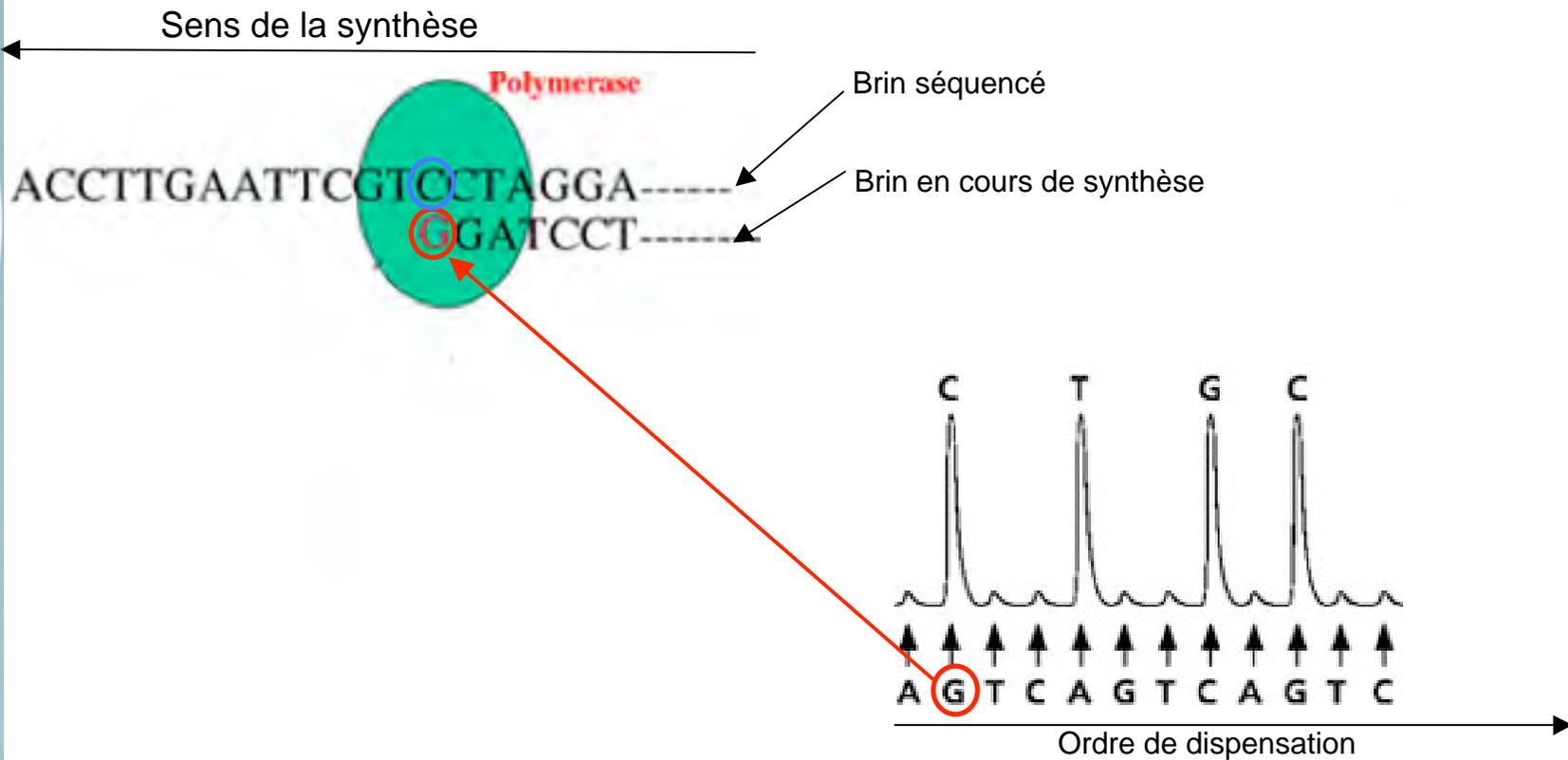
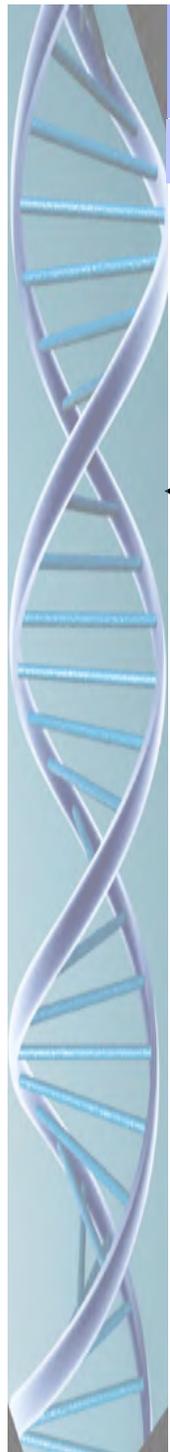
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

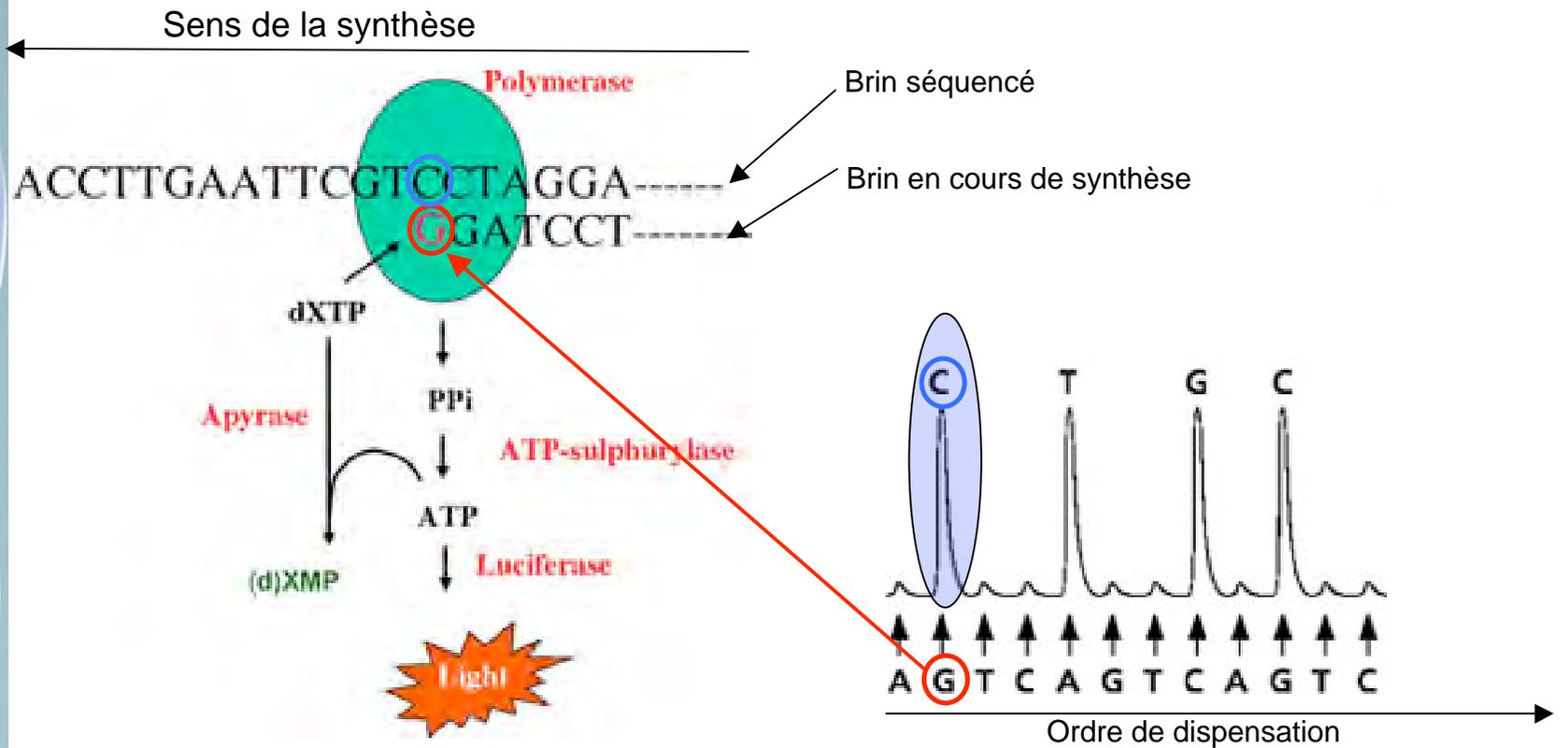
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

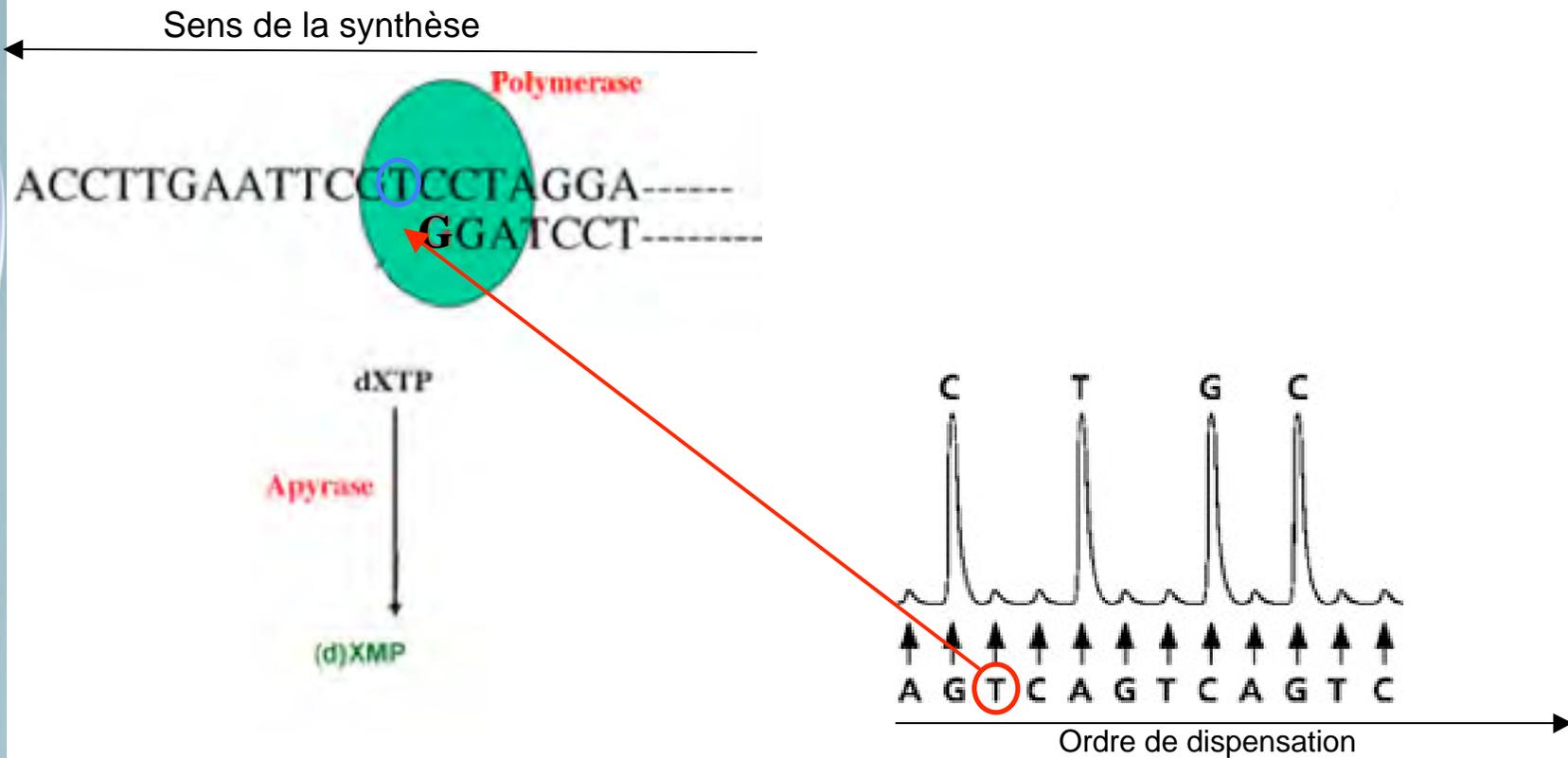
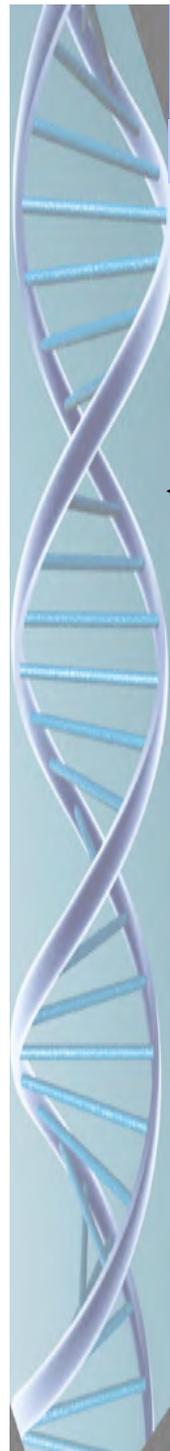
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

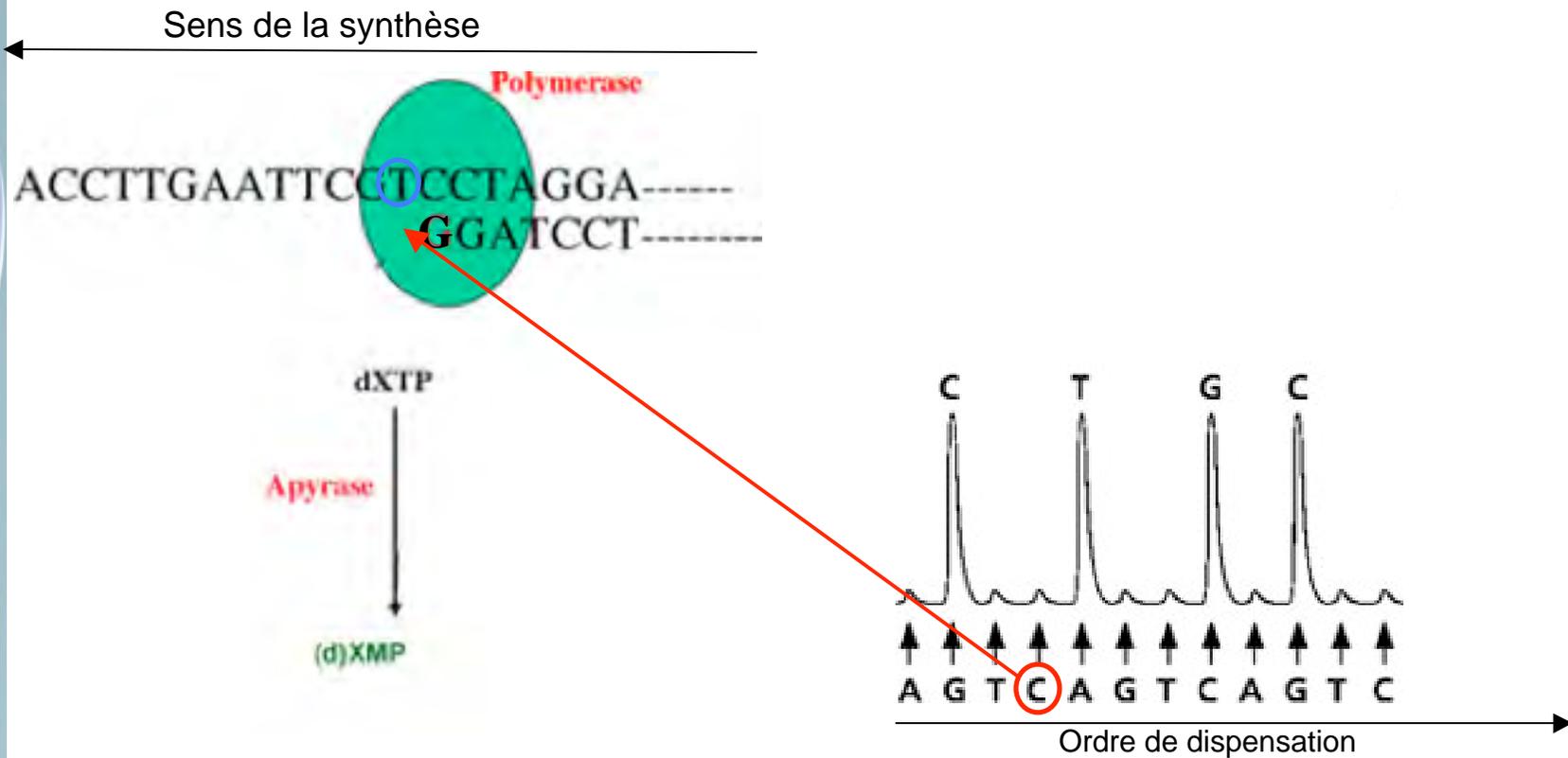
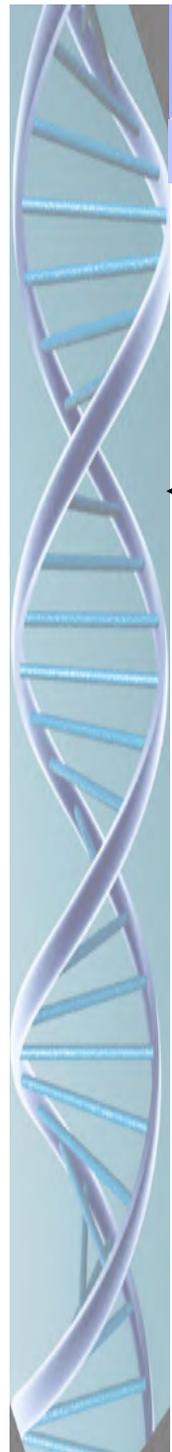
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

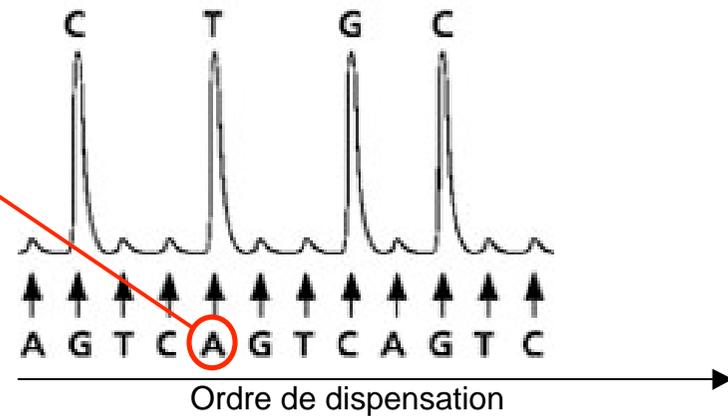
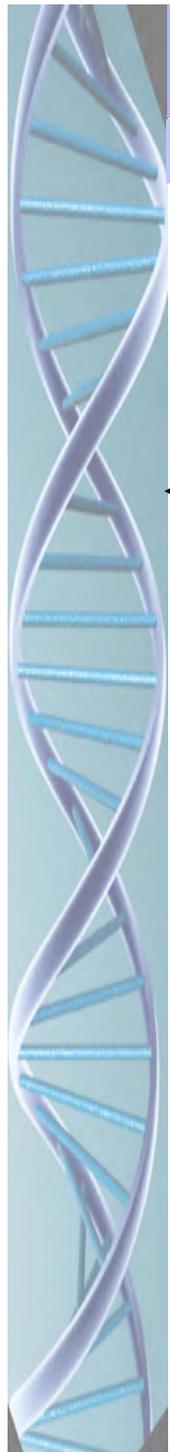
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

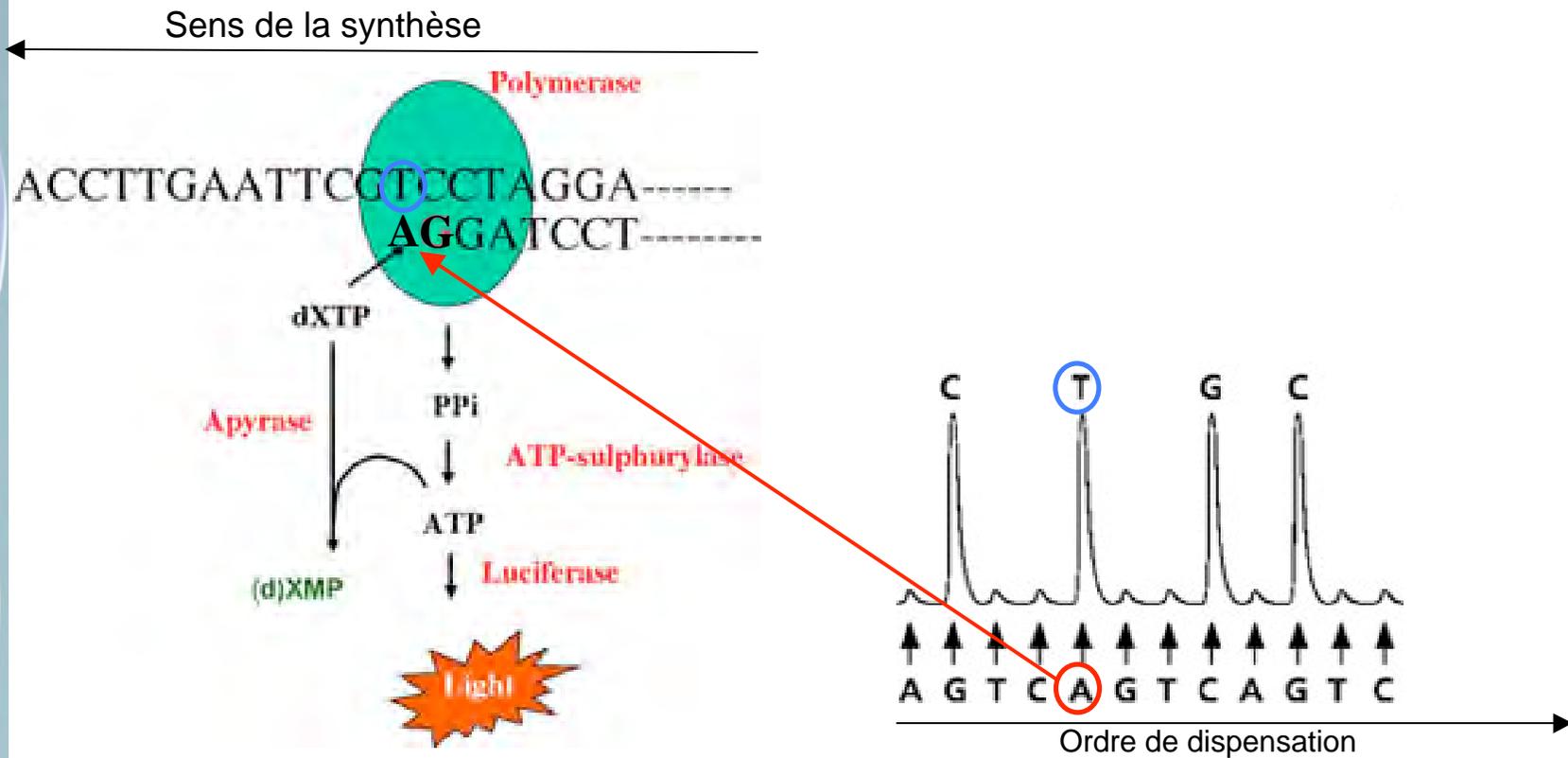
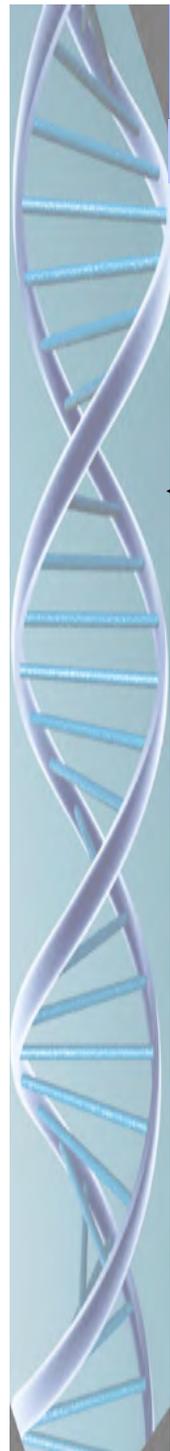
Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

Principe du pyroséquençage



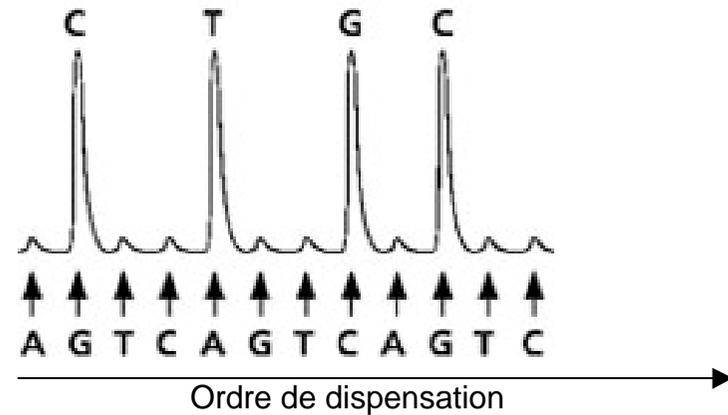
Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

Principe du pyroséquençage



Sens de la synthèse



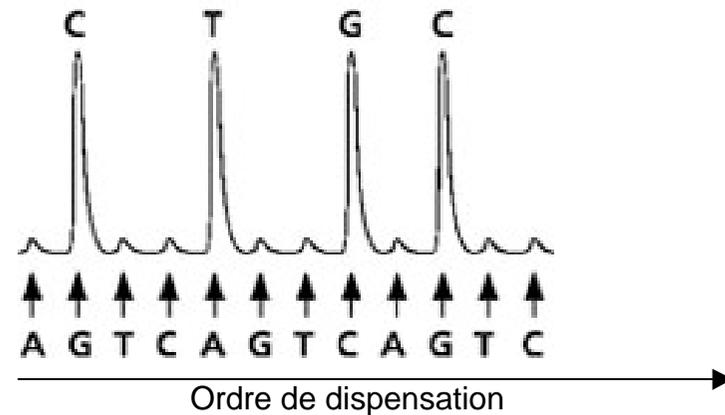
Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

Principe du pyroséquençage



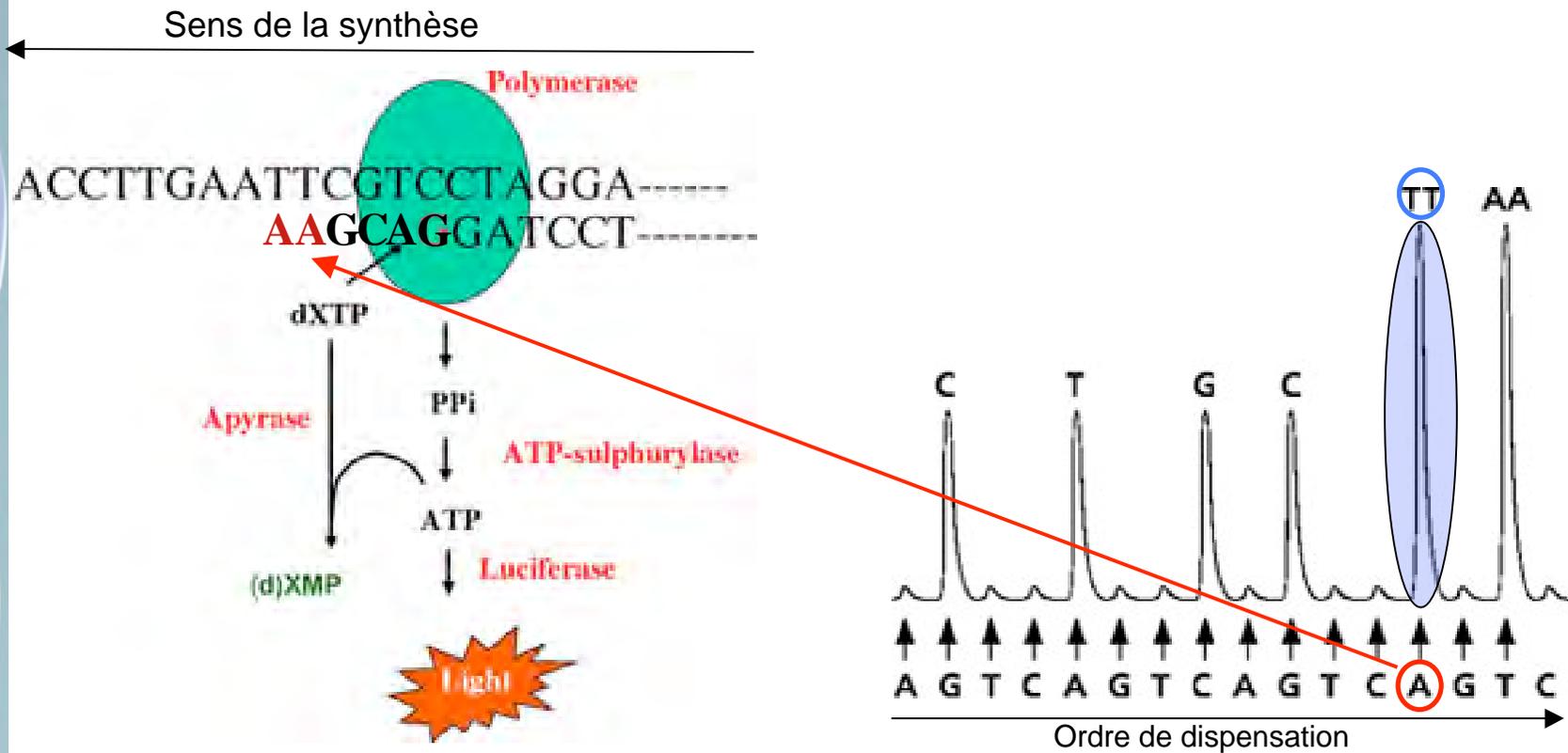
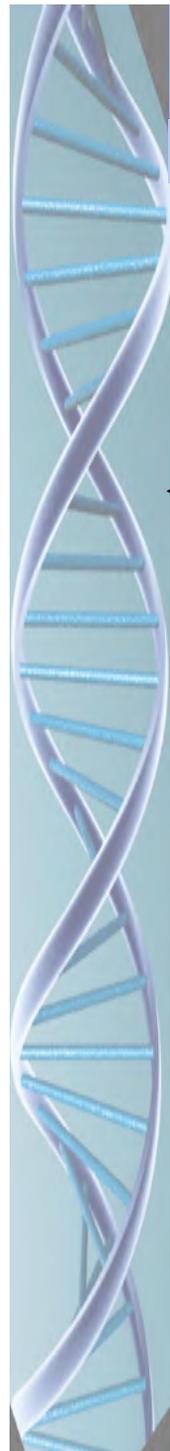
Sens de la synthèse



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

Principe du pyroséquençage



Une dispensation cyclique (A, G, T, C, A, G, T, C, ...) est utilisée

2 - Introduction au séquençage de nouvelle génération

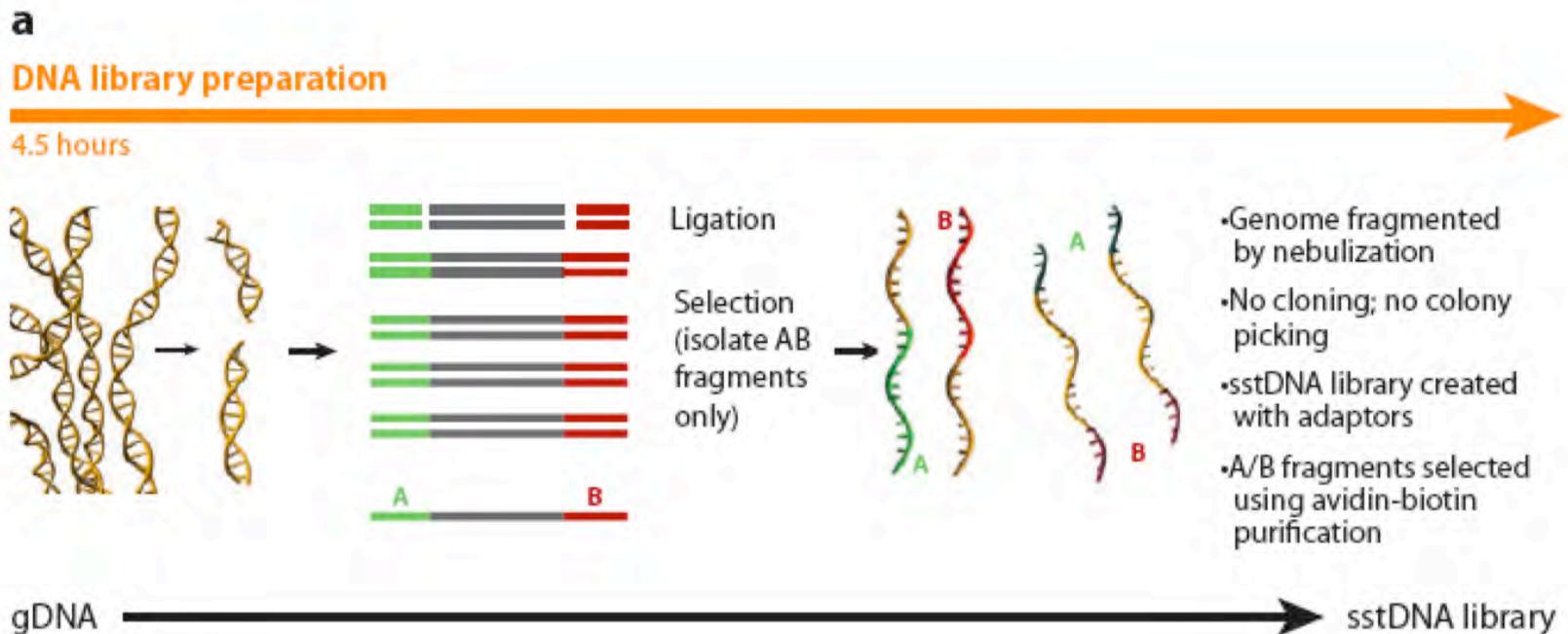
Détails de la plateforme Roche/454



2 - Introduction au séquençage de nouvelle génération

Détails de la plateforme Roche/454

Préparation de la librairie = collection de fragments d'ADN à séquencer



2 - Introduction au séquençage de nouvelle génération

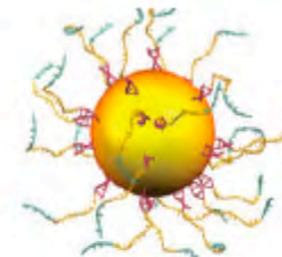
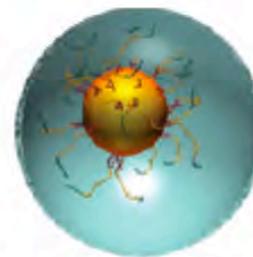
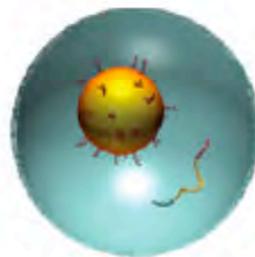
Détails de la plateforme Roche/454

- Amplification de l'ADN à séquencer: PCR en émulsion
- Un brin d'ADN par bille
- Une bille par goutte d'émulsion

b

Emulsion PCR

8 hours



Anneal sstDNA to an excess of DNA capture beads

Emulsify beads and PCR reagents in water-in-oil microreactors

Clonal amplification occurs inside microreactors

Break microreactors and enrich for DNA-positive beads

sstDNA library

Bead-amplified sstDNA library

2 - Introduction au séquençage de nouvelle génération

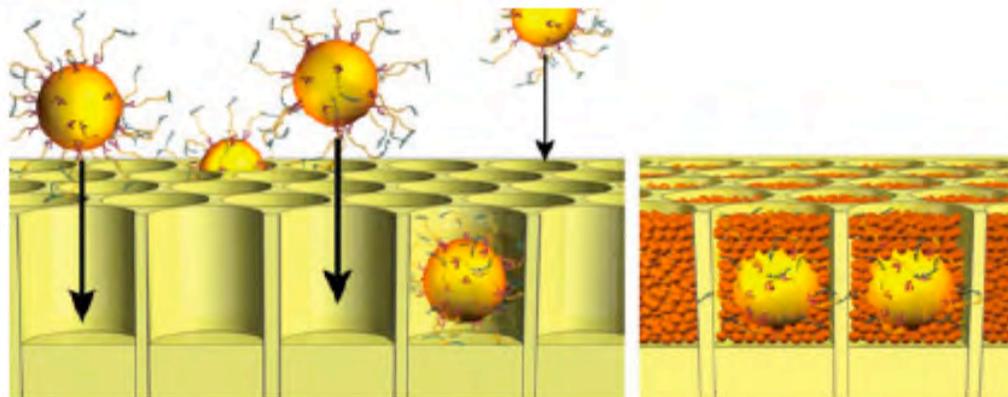
Détails de la plateforme Roche/454

- Séquençage dans une plaque à micropuits (400 000 par plaque)
- Une bille par puit

C

Sequencing

7.5 hours



- Well diameter: average of 44 μm
- 400,000 reads obtained in parallel
- A single cloned amplified sstDNA bead is deposited per well

2 - Introduction au séquençage de nouvelle génération

Détails de la plateforme Roche/454

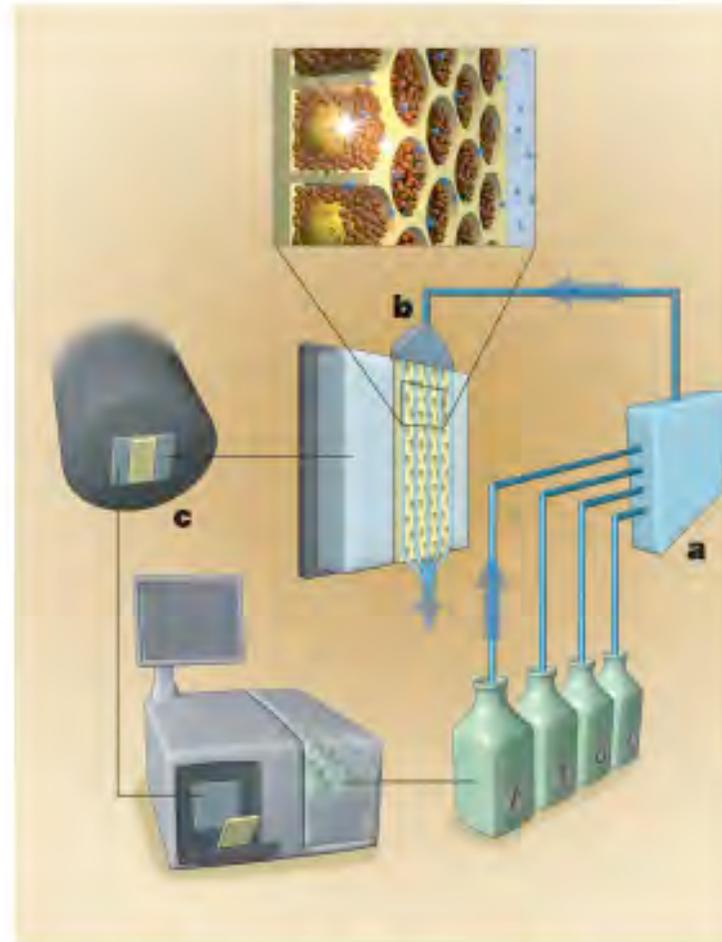
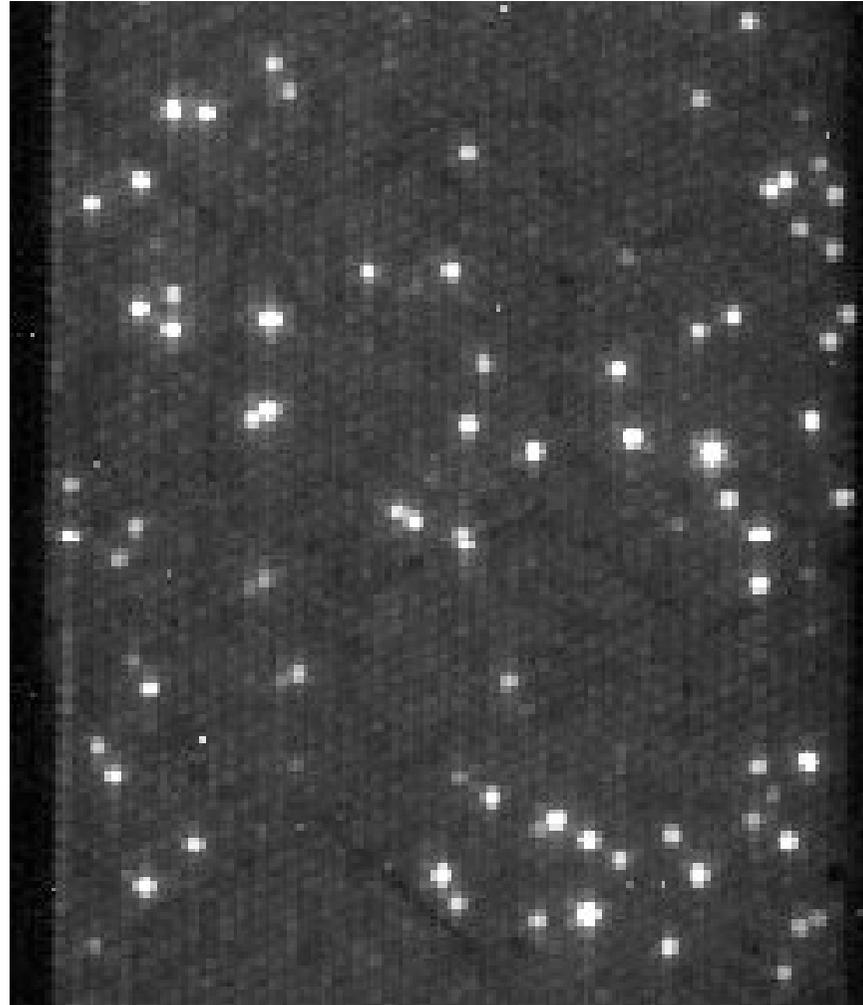
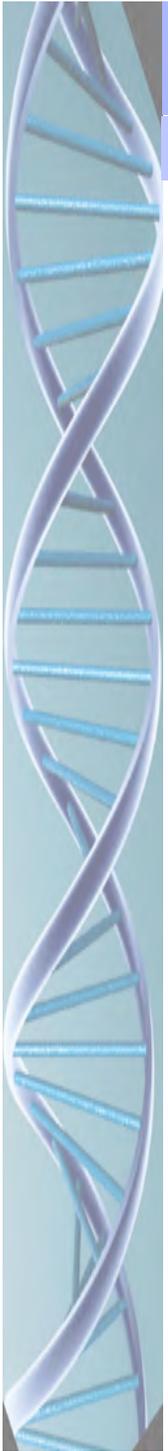


Figure 2 | Sequencing instrument. The sequencing instrument consists of the following major subsystems: a fluidic assembly (a), a flow chamber that includes the well-containing fibre-optic slide (b), a CCD camera-based imaging assembly (c), and a computer that provides the necessary user interface and instrument control.

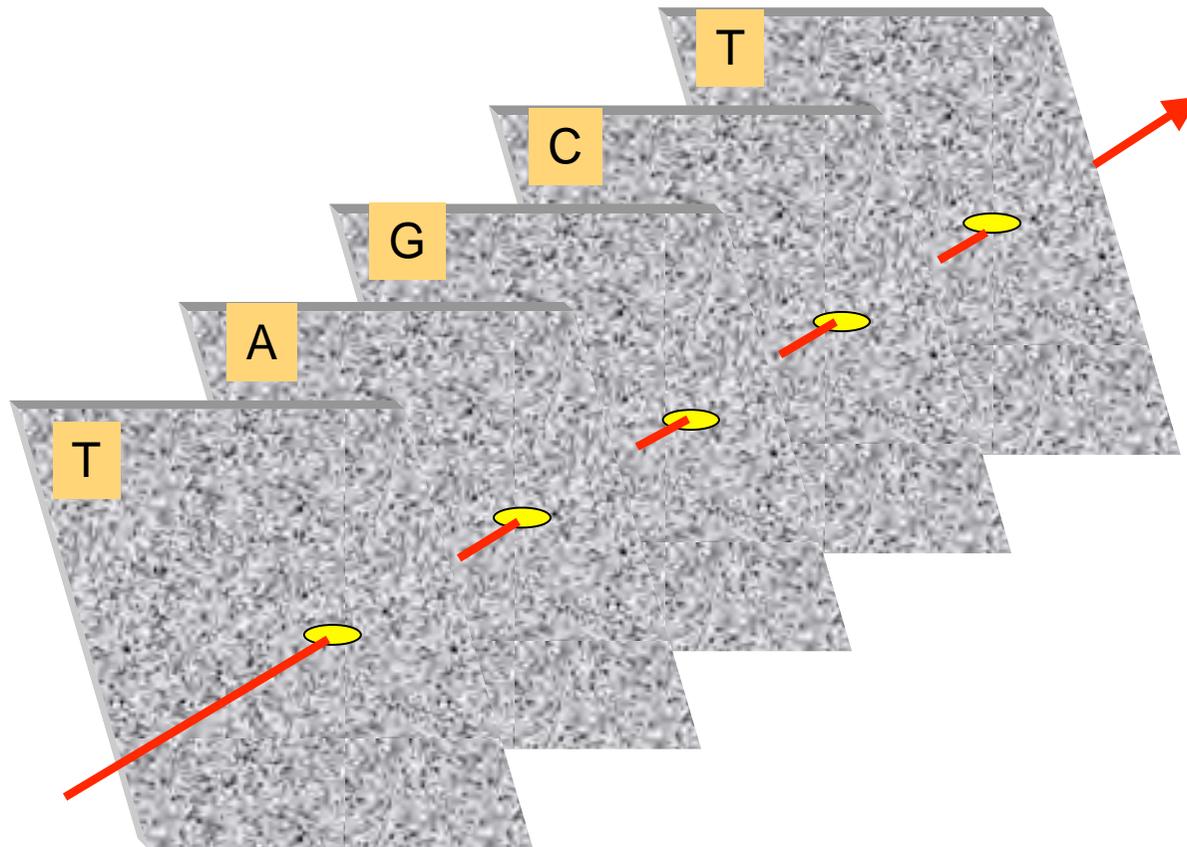
2 - Introduction au séquençage de nouvelle génération

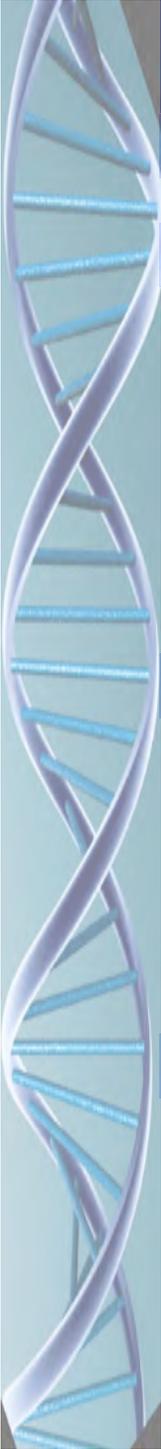
Détails de la plateforme Roche/454



2 - Introduction au séquençage de nouvelle génération

Détails de la plateforme Roche/454





Plan

•1 - Introduction à l'écologie moléculaire

- But et histoire
- Généralités et rappels de génétique moléculaire
 - Clonage
 - PCR
 - Séquençage traditionnel

•2 - Introduction aux techniques de séquençage de nouvelle génération

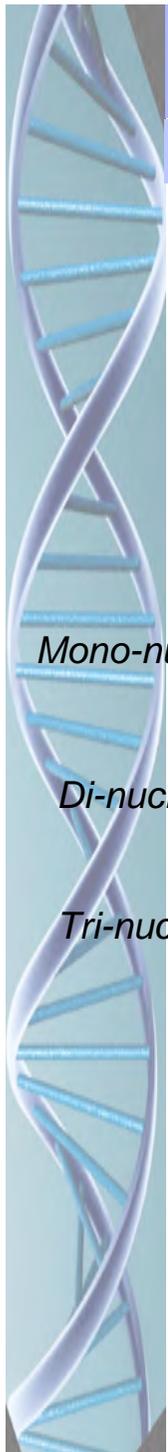
- Généralités
- Plateformes principales
- Principe du pyroséquençage
- Détails de la plateforme Roche/454

•3 - Quelques applications en écologie moléculaire

- Développement de marqueurs microsatellites
- Caractérisation de régimes alimentaires
- Détermination de la biodiversité à partir d'échantillonnages environnementaux
- Détermination de la biodiversité présente dans un environnement passé

3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites



Mono-nucleotide 5'

...AGCAGTACGT.....**AAAAAAAAAAAA**.....AGGTACCAC... 3'
(A)₁₂

Di-nucleotide 5'

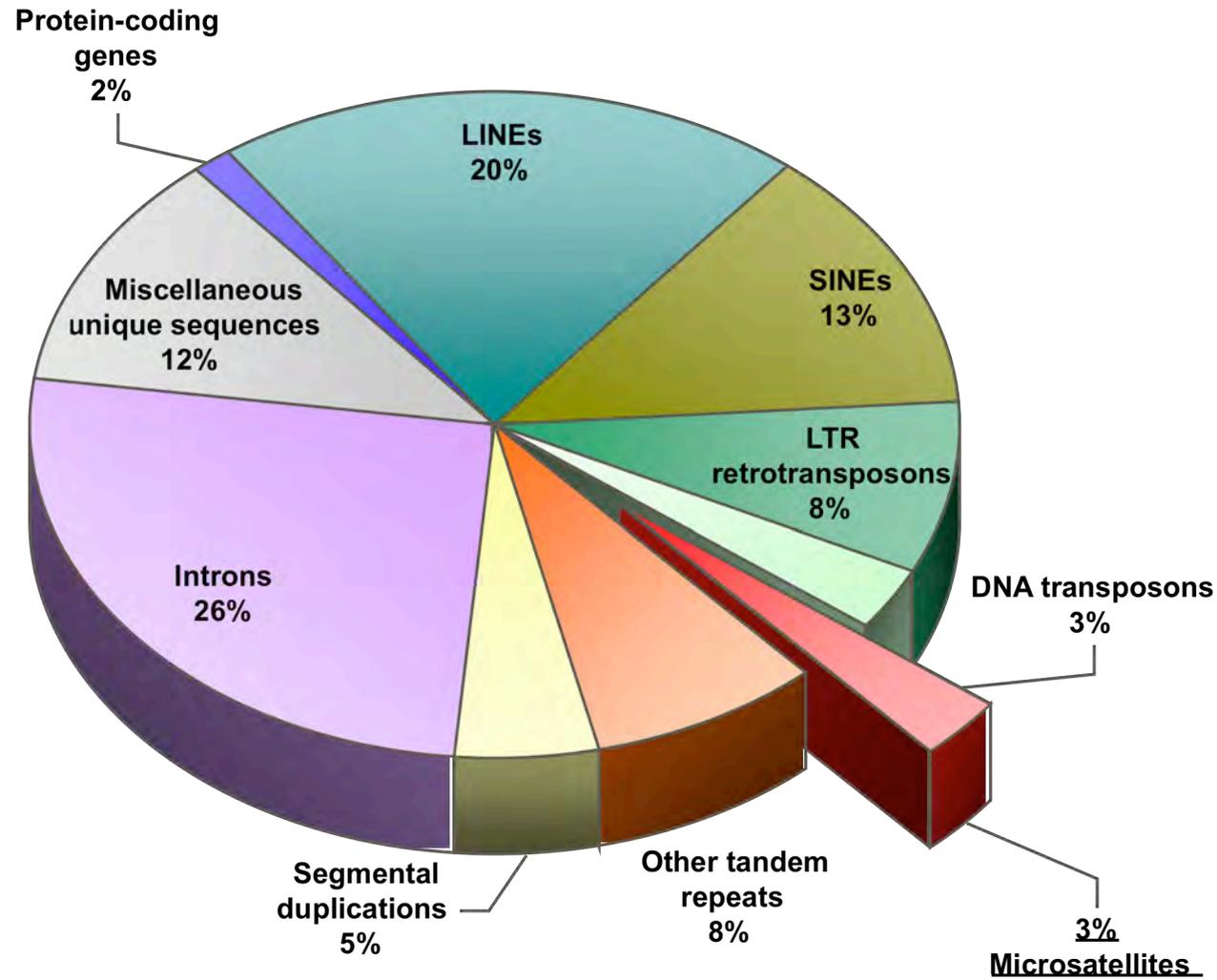
...AGCAGTACGT.....**GTGTGTGTGTGTGT**.....AGGTACCAC... 3'
(GT)₇

Tri-nucleotide 5'

...AGCAGTACGT.....**GGCGGCGGCGGC**.....AGGTACCAC... 3'
(GGC)₄

3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites



3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites

Transmission Mendélienne classique
Fort taux de mutation
Multi-allelique
Co-dominant
Polymorphisme de taille facilement déterminé par PCR

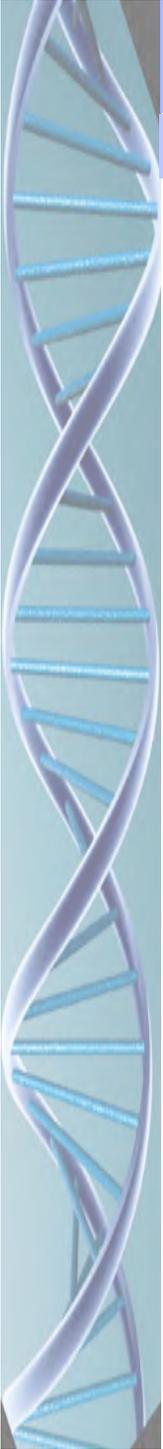


Fort polymorphisme
Interprétation simple et directe des génotypes observés



Marqueurs de choix en écologie moléculaire





3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites

Permettent de déterminer les relations biologiques à différents niveaux:

- Structure génétique à une échelle large (phylogeographie)
- Structure génétique à une échelle fine (génétique des pop)
- Niveau individuel (parenté, médico-légal...)

3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites

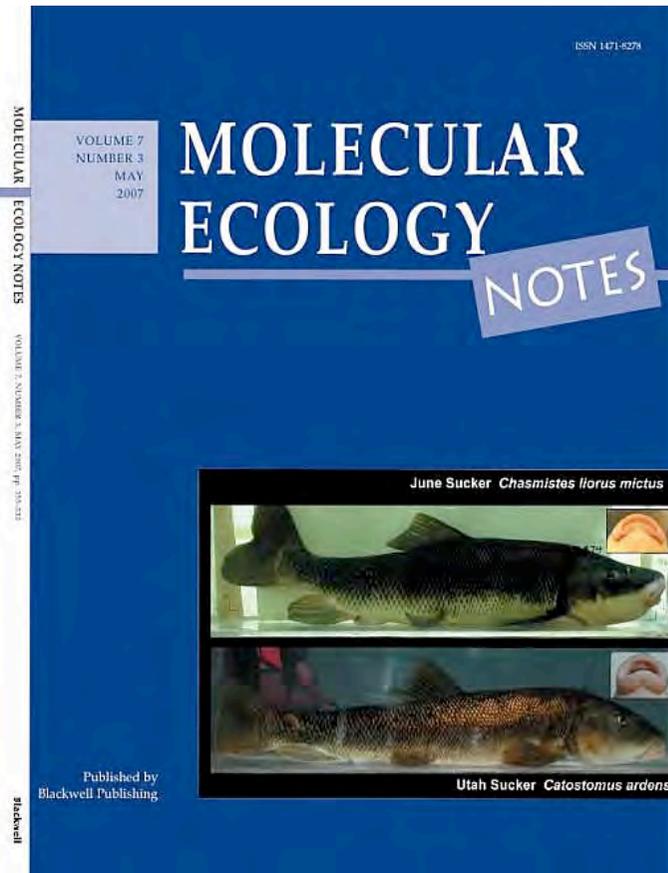


Débute en 1992

Primer note = article présentant le développement de nouveaux marqueurs pour une espèce donnée

3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites



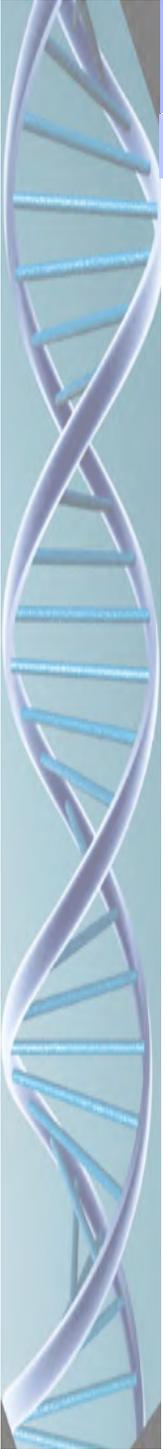
Débute en 2001

tous les 3-4mois
150 pages
30-40 primer notes

Aujourd'hui

tous les 2 mois
400 pages
100+ primer notes

Primer note = article présentant le développement de nouveaux marqueurs pour une espèce donnée

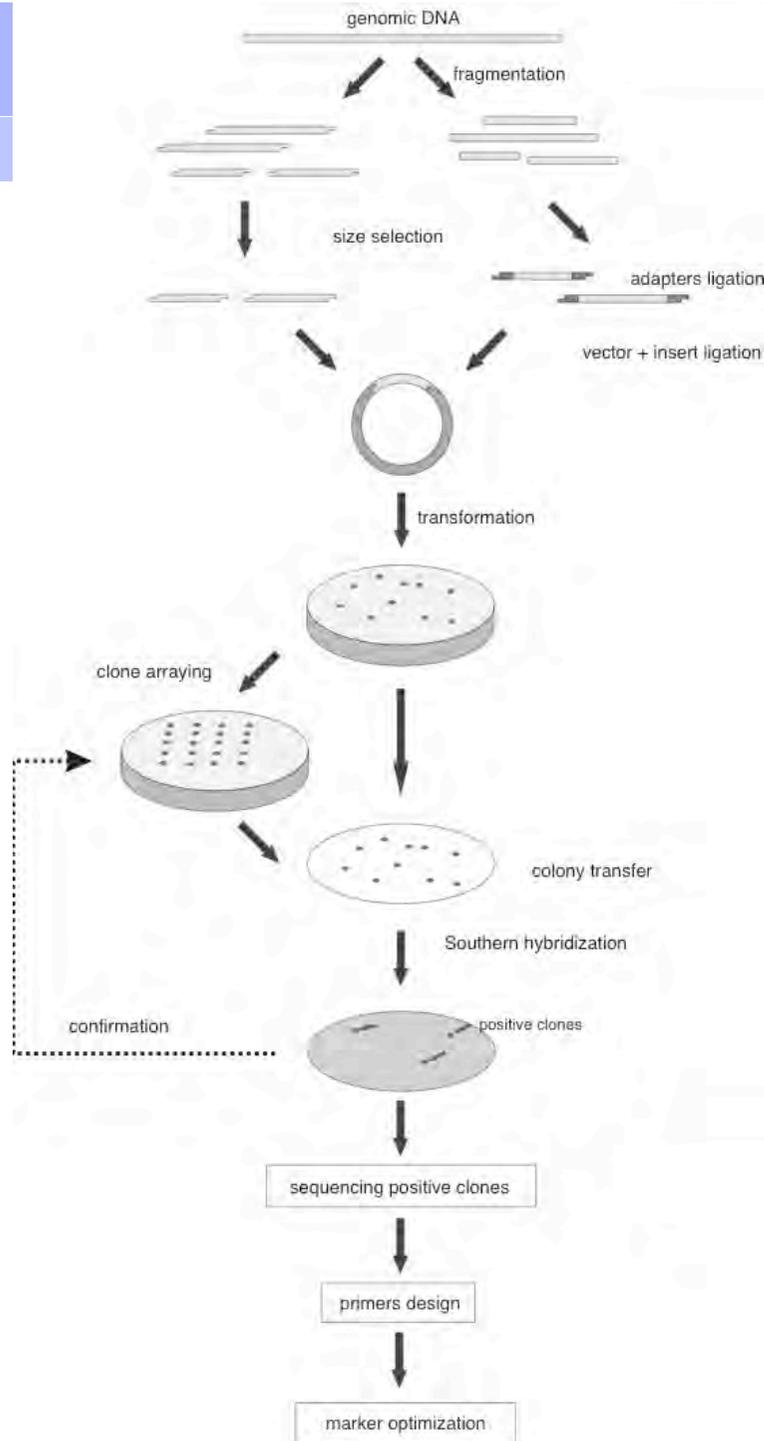


3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites

Stratégie d'obtention de marqueurs microsatellites

- Test de marqueurs développés pour une espèce proche
 - Pas toujours disponibles
 - Mutations dans la région des amorces de PCR
 - Généralement moins polymorphes que pour les espèces d'origines
- Développement de marqueurs de novo par vous même
 - Nécessite généralement du clonage et d'autres manip de laboratoire longues et laborieuses
- Utiliser un service commercial pour le faire pour vous
 - Peut être vite très chers
 - Délais souvent longs



•Fragmentation de l'ADN

•Ligation (vecteur)

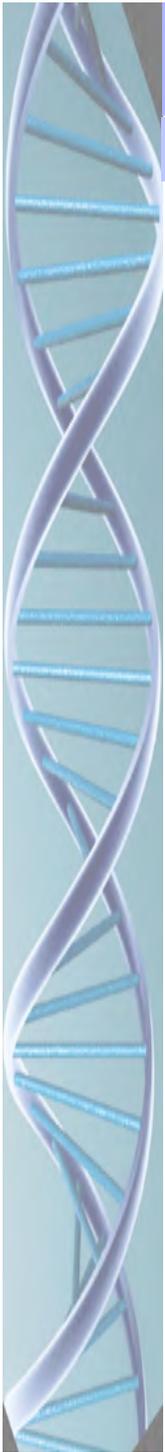
•Transformation des bactéries

•Hybridation (enrichissement) (sondes avec motifs microsatellites marquées)

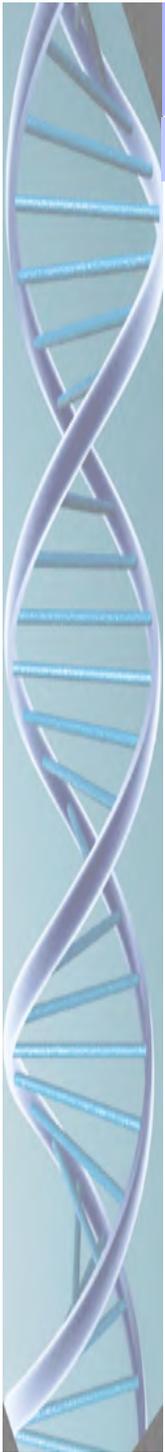
•Séquençage des clones positifs (Par séquençage Sanger)

•Design des primers (pour amplifier les microsatellites par PCR)

•Optimisation (Optimisation de l'amplification, test de polymorphisme)



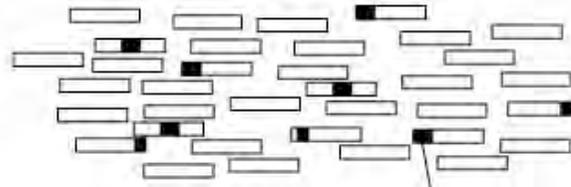
DNA extraction

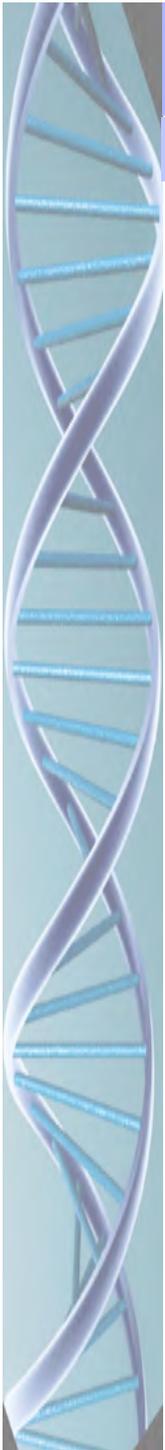


DNA extraction



Shotgun sequencing using Roche 454 technology





DNA extraction



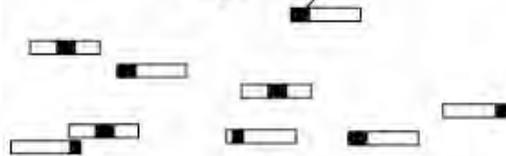
Shotgun sequencing using Roche 454 technology

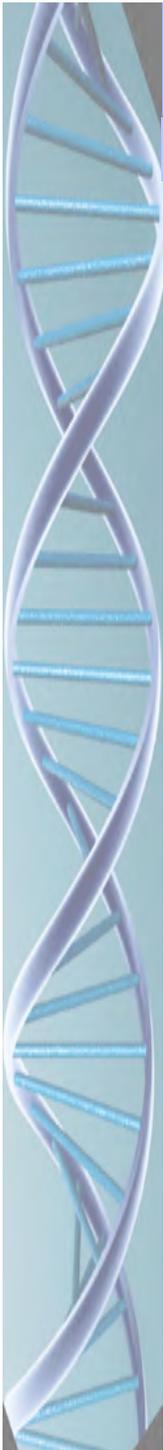


Bioinformatic detection of reads containing microsatellite repeated motifs



microsatellite motifs





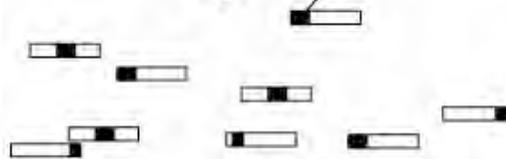
DNA extraction



Shotgun sequencing using Roche 454 technology



Bioinformatic detection of reads containing microsatellite repeated motifs



microsatellite motifs

Selection of repeated motifs enabling the design of primers
(i.e. enough space on each side of the repeated motif)

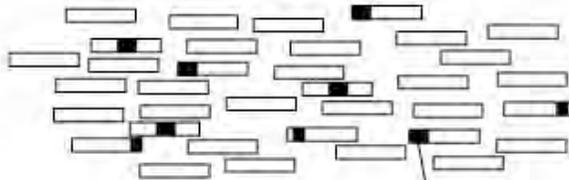




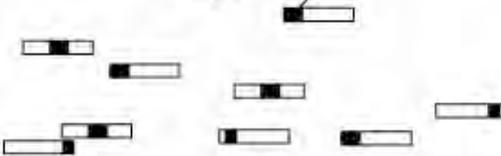
DNA extraction



Shotgun sequencing using Roche 454 technology



Bioinformatic detection of reads containing microsatellite repeated motifs



Selection of repeated motifs enabling the design of primers (i.e. enough space on each side of the repeated motif)



Primer design and test for polymorphism



Suffisamment de marqueurs polymorphes

3 - Quelques applications en écologie moléculaire

Développement de marqueurs microsatellites



DNA extraction

Genomic DNA (5 µg)

Shotgun sequencing using Roche 454 technology
(1/16th of a plate)

17,215 Reads
Average size 243 bp

Bioinformatic detection of reads containing microsatellite repeated motifs

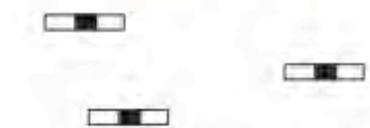
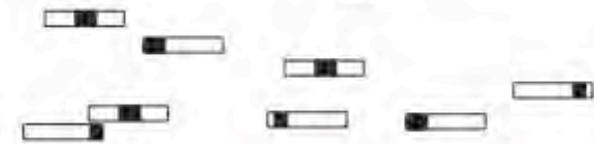
231 reads containing:
73 dinucleotides
107 trinucleotides
51 tetranucleotides

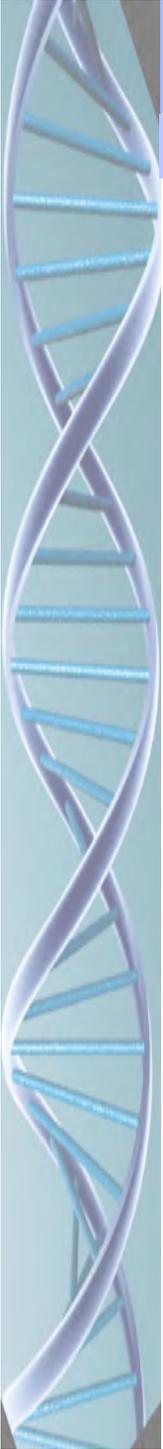
Selection of repeated motifs enabling the design of primers
(i. e. enough space on each side of the repeated motif)

24 reads containing:
17 dinucleotides
5 trinucleotides
2 tetranucleotides



microsatellite motifs





3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

- Problématique: Il peut être difficile de déterminer le régime alimentaire de certaines espèces:
 - Difficultés d'observation de l'animal
 - Contenu stomacal difficilement accessible!
 - Difficulté de détermination classique des restes alimentaires à partir de fèces

Comment mettre a profit les NGS pour rendre cette tâche plus simple?

3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

Molecular Barcoding

• Barcoding = une approche standardisée pour identifier des organismes par le séquençage de courts fragments d'ADN appelés barcodes ou code-barre moléculaire.

Le gène choisi doit présenter une forte variation entre espèces (forte variabilité interspécifique) et peu de variation au sein des espèces (faible variabilité intra-spécifique)

Pour les animaux: Cytochrome oxidase I (COI)

TREE-1026: No of Pages 8

ARTICLE IN PRESS

Review

Cell
PRESS

DNA barcoding for ecologists

Alice Valentini^{1,2}, François Pompanon¹ and Pierre Taberlet¹

¹ Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

² Dipartimento di Ecologia e Sviluppo Economico Sostenibile, Università della Tuscia, Via S. Giovanni Decollato 1, 01100 Viterbo, Italy

Trends in Ecology and Evolution



3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

On peut tenter de déterminer les espèces ingérées par un animal en déterminant les barcodes moléculaires présents dans le mélange d'ADN extrait des contenus stomacaux ou des fèces.

Molecular Ecology Resources (2009) 9, 51–60

doi: 10.1111/j.1755-0998.2008.02352.x

TECHNICAL ADVANCES

New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach

ALICE VALENTINI,^{*,†} CHRISTIAN MIQUEL,^{*} MUHAMMAD ALI NAWAZ,^{‡,§} EVA BELLEMAIN,^{*} ERIC COISSAC,^{*} FRANÇOIS POMPANON,^{*} LUDOVIC GIELLY,^{*} CORINNE CRUAUD,[¶] GIUSEPPE NASCETTI,[†] PATRICK WINCKER,[¶] JON E. SWENSON^{‡,**} and PIERRE TABERLET^{*}

3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires



Deosai National Park, Pakistan



Golden marmot



Brown bear



3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

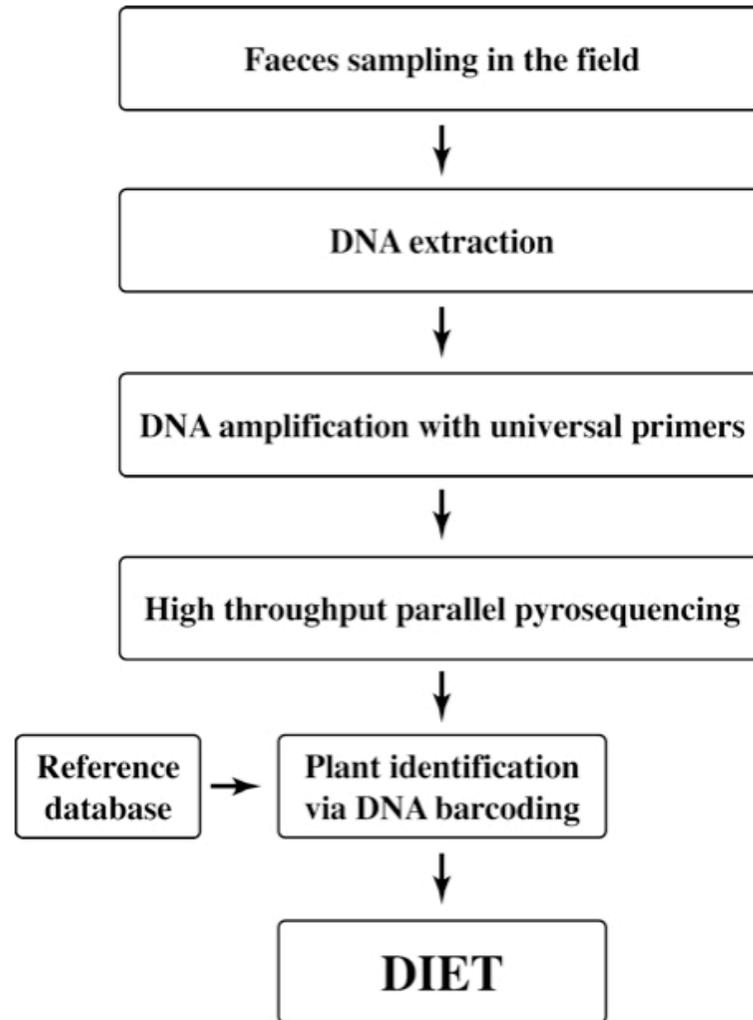


Fig. 1 Flowchart diagram showing the main steps of the *trnL* approach for assessing diet composition using faeces.

3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires



Deosai National Park, Pakistan



Golden marmot



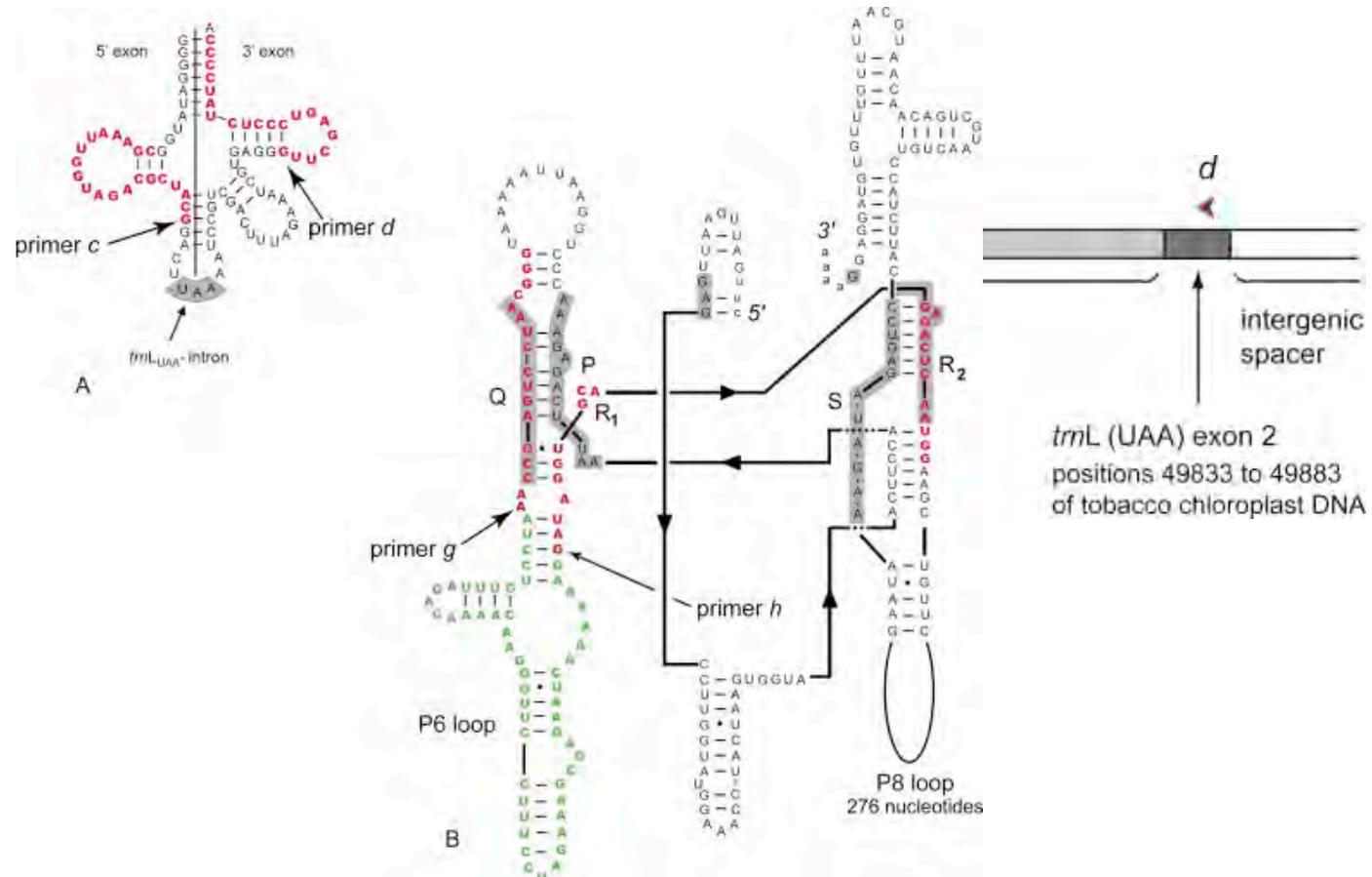
Brown bear

- Matériel biologique
 - 12 fèces récoltés pour chaque espèce
 - ADN extrait à partir de 10 mg de chaque échantillon

3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

- Marqueur génétique utilisé
 - P6 loop de l'intron du gène chloroplastique *trnL* (20 à 85bp)



3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires



Deosai National Park, Pakistan



Golden marmot



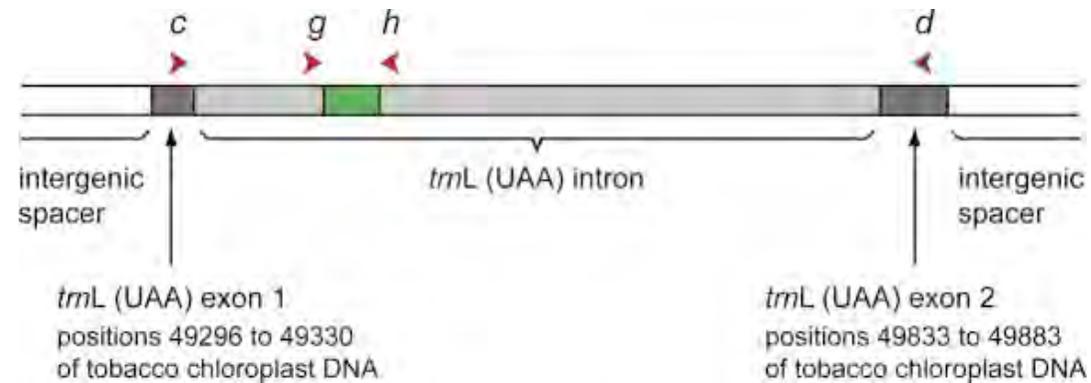
Brown bear

- Amplification par PCR
- Séquençage sur GS FLX (Roche/454)
- 97 737 séquences obtenues

3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

- Marqueur génétique utilisé
 - P6 loop de l'intron du gène chloroplastique *trnL*



3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

Table 3 Plant taxa identified in the diet of the Himalayan brown bear (*Ursus arctos*) and of the golden marmot (*Marmota caudata*) in Deosai National Park (Pakistan), based on sequence variation of the P6 loop of the chloroplast *trnL* (UAA) intron using faeces as a source of DNA

Family	Plant taxon	Level of identification	<i>Ursus arctos</i>												<i>Marmota caudata</i>													
			Faeces sample												Faeces sample													
			1	2	3	4	5	6	7	8	9	10	11	12	Total	1	2	3	4	5	6	7	8	9	10	11	12	Total
Apiaceae	Apoideae	subfamily				x									1													—
	<i>Heracleum candicans</i>	species	x	x								x	x		5	x							x		x			3
	<i>Pleurospermum hookeri</i>	species													—					x	x	x				x		4
Araceae	Araceae*	family												—			x										1	
Asteraceae	<i>Anaphalis nepalensis</i>	species												—											x		1	
	Anthemideae_1*	tribe				x								1	x			x	x			x	x	x	x	x	8	
	Anthemideae_2*	tribe												—				x	x					x			4	
	<i>Aster falconeri</i>	species												—				x				x	x		x	x	5	
	Asteraceae_1*	family												—					x								1	
	Asteraceae_2*	family			x									x	2				x	x	x	x			x	x	6	
	Asteraceae_3*	family												—				x	x								2	
	Asteraceae_4*	family												—								x					2	
	Asteraceae_5*	family												—								x					2	
	Asteraceae_6*	family												—											x		1	
	Asteroideae_1*	subfamily												—		x	x			x	x	x	x	x	x	x	8	
	Asteroideae_2*	subfamily												—						x			x			x	4	
	Asteroideae_3*	subfamily												—							x						1	
	Asteroideae_4*	subfamily												—						x							1	
	Coreopsideae*	tribe												—			x			x	x						3	
	Gnaphalieae*	tribe												—							x						1	
	Inuleae*	tribe												1			x						x			x	4	
		<i>Leontopodium brachyactis</i>	species											—									x				1	
	Brassicaceae	Brassicaceae	family											—												x	1	
		<i>Draba oreades</i>	species											—				x									2	
<i>Thlaspi andersonii</i>		species											—	x											x	2		



3 - Quelques applications en écologie moléculaire

Caractérisation de régimes alimentaires

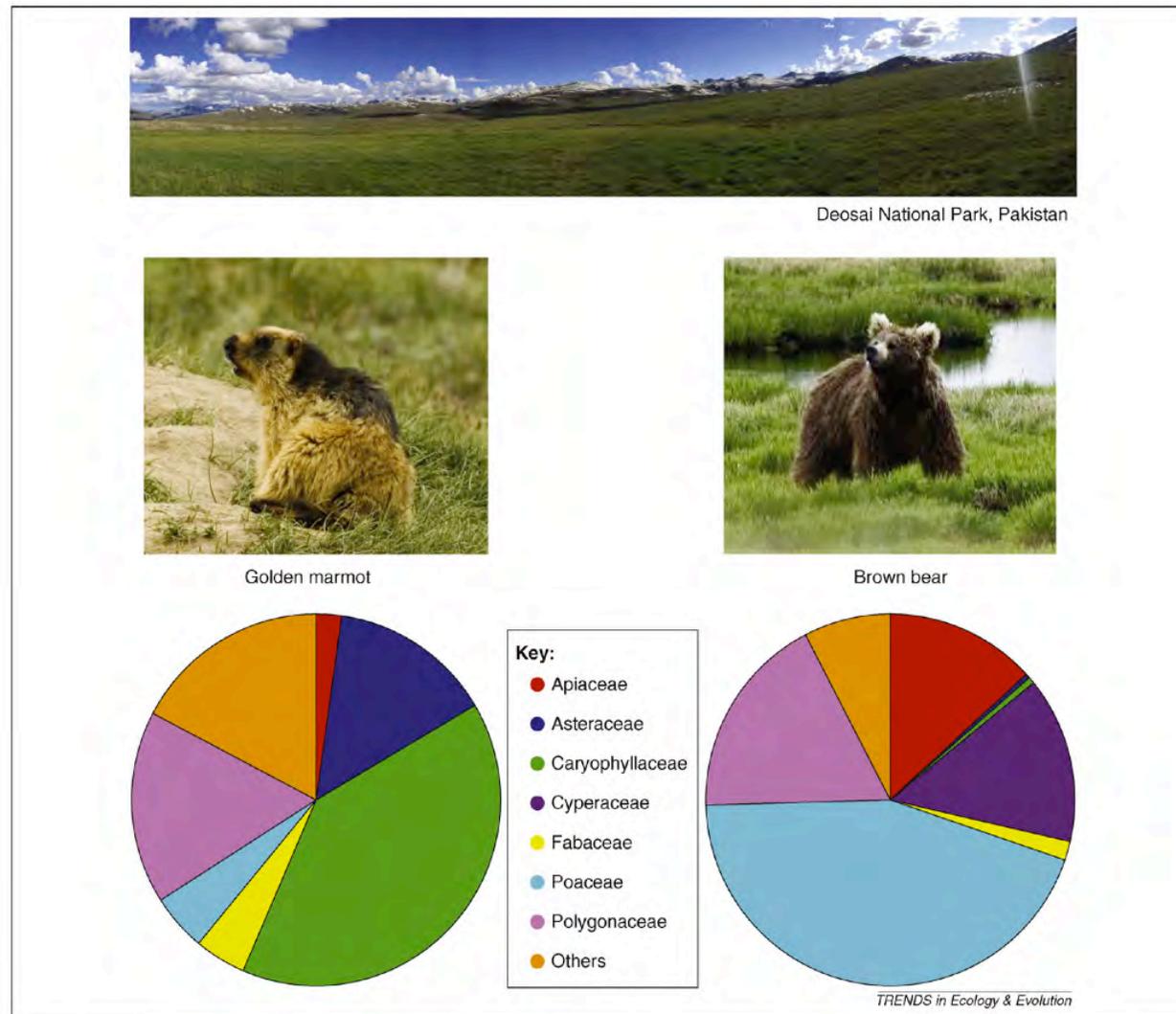


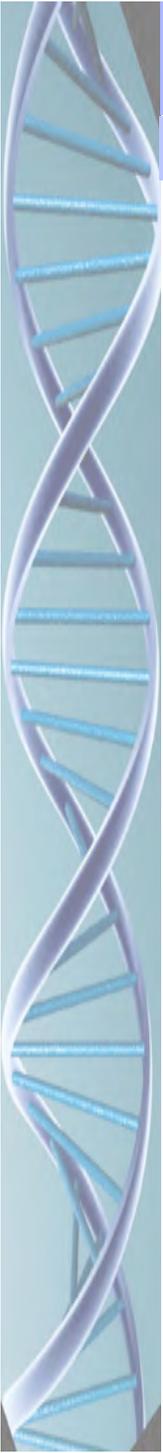
Figure 2. Example of feces analysis for estimating diet composition using a DNA barcoding approach (from Ref. [70]). Comparison of brown bear (*Ursus arctos*) and golden marmot (*Marmota caudata*) plant diets in the Himalayan environment. Twelve feces of each species were collected in the Deosai National Park (Pakistan). After DNA extraction and amplification with universal primers targeting a short fragment of the chloroplast *trnL* (UAA) intron [69], the amplicons were analyzed on the 454 GS FLX sequencer. More than 2000 DNA sequences were obtained per feces. The plant taxa eaten by both species were identified by comparison with available reference sequences. The results show that the diets of the two species are different: bears prefer Poaceae, whereas marmots prefer Caryophyllaceae.

3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

- Problématique : Il est difficile de déterminer la diversité des organismes présents dans un milieu donné. Particulièrement, l'utilisation de techniques moléculaires a montré que cette diversité était largement sous-évaluée pour ce qui est des microorganismes
- Métagénomique = terme récent désignant un champs de recherche visant à identifier le matériel génétique issue directement d'échantillons environnementaux (génomique environnementale, écogénomique, génomique des communautés)
- Echantillonnage environnemental = mélange de matériel organique et inorganique prélevé dans l'environnement (sol, fèces). Il peut contenir des organismes vivants (microorganismes ou petits macroorganismes, nematodes etc...) and des restes de macro organismes présent dans le site d'échantillonnage.





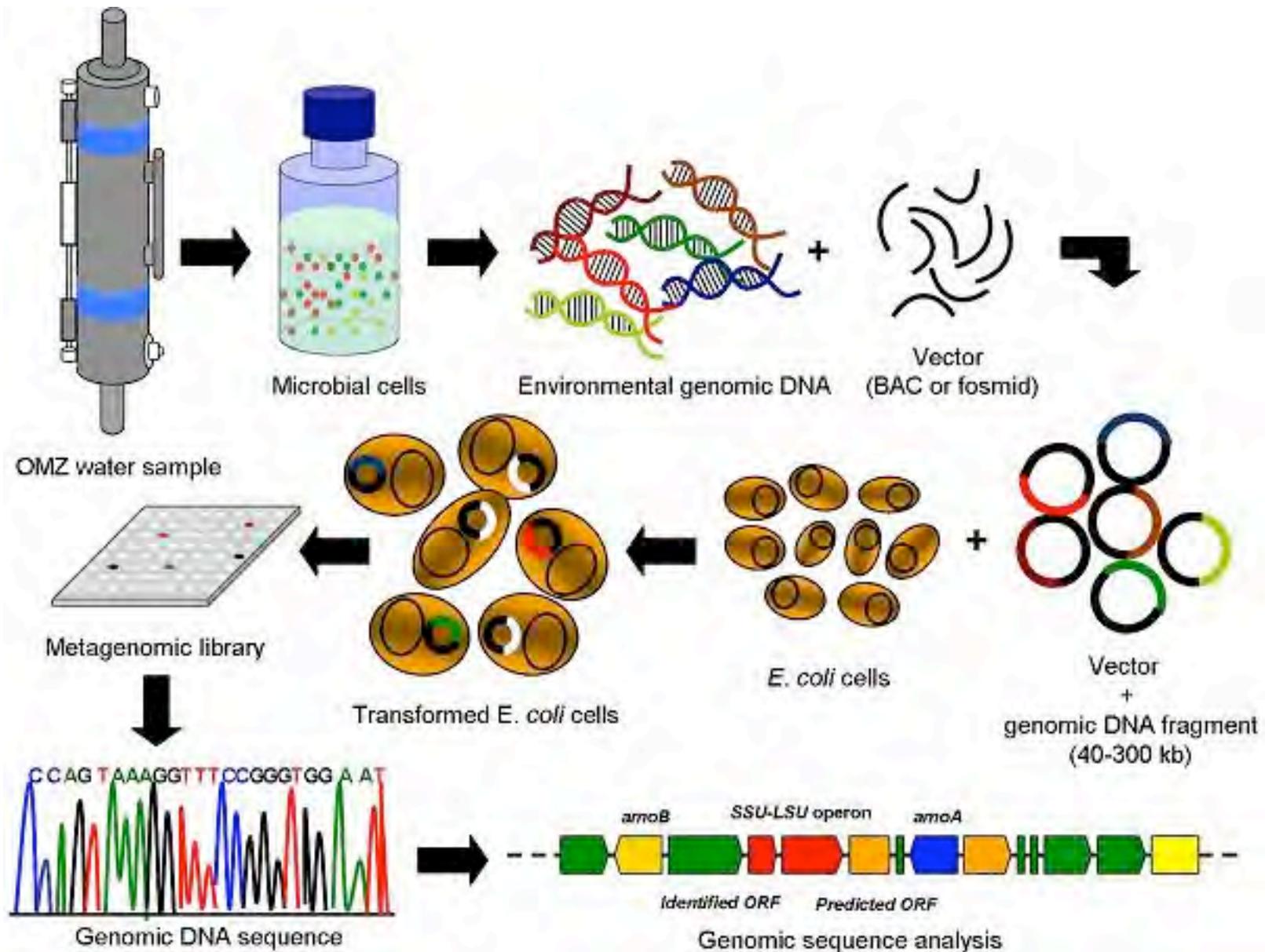
3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

- Traditionnellement, l'approche consiste à utiliser une combinaison de clonage aléatoire de l'ADN présent dans le mélange extrait de l'échantillon environnemental, puis de séquençage indépendant de ces clones.
- Les NGS permettent d'éviter ces étapes de clonage et d'obtenir une estimation non-biaisée (moins-biaisée) de la diversité présente par le séquençage massif des fragments d'ADN présents dans le mélange.

3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux



3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

MOLECULAR ECOLOGY

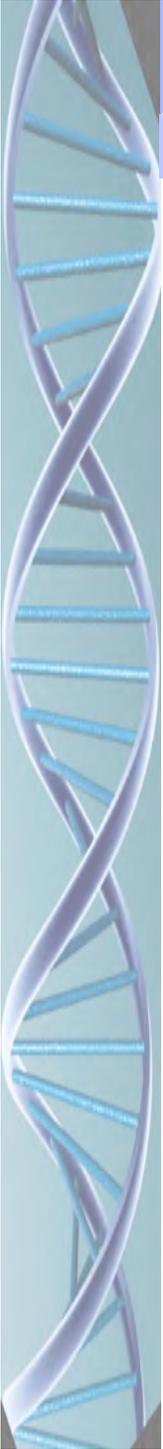
Molecular Ecology (2010), 19 (Suppl. 1), 21–31

doi: 10.1111/j.1365-294X.2009.04480.x

Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water

THORSTEN STOECK,* DAVID BASS,† MARKUS NEBEL,‡ RICHARD CHRISTEN,§ MEREDITH D. M. JONES,¶ HANS-WERNER BREINER* and THOMAS A. RICHARDS¶

*Department of Ecology, University of Kaiserslautern, D-67653 Kaiserslautern, Germany, †Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK, ‡Department of Computer Science, University of Kaiserslautern, D-67653 Kaiserslautern, Germany, §Université de Nice et CNRS UMR 654, Laboratoire de Biologie Virtuelle, Centre de Biochimie, Parc Valrose, Faculté des Sciences, Nice, F 06108, France, ¶Centre for Eukaryotic Evolutionary Microbiology, School of Biosciences, University of Exeter, Devon EX4 4QD, UK



3 - Quelques applications en écologie moléculaire

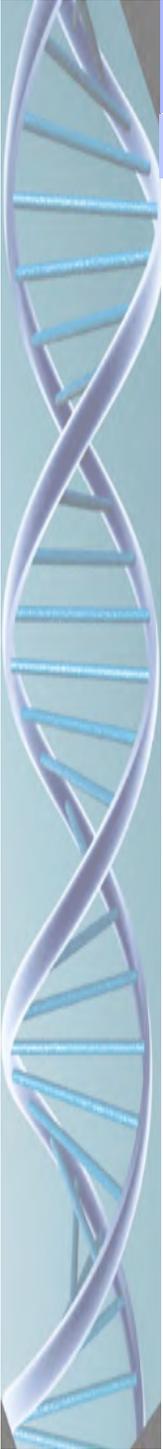
Détermination de la biodiversité à partir d'échantillonnages environnementaux

But: Caractériser l'assemblage de microorganismes eucaryotes présents dans la couche anoxique d'eau de Fjord Norvégiens.

Point de départ: Les prémices des techniques moléculaires ont montrés que la diversité des communautés de microorganismes eucaryotes était bien plus importante que ce que laissaient présager les estimations faites sur des bases morphologiques

Problème: les approches par clonage sont longues, relativement chères, et sont sujettes à de nombreux biais, notamment dans la détection d'organismes rares

Solution: L'utilisation de NGS doit permettre d'accélérer ce processus en le rendant plus simple par la génération d'un grand nombre de séquences



3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

Echantillon:

10 L d'eau provenant de la couche anoxique située à 18m de profondeur d'un Fjord Norvégien

Extraction:

L'ADN des organismes présents est isolé et extrait de façon globale

Marqueurs choisis:

2 régions variables de la petite sous-unité de l'ADNr, V4 et V9 (200-277bp)

Plateforme choisie:

GS FLX (Roche/454)

3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

Séquences obtenues:

271 197 pour V4

330 873 pour V9

Séquences comparées à une base de donnée

NCBI nucleotide database

Par BLAST (BLASTn)

Table 1 SSU rDNA data and environmental metadata of the Norwegian Framvaren Fjord anoxic water sample

Parameter	V4	V9
N 454-reads total ^a	271 197	330 873
Total (unique)	254 521	270 454
high quality ^b tags	(39 396)	(17 190)
Total (unique ^c) protist	245 935	264 008
tags including fungi	(38 116 ^c)	(15 156 ^c)
Total (unique) Metazoa tags	7627 (662)	3455 (1373)
Total (unique) Archaea tags	0 (0)	152 (69)
Total (unique) Bacteria tags	5 (4)	1610 (312)
Total (unique) unassignable ^d tags	954 (587)	1229 (280)

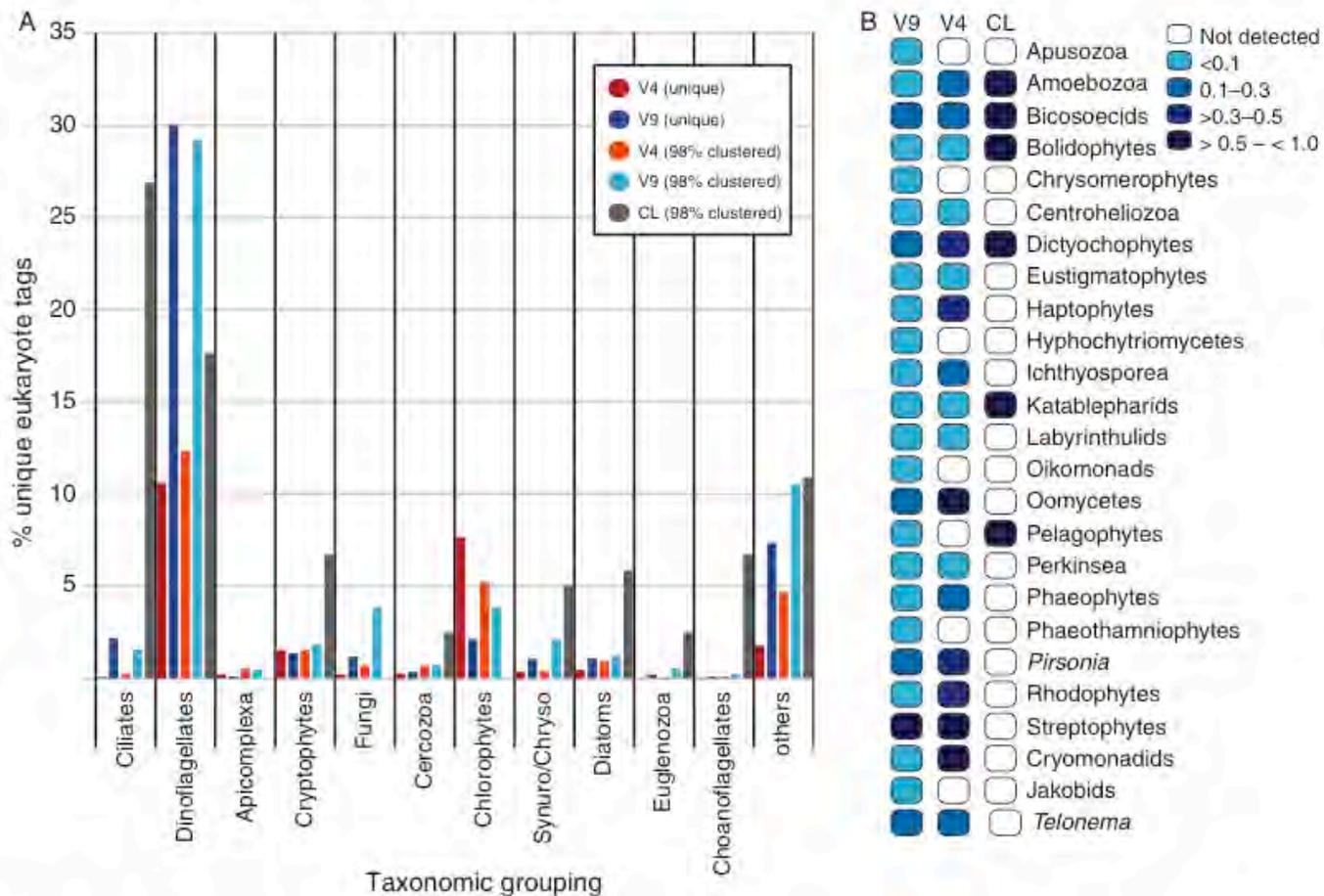


Fig. 4 Taxonomic diversity profile of V4 and V9 454 tags clustered at 100% and 98% similarity in comparison with a clone library-derived diversity profile, clustered at 98%. (a) Higher rank taxon groups that were represented by a proportion $\geq 1\%$ of all unique tags in at least one of the two sets of amplicons used for 454 sequencing is shown. The category 'others' denotes tags that could not be assigned to a taxonomic entity based on an 80% BLASTN similarity threshold and tags which fell into defined taxon groups but were represented by $<1\%$ of the diversity of unique tags in either of the two PCR amplicon libraries used for 454 sequencing (Fig. 4b). Only 0.42% of V4 tags and 0.46% of V9 tags could not be assigned to a taxon group using the 80% BLAST similarity threshold (Table 1). Clustering at 98% did not substantially change the taxonomic distribution of tags. (b) Higher rank taxon groups that were present in at least one of the two PCR amplicon libraries used for 454 sequencing or the clone libraries but that were represented by $<1\%$ of unique tags. It appears that although the clone-library was subjected to relatively large sequencing effort (for clone library studies; c. 1000 clones), this approach missed a substantial number of higher taxonomic groups, most of which were only recovered in low abundance within 454-derived tags. The V9 amplicon library retrieved a greater number of higher rank taxonomic groups than the V4 amplicon library. CL=clone library-derived data (from Behnke *et al.* 2006 and Behnke and Stoeck, unpublished).

3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité à partir d'échantillonnages environnementaux

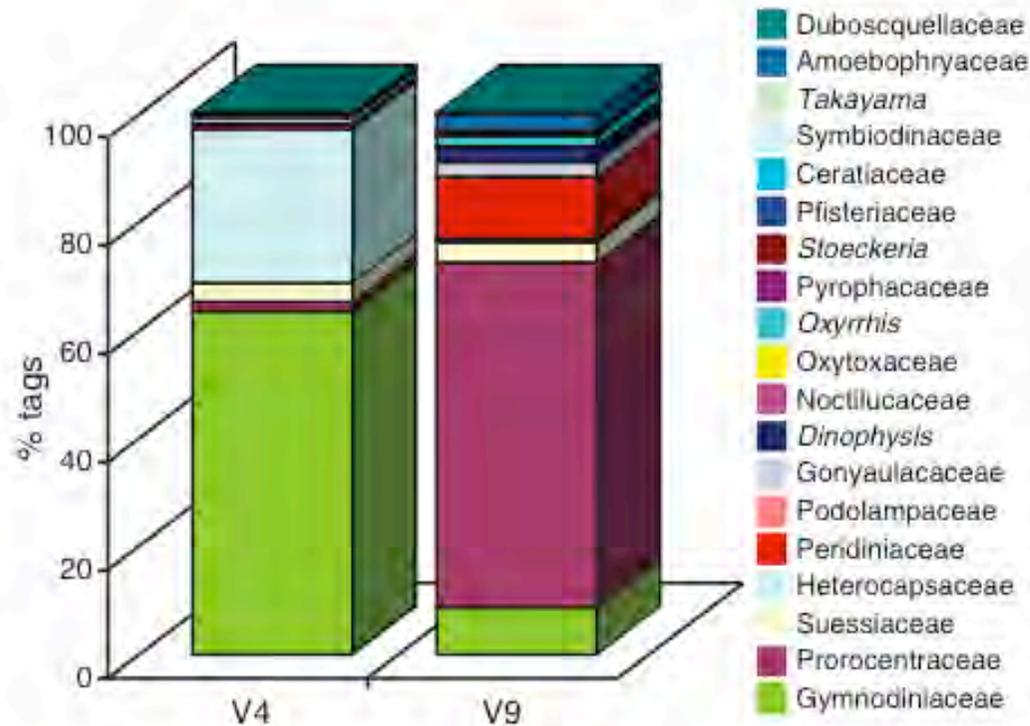


Fig. 5 Comparison of taxonomic distributions of V4 and V9 tags within dinoflagellates. The 454 sequencing results demonstrated that the dinoflagellates were the most diverse taxonomic group sampled (Fig. 4a). The V4 and V9 primer pairs detect very different taxonomic profiles, further confirming the extensive diversity of dinoflagellate sequences within the sample.

3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité présente dans un environnement passé

But: Caractériser la communauté végétale présente dans un paléoenvironnement

Point de départ: La détermination des paléoenvironnement se fait traditionnellement par une étude morphologique des pollens et des macro-fossiles

Problème: Ces approches peuvent être longues et laborieuses, et révèlent souvent une faible résolution taxonomique et de forts biais d'échantillonnage

MOLECULAR ECOLOGY RESOURCES

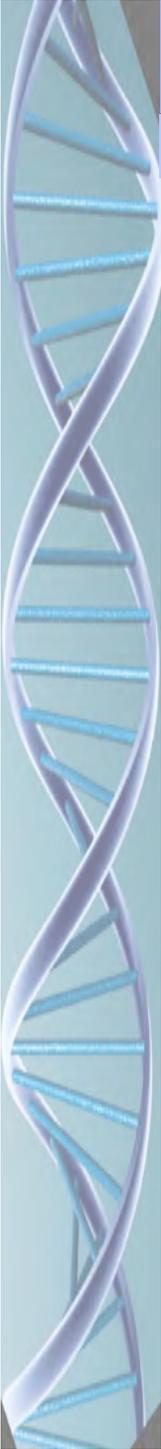
Molecular Ecology Resources (2010)

doi: 10.1111/j.1755-0998.2010.02855.x

TECHNICAL ADVANCES

Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate

J. H. SØNSTEBØ,* L. GIELLY,+ A. K. BRYSTING,† R. ELVEN,* M. EDWARDS,§ J. HAILE,¶**
E. WILLERSLEV,¶ E. COISSAC,+ D. RIOUX,+ J. SANNIER,* P. TABERLET+ and C. BROCHMANN*



3 - Quelques applications en écologie moléculaire

Détermination de la biodiversité présente dans un environnement passé

Echantillon:

6 grammes de sol provenant d'un permafrost de la région arctique de Russie
Echantillon daté au C14 a plus de 15000 ans (quaternaire ancien)

Extraction:

L'ADN des organismes présents est isolé et extrait de façon globale

Marqueurs choisis:

P6 loop d'un gène chloroplastique 'trnL', très variable (13-156bp)
Comparaison avec une détermination basée sur le pollen

Plateforme choisie:

GS FLX (Roche/454)



Référence taxonomique:

Base de donnée comprenant 842 espèces de la flore arctique, séquencées pour l'intégralité du gène trnL

Résultat:

La P6 loop suffit à identifier toutes les familles présentes, la plupart des genres (>75%), et 33% des espèces

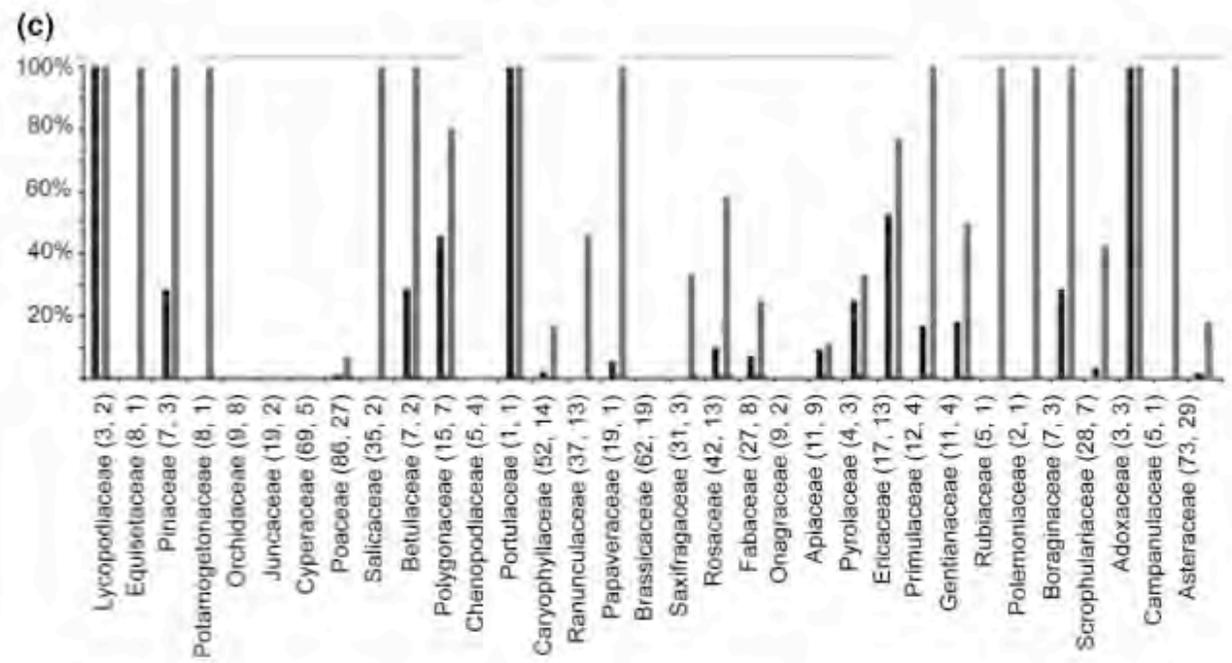
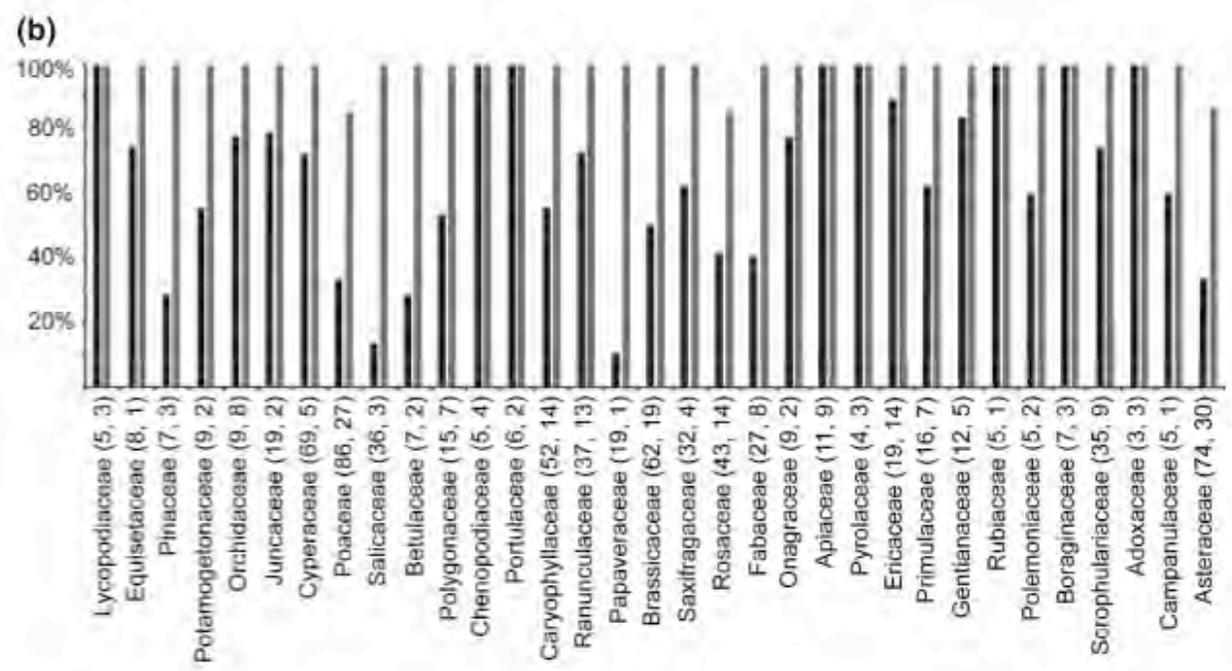
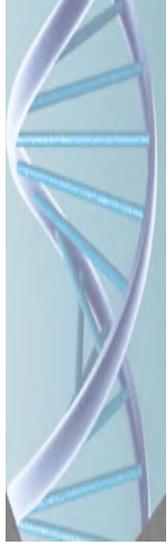
Ce niveau d'identification de résolution taxonomique est bien supérieur à celui obtenu par l'étude du pollen

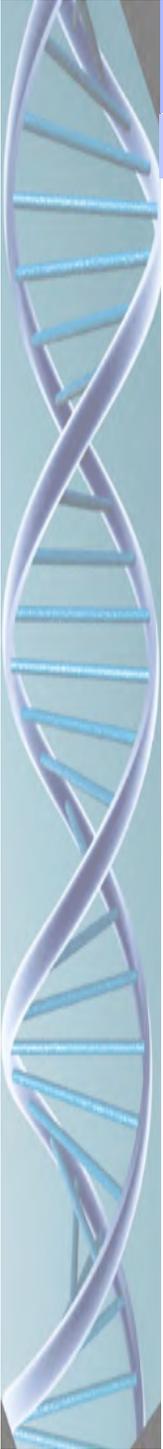
Table 2 Lists of molecular taxonomic units based on DNA from the two Chukotkan sediment samples, showing the molecular taxonomic units identified and their frequency of occurrence (% of sequences obtained in the 454 run)

	%
22 960 ± 120 years BP	
<i>Bistorta vivipara</i>	47.25
<i>Equisetum arvense</i> / <i>E. fluviatile</i> / <i>E. sylvaticum</i>	24.31
<i>Salix</i> sp./ <i>Chosenia arbutifolia</i> / <i>Populus balsamifera</i>	4.74
<i>Armeria scabra</i>	3.03
<i>Thymus oxyodontus</i>	2.77
<i>Lagotis glauca</i>	2.17
Asteraceae 1*	1.87
<i>Avenella flexuosa</i>	1.77
<i>Aconogonon alaskanum</i> / <i>A. ocreatum</i> / <i>A. tripterosperrum</i>	1.36
<i>Rumex</i> sp.	1.31
<i>Packera</i> sp./ <i>Senecio</i> sp.	0.96
Poaceae 1†	0.96
<i>Ranunculus acris</i> / <i>R. subborealis</i> / <i>R. turneri</i>	0.81
<i>Festuca</i> sp.	0.76
<i>Hulteniella integrifolia</i>	0.66
<i>Saxifraga hirculus</i>	0.55
<i>Trientalis europaea</i>	0.45
Asteraceae 2‡	0.40
<i>Valeriana capitata</i> / <i>V. officinalis</i> agg.	0.35
<i>Myosotis alpestris</i>	0.30
Asteraceae 3§	0.30
<i>Empetrum sibiricum</i> / <i>E. subholarcticum</i>	0.30
<i>Anthoxanthum nipponicum</i>	0.25
<i>Crepis chrysanta</i>	0.25
<i>Saxifraga bracteata</i> / <i>S. cernua</i> / <i>S. hyperborea</i> / <i>S. radiata</i> / <i>S. rivularis</i>	0.20
<i>Papaver</i> sp.	0.15
<i>Elymus</i> sp./ <i>Leymus</i> sp.	0.15
<i>Trollius</i> sp.¶	0.15
<i>Koeleria asiatica</i> / <i>Trisetum spicatum</i>	0.15
<i>Pedicularis oederi</i>	0.10
<i>Viola biflora</i>	0.10
<i>Claytonia arctica</i> / <i>C. scammaniana</i>	0.10
<i>Sanguisorba officinalis</i>	0.10
<i>Vaccinium uliginosum</i>	0.10
<i>Calamagrostis</i> sp.	0.10
<i>Potentilla</i> sp.	0.10
<i>Pulsatilla patens</i>	0.05
<i>Beckmannia syzigachne</i>	0.05
<i>Cardamine pratensis</i>	0.05
<i>Trisetum sibiricum</i>	0.05
<i>Vaccinium alaskense</i> / <i>V. myrtilus</i>	0.05
<i>Castilleja elegans</i> / <i>C. hyperborea</i>	0.05
<i>Deschampsia</i> sp.	0.05
<i>Parrya arctica</i> / <i>P. nudicaulis</i>	0.05
<i>Astragalus alpinus</i> / <i>A. umbellatus</i>	0.05
<i>Thalictrum alpinum</i> / <i>T. minus</i> / <i>T. sparsiflorum</i>	0.05
<i>Caltha arctica</i> / <i>C. palustris</i>	0.05



Fig. 1 Power of the P6 loop (a), the *trnL* intron (b) and pollen (c) for discrimination between genera and species in plant families with more than two genera or five species included in the taxonomic reference data base. Black and grey histograms show the percentages of discriminated species and genera, respectively. Numbers in parentheses refer to number of species and number of genera analysed.





Bibliographie

- Schuster 2008. Next-generation sequencing transforms today's biology. *Nature Methods* - 5, 16 - 18 (2008). Published online: 19 December 2007; | doi:10.1038/nmeth1156.
- Mardis ER. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9:387-402. Review.
- Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008 Mar;24(3):133-41. Epub 2008 Feb 11. Review.
- Shendure and Ji. 2008 Next-generation DNA sequencing. *Nat Biotechnol.* 2008 Oct;26(10):1135-45.
- Wheeler DA et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 2008 Apr 17;452(7189):872-6.



Problématique générale?

Justification de la pertinence du choix de l'échantillon

Quel(s) gène(s) a été choisi, pourquoi?

Quelle plateforme a été utilisé, pourquoi?

Combien de séquences ont été générées?

Quelles références ont été utilisées pour identifier les organismes présents?

Quels sont les résultats principaux?

Les auteurs font-ils part de limitations de l'approche choisie?