

TD1 : Lire l'ADN : des premières séquences au génome humain et aux fossiles

Aujourd'hui, il est possible de lire l'intégralité du génome d'un être humain en quelques heures, de comparer des milliers de génomes grâce à l'informatique, et même d'analyser l'ADN d'êtres humains disparus depuis des dizaines de milliers d'années. Ces avancées spectaculaires sont pourtant le résultat d'une **évolution progressive des méthodes de séquençage de l'ADN**.

Les biologistes sont passés d'une lecture lente et fragmentaire de l'ADN à une lecture massive, rapide, puis à une lecture directe de longues molécules. Chaque innovation technologique a permis de répondre à de **nouvelles questions scientifiques**, notamment sur le **génome humain**, sa diversité et son histoire évolutive.

Ce travail vise à comprendre **comment fonctionnent les principales méthodes de séquençage**, pourquoi elles se sont succédé, et dans quels contextes chacune est la plus adaptée.

Comment l'évolution des méthodes de séquençage de l'ADN a-t-elle permis de lire le génome humain à différentes échelles,

et pourquoi plusieurs technologies sont-elles nécessaires pour répondre aux questions biologiques actuelles ?

À l'issue de ce travail, vous devrez être capables de :

- expliquer le principe de fonctionnement de plusieurs méthodes de séquençage de l'ADN ;
- comparer leurs avantages et leurs limites ;
- relier chaque méthode à un type de question scientifique (gène, génome entier, ADN fossile) ;
- montrer que les progrès techniques ont profondément modifié notre compréhension du génome humain.

PARTIE 1 : Séquencer un fragment d'ADN : la méthode de Sanger (1977)

La méthode de Sanger, mise au point en 1977, repose sur la copie d'un brin d'ADN par une enzyme, l'ADN polymérase.

La réaction contient :

- des dNTP normaux, nécessaires à la synthèse ;
- des ddNTP, nucléotides modifiés qui bloquent l'élongation du brin lorsqu'ils sont incorporés.

L'incorporation aléatoire des ddNTP produit une collection de fragments d'ADN de tailles différentes, chacun se terminant par une base connue. La séparation de ces fragments permet de reconstituer la séquence.

A l'aide du document 1, répondre aux questions:

1. Identifier la molécule qui sert de modèle lors du séquençage de Sanger.
2. Expliquer le rôle des **ddNTP** dans cette méthode.
3. Montrer comment l'obtention de fragments de tailles différentes permet de reconstituer la séquence d'ADN.
4. Expliquer pourquoi cette méthode est adaptée au séquençage d'un **gène**, mais pas d'un **génome entier**.

PARTIE 2 : Lire un génome entier : le séquençage haut débit (NGS)

À partir des années 2000, de nouvelles technologies permettent de séquencer **des millions de fragments d'ADN simultanément**.

NGS : Next-Generation Sequencing, que l'on traduit en français par **séquençage de nouvelle génération** ou **séquençage haut débit**.

Le terme « nouvelle génération » indique une **rupture technologique** par rapport aux méthodes plus anciennes (comme Sanger).

Le principe repose sur :

- la **fragmentation du génome** en petits morceaux,
- l'ajout d'**adaptateurs**,
- la lecture parallèle de ces fragments,
- puis leur **assemblage par des outils informatiques**, à l'aide d'un génome de référence.
-

Questions :

5. Expliquer pourquoi il est nécessaire de fragmenter l'ADN avant le séquençage.
6. Comparer le nombre de fragments analysés simultanément dans la méthode de Sanger et dans le NGS.
7. Montrer que l'informatique est indispensable au fonctionnement du séquençage haut débit.
8. Expliquer en quoi cette méthode a rendu possible le **séquençage complet du génome humain**.

PARTIE 3 : Lire de longues molécules et des ADN anciens : méthodes de troisième génération

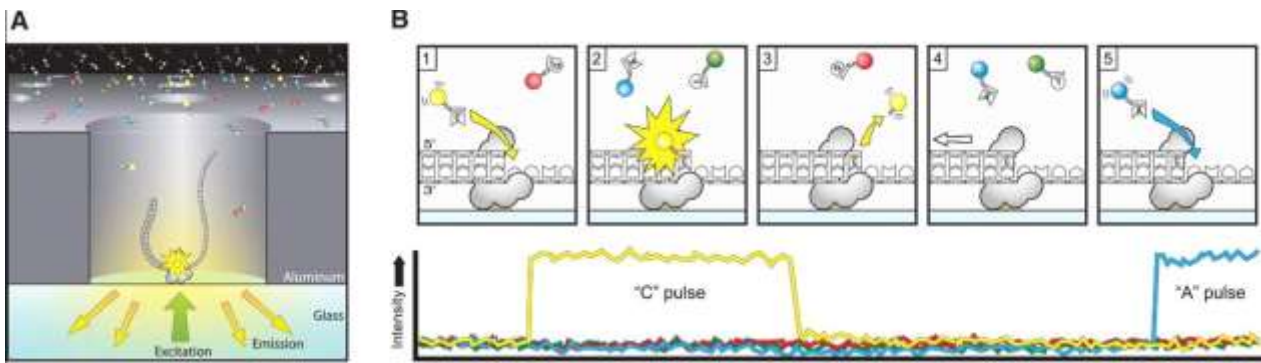
Les méthodes de troisième génération, développées à partir des années 2010, permettent de lire **de longues molécules d'ADN**, parfois sans amplification :

- **PacBio « SMRT* »** détecte la fluorescence des nucléotides lors de la synthèse. (doc.3). Cette méthode permet d'observer en temps réel la synthèse d'un brin d'ADN grâce à des nucléotides marqués par fluorescence. Chaque incorporation de nucléotide est détectée dans un espace très petit appelé ZMW**.

*SMRT signifie *Single Molecule Real-Time*, ce qui indique que la méthode permet de lire la séquence d'ADN molécule par molécule, en temps réel, pendant la synthèse de l'ADN.

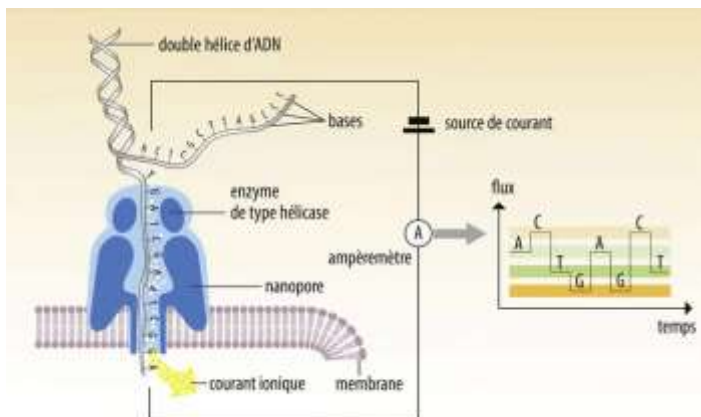
**ZMW signifie *Zero-Mode Waveguide* (guide d'onde en mode zéro) et correspond à une nanostructure qui permet d'observer l'activité d'une seule ADN polymérase lors de la synthèse de l'ADN.

- **Le séquençage par nanopore** détecte les variations de courant électrique produites par le passage de l'ADN. (doc.4)



Doc.3 : Principe du séquençage "SMRT" A: schéma du puits enfermant la molécule d'ADN liée à la polymérase; B: incorporation des bases, illumination et détection du signal ZMW

(<https://science.sciencemag.org/content/323/5910/133>)



Le séquençage de l'ADN par passage au travers d'un nanopore est assez sensible pour permettre l'analyse d'une molécule individuelle d'ADN.

Une enzyme (*unwinding enzyme*) de type hélicase sépare les deux brins de l'hélice d'ADN et linéarise l'ADN simple brin.

Ce monobrin va alors passer à l'intérieur du nanopore, lui-même inclus dans une membrane et traversé par un champ électrique.

Les variations spécifiques du courant engendrées par le passage successif des différents nucléotides dans le canal (chaque nucléotide possède une charge électrique caractéristique) sont enregistrées au fur et à mesure et il sera ainsi possible de reconstituer la séquence initiale de la molécule.

Doc.4 : Séquençage de troisième génération par nanopore

(Encyclopædia Universalis France)

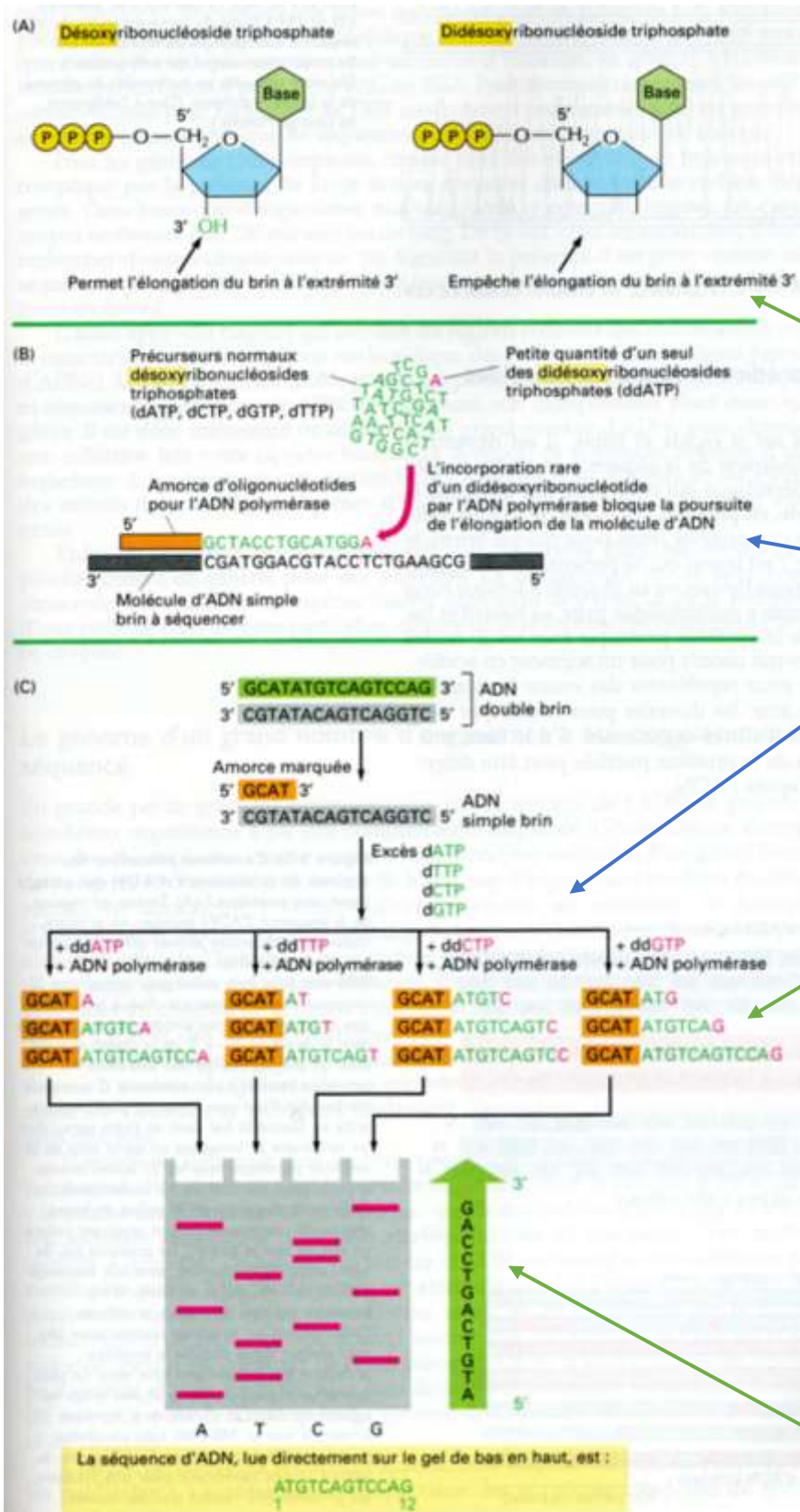
Questions

- Identifier la différence entre le type de signal mesuré par PacBio et par Nanopore.
- Expliquer pourquoi la lecture de longues molécules constitue un avantage scientifique.
- Justifier l'utilisation de ces méthodes pour l'étude de **l'ADN fossile**.
- Compléter le tableau en justifiant chaque choix.
- Montrer que le développement de nouvelles méthodes répond à de nouveaux besoins scientifiques.

Question scientifique	Méthode la plus adaptée
Séquencer un gène unique	
Comparer des milliers de génomes humains	
Étudier un ADN fossile fragmenté	
Identifier rapidement une mutation	

PARTIE 4 : synthèse

À partir des documents, montrer que les méthodes de séquençage de l'ADN ont évolué pour répondre à des questions scientifiques de plus en plus complexes, notamment l'étude du génome humain.



La méthode de séquençage de Sanger

1. Principe général

La méthode de Sanger permet de **déterminer l'ordre des nucléotides (A, T, C, G)** d'un fragment d'ADN.

Elle repose sur l'utilisation de **nucléotides particuliers**, appelés **didésoxynucléotides (ddNTP)**, qui **arrêtent la synthèse** de l'ADN lorsqu'ils sont incorporés.

2. Nucléotides normaux et nucléotides « stop »

- Les **désoxyribonucléotides (dNTP)** normaux possèdent un **groupement OH en 3'**, indispensable pour l'allongement du brin d'ADN.
- Les **didésoxyribonucléotides (ddNTP)** n'ont pas de groupement OH en 3'.

Lorsqu'un ddNTP est incorporé, **la synthèse du brin d'ADN s'arrête**.

3. Mise en place de la réaction

On utilise :

- un **brin d'ADN matrice** à séquencer,
- une **amorce**,
- une **ADN polymérase**,
- des **dNTP en grande quantité**,
- et une **faible quantité d'un seul type de ddNTP**.

On réalise **4 réactions séparées** :

- une avec **ddATP** (tube A),
- une avec **ddTTP** (tube T),
- une avec **ddCTP** (tube C),
- une avec **ddGTP** (tube G).

4. Formation de fragments d'ADN de tailles différentes

Dans chaque tube :

- la **synthèse de l'ADN commence normalement**,
- mais **s'arrête au hasard** dès qu'un ddNTP est incorporé.

On obtient donc **de nombreux fragments d'ADN de tailles différentes**, tous terminés par la **même base (A, T, C ou G selon le tube)**.

5. Séparation des fragments par électrophorèse

Les fragments d'ADN obtenus dans **chaque tube** sont déposés dans une **piste différente** d'un gel de polyacrylamide.

Chaque piste correspond à une réaction contenant un **didésoxynucléotide particulier** : A, T, C ou G.

Lors de l'électrophorèse :

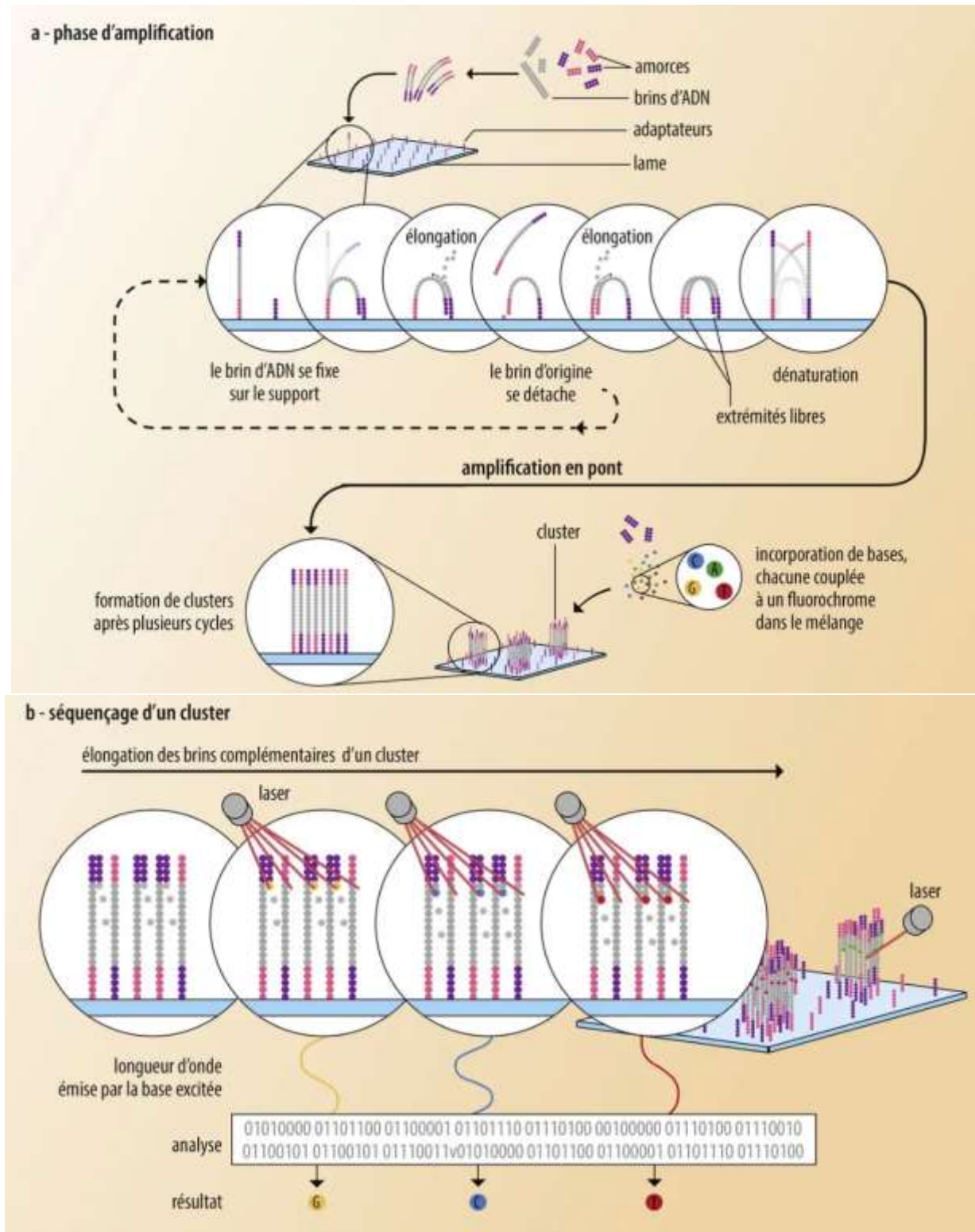
- les **fragments les plus courts migrent le plus loin**,
- les **fragments les plus longs migrent moins** et restent plus haut dans le gel.

6. Lecture de la séquence

- On lit le gel **de bas en haut**, car les fragments d'ADN les plus courts ont migré le plus loin.
- **À chaque position du gel**, on identifie la **piste (A, T, C ou G)** dans laquelle apparaît une bande.
- La succession des bandes, lues de bas en haut, permet de **reconstituer l'ordre des nucléotides du brin synthétisé**, puis celui du fragment d'ADN étudié.

Document 1

La méthode de séquençage de Sanger
(Biologie moléculaire de la cellule, 4^{ème} Ed., 2004)



Document 2 : séquençage haut débit

© Encyclopædia Universalis France - Conditions d'utilisation

Un **cluster** : est un **amas de copies identiques d'un même fragment d'ADN**, formé sur le support de séquençage lors du séquençage haut débit (NGS).

Selon les technologies de séquençage haut débit (NGS) :

- **Quelques millions de fragments**
→ plateformes de taille intermédiaire (usage courant en recherche)
- **Plusieurs centaines de millions de fragments**
→ plateformes très utilisées pour les génomes humains
- **Jusqu'à plusieurs milliards de fragments en parallèle**
→ plateformes industrielles de très haute capacité