

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de RELIZANE
Faculté des Sciences et de la Technologie
Département des Sciences Biologiques



POLYCOPIÉ DE COURS

Destiné aux étudiants de la 1^{ière} année Master Biochimie
appliquée

Intitulé

Génie génétique & outils bio-informatiques

Elaboré par :

Dr. BELGUENDOZ Amina

Année universitaire : 2022/2023

Figure 1.1 Séparation du sang par centrifugation en gradient de Ficoll et représentation schématique de l'isolement des granulocytes pour l'extraction de l'ADN.....	1
Figure 1.2 Un exemple de qualité d'ADN évaluée par amplification PCR.....	8
Figure 1.3 Extraction au phénol pour isoler les acides nucléiques sans contamination protéique.....	9
Figure 2.1 Hybridation Southern blot.....	12
Figure 3.1 Amplification par PCR d'une région définie de l'ADN via plusieurs cycles de dénaturation, d'hybridation d'amorce et d'extension.....	15
Figure 3.2 L'hybridation des amorces oligonucléotidiques au cours de l'étape d'annelage de la PCR définit la région d'amplification.	16
Figure 3.3 Une réaction PCR typique comporte trois étapes : dénaturation, hybridation et extension.....	17
Figure 3.4 Présentation de la PCR de transcription inverse.	19
Figure 3.5 PCR nichée versus semi-nichée.	20
Figure 3.6 Présentation de la PCR multiplex.	20
Figure 3.7 Profil de température pour les étapes de PCR.....	21
Figure 3.8 L'effet de la température sur l'hybridation de l'amorce 5 en PCR.....	23
Figure 3.9 Un exemple de tracé d'amplification PCR, où le cycle de seuil (Ct) est le nombre de cycles auquel la fluorescence franchit le seuil d'amplification.....	24
Figure 4.1 Le principe de la terminaison de chaîne de la synthèse de l'ADN.....	25
Figure 4.2 La procédure schématique du séquençage de terminaison de chaîne.	27
Figure 4.3 Le processus de séquençage automatisé.	30
Figure 4.4 Principe du pyroséquençage.....	32
Figure 4.5 Procédure automatisée de pyroséquençage.....	33
Figure.5.1 Disciplines scientifiques associées à la bio-informatique et à ses systèmes de soutien.....	34
Figure .5.2 Application de la bio-informatique dans différents domaines des sciences biologiques, agricoles et médicales.	37

Tableau 1.1 Chelex-100 et InstaGene Matrix Méthodes d'extraction d'ADN à partir de sang séché sur papier filtre.....	4
Tableau 5.1 Liste des ressources de base de données importantes et populaires pour la bio-informatique.	35

1	Extraction et purification des acides nucléiques	1
1.1	Extraction d'ADN d'un échantillon de sang total	1
1.1.1	Gradient de densité dirigé par Ficoll par centrifugation.....	1
1.1.2	Extraction rapide grâce à la protéinase K et au phénol	2
1.2	Extraction d'ADN des taches de sang sec	2
1.2.1	Préparation de taches de sang séché.....	2
1.2.2	Libérez l'ADN de la tache de sang séché	3
1.3	Isolement non invasif de l'ADN humain	4
1.3.1	Purification de l'ADN génomique des cheveux	4
1.3.2	ADN génomique purifié à partir de la salive	5
1.3.3	Extraction d'ADN génomique à partir d'un échantillon d'urine	5
1.3.4	Comparaison de différentes méthodes d'extraction.....	6
1.4	Préparation d'ADN à partir de micro-organismes	6
1.4.1	Méthodes chimiques.....	6
1.4.2	Méthode Chelex-100	7
1.5	Purification des acides nucléiques.....	8
1.5.1	Purification organique	9
1.5.2	Purification inorganique	10
2	Hybridation moléculaire.....	11
2.1	Les différentes techniques d'hybridation	11
2.1.1	Southern blotting ou hybridation ADN/AND	11
2.1.2	Northern blotting ou hybridation ARN/ADN	13
2.1.3	Dot-blotting et Slot-blotting	13
2.1.4	L'hybridation in situ.....	13
3	Amplification des acides nucléiques	14
3.1	La réaction en chaîne par polymérase	14
3.1.1	Les étapes d'une réaction PCR	16
3.1.2	Analyse des produits PCR.....	18
3.2	Techniques d'amplification des acides nucléiques	18
3.2.1	Réaction en chaîne par polymérase transcriptase inverse	18
3.2.2	Réaction en chaîne par polymérase nichée.....	19
3.2.3	Réaction en chaîne par polymérase multiplex.....	19
3.3	Paramètres affectant la PCR.....	20
3.3.1	La taille du produit PCR.....	21

3.3.2	Les longueurs des amorces.....	21
3.3.3	Température d'hybridation.....	21
3.4	PCR quantitative	23
4	Techniques de séquençage.....	25
4.1	Principe de la procédure didésoxynucléotide.....	25
4.1.1	Procédure.....	25
4.1.2	ADN polymérase utilisée dans le séquençage.....	28
4.1.3	Modèle de séquençement de terminaison de chaîne.....	28
4.1.4	Séquençage automatisé de l'ADN	28
4.2	Séquençage à haut débit : Le pyroséquençage	31
4.2.1	Principe du pyroséquençage.....	31
4.2.2	Avantage du pyroséquençage.....	33
5	Les outils bio-informatiques.....	34
5.1	Introduction à la Bio-informatique.....	34
5.2	But de la bio-informatique	36
5.3	Les défis de la bio-informatique.....	36
5.4	Application de la bio-informatique	37
5.4.1	Analyse de séquence	37
5.4.2	Analyse phylogénétique	38
5.4.3	Évaluation et validation du modèle de protéines prédites	38
5.4.4	Découverte et conception de petites molécules menant au développement de médicaments/agrochimiques	38
5.4.5	Analyse des données de séquençage de nouvelle génération.....	38
5.4.6	Identification SNP et SSR.....	39
5.4.7	Pharmacogénomique	39
5.4.8	Métabolomique et analyse des flux métaboliques.....	40
5.4.9	Biologie des systèmes et intégration de données omiques.....	40

Références bibliographiques

1 Extraction et purification des acides nucléiques

1.1 Extraction d'ADN d'un échantillon de sang total

L'ADN génomique obtenu à partir d'échantillons de sang total de patients est un élément clé pour le diagnostic moléculaire. L'isolement de l'ADN génomique est la première étape de l'exécution de tests de diagnostic moléculaire. Ainsi, il est essentiel d'obtenir de l'ADN génomique hautement purifié à partir de populations d'échantillons en utilisant des techniques d'isolement d'ADN appropriées.

1.1.1 Gradient de densité dirigé par Ficoll par centrifugation

Le sang total peut être prélevé en routine pour l'isolement des cellules mononucléaires du plasma et du sang périphérique en utilisant la centrifugation et un milieu à gradient de densité tel que Ficoll-Hypaque. Le sang frais est prélevé en présence d'anticoagulants tels que l'EDTA ou le citrate. Après centrifugation, le sang total est séparé en quatre couches, avec du plasma en haut suivi de globules blancs contenant des cellules mononucléaires du sang périphérique, une couche de milieu Ficoll-Hypaque et une couche inférieure contenant des érythrocytes et des granulocytes (Fig. 1.1). En effet, les globules rouges (GR) et les granulocytes ont une densité plus élevée que le Ficoll et se sédimentent au fond de la couche de Ficoll-Hypaque.

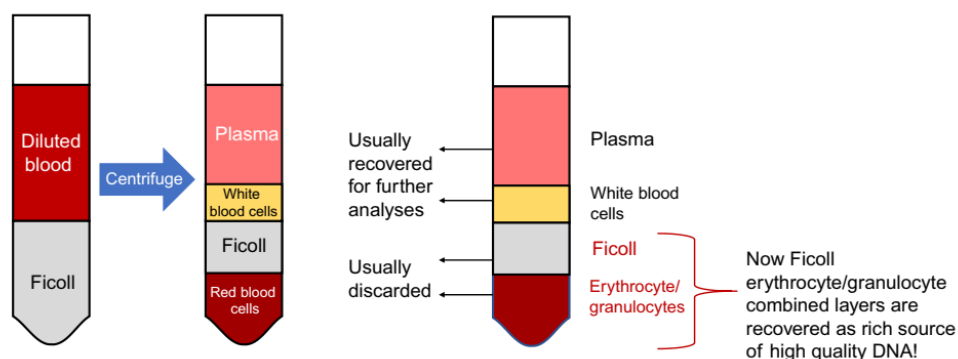


Figure 1.1 Séparation du sang par centrifugation en gradient de Ficoll et représentation schématique de l'isolement des granulocytes pour l'extraction de l'ADN.

L'ADN génomique peut être obtenu à partir du niveau plasmatique par une purification supplémentaire de l'ADN, comme décrit plus loin dans ce chapitre. Le rendement typique est d'environ 100 à 150 ng par ml de plasma. Pour les globules blancs, les plaquettes sont séparées des cellules mononucléaires par centrifugation à travers un

gradient de coussin de sérum bovin fœtal qui permet la pénétration des cellules mononucléaires mais pas des plaquettes. Après centrifugation, le surnageant contenant les plaquettes est jeté et les cellules mononucléaires du sang périphérique peuvent être collectées pour une purification supplémentaire de l'ADN.

Après l'élimination du plasma et des globules blancs, le Ficoll restant, qui est maintenant la couche supérieure, peut être retiré des érythrocytes/granulocytes résiduels, une riche source d'ADN génomique pour les études de génétique moléculaire. Cependant, les érythrocytes/granulocytes résiduels sont souvent jetés. En fait, les étapes de purification des granulocytes peuvent comprendre l'ajout de chlorure d'ammonium aux érythrocytes/granulocytes suivi de la lyse des érythrocytes, du lavage et de la remise en suspension du culot de granulocytes. Normalement, la purification des granulocytes doit être effectuée immédiatement après la séparation du sang. De cette manière, l'ADN peut être extrait des granulocytes.

1.1.2 Extraction rapide grâce à la protéinase K et au phénol

Deux autres méthodes d'extraction rapide d'ADN génomique qui ne sont pas impliquées dans la technique du gradient de densité Ficoll sont les méthodes à la protéinase K et au phénol. Pour le protocole de protéinase K, le sang total est mélangé avec du Tris, de l'EDTA, du dodécylsulfate de sodium (SDS), du $MgCl_2$ et de la protéinase K en présence de sel élevé pour une digestion nocturne à 37°C. Une fois la digestion terminée, les échantillons peuvent être purifiés davantage via un protocole de purification standard, comme décrit plus loin dans la section sur la purification de l'ADN. Pour la méthode au phénol, le sang total est mélangé avec du phénol saturé de Tris-HCl (pH 8,0) et de l'eau, puis agité pendant 4 h à température ambiante. Après centrifugation, la phase aqueuse est recueillie pour une purification standard supplémentaire comme décrit plus loin dans la section sur la purification de l'ADN.

1.2 Extraction d'ADN des taches de sang sec

1.2.1 Préparation de taches de sang séché

Généralement, l'ADN génomique extrait du sang total a été largement utilisé pour le diagnostic, mais les gouttes de sang séché (DBS) sur papier filtre sont apparues comme un moyen pratique de collecter et de transférer des échantillons pour l'extraction d'ADN génomique dans les zones rurales où les chaînes du froid efficaces font défaut et les

transports sont difficiles. L'utilisation de DBS pour la préparation d'ADN génomique présente de nombreux avantages. Par exemple, le sang prélevé dans des tubes doit idéalement être conservé à 4°C et traité dès que possible pour obtenir un rendement et une qualité d'ADN maximum. En revanche, les taches de sang collectées peuvent être séchées à température ambiante et expédiées sans réfrigération. De plus, le DBS peut être collecté à partir de prélèvements au doigt ou, pour les nourrissons et les jeunes enfants, à partir de prélèvements au talon.

Un autre avantage est que le DBS peut être conservé dans une installation de stockage à long terme (en dessous de 20 ° C) pendant des années avant l'extraction de l'ADN sans perte significative de qualité. Par exemple, l'ADN extrait de taches de sang stockées jusqu'à 25 ans a été utilisé avec succès dans l'amplification du génome entier pour diverses investigations génétiques, y compris le séquençage direct et les études d'association à l'échelle du génome.

La procédure de collecte simple et la stabilité de l'ADN dans les taches de sang permettent à la collecte d'être manipulée par des adultes sans obliger les participants à se rendre dans un établissement médical. Cela peut être particulièrement utile dans les grandes études épidémiologiques où les participants sont invités à donner des échantillons biologiques à plusieurs moments et/ou lorsque les participants à l'étude sont géographiquement dispersés.

1.2.2 Libérez l'ADN de la tache de sang séché

Le DBS peut être perforé à partir d'un papier filtre et est traité avec 0,5 % de saponine ou une solution saline tamponnée au phosphate (PBS) pendant une nuit à 4°C. Par la suite, le papier filtre perforé peut être soumis à la méthode Chelex-100 ou à la méthode InstaGene Matrix (Tableau 1.1). Pour Chelex-100, le poinçon est placé dans 5% de Chelex-100 et soumis à 100°C pendant 8min suivi par centrifugation pour recueillir le surnageant dans lequel l'ADN a été libéré. D'autre part, le poinçon peut être incubé avec InstaGene Matrix à 56°C pendant 30min suivi d'une centrifugation pour recueillir le surnageant. La sensibilité et la qualité de l'ADN de ces deux méthodes d'extraction d'ADN sont bonnes. Par conséquent, des méthodes simples et peu coûteuses telles que Chelex100 et InstaGene Matrix peuvent être sensibles et utiles pour extraire l'ADN de DBS. Un DBS de 11 à 15 mm de diamètre peut généralement produire de 0,8 à 1,6 µg d'ADN génomique. Pour l'application de la technologie de séquençage de nouvelle génération (NGS) dans les

études d'association à l'échelle du génome (GWAS) ou les études d'association à l'échelle du méthylome (MWAS), aussi peu que 1 µg de matériel de départ peut être utilisé sans sacrifier la qualité des données de séquençage et sans biaiser le profil de méthylation.

Tableau 1.1 Chelex-100 et InstaGene Matrix Méthodes d'extraction d'ADN à partir de sang séché sur papier filtre

Chelex-100			InstaGene Matrix	
Soaking in Saponin	Soaking in PBS	No Soaking	Soaking in PBS	No Soaking
1. Place punches in 1 mL 0.5% saponin	1. Place punches in 100µL PBS	1. Place punches in 180µL 5% Chelex-100 already heated to 100°C	1. Place punches in 100µL PBS	1. Place punches in a clean tube
2. Incubate at 4°C overnight	2. Incubate at 4°C overnight	2. Vortex 30 s	2. Incubate at 4°C overnight	2. Add 200µL InstaGene Matrix
3. Remove saponin and add 1 mL PBS	3. Centrifuge at 18,000g for 2 min	3. Incubate at 99°C for 10 min	3. Centrifuge at 18,000g for 2 min	3. Incubate at 56°C for 30 min. Vortex carefully after 15 min and after completed incubation
4. Incubate at 4°C for 30 min	4. Remove PBS and place punches in 100µL PBS	4. Centrifuge at 12,000g for 1.5 min	4. Remove and discard supernatant and add 100µL PBS	4. Boil samples at 100°C for 8 min
5. Remove PBS and place punches in 100µL 5% Chelex-100	5. Centrifuge at 18,000g for 2 min	5. Remove the supernatant and transfer to a clean tube	5. Centrifuge at 18,000g for 2 min	5. Centrifuge at 15,000g for 2 min
6. Incubate at 100°C for 8 min	6. Remove PBS and place punches in 100µL 5% Chelex-100	6. Centrifuge the supernatant at 12,000g for 1.5 min	6. Remove and discard supernatant	6. Carefully remove and store supernatant at -20°C if the extract is not used promptly
7. Centrifuge at 10,600g for 2 min	7. Incubate at 100°C for 8 min	7. Carefully remove and store supernatant at -20°C if the extract is not used promptly	7. Add 200µL InstaGene Matrix	
8. Carefully remove and store supernatant at -20°C if the extract is not used promptly	8. Centrifuge at 10,600g for 2 min		8. Incubate at 56°C for 30 min. Vortex carefully after 15 min and after completed incubation	
	9. Carefully remove and store supernatant at -20°C if the extract is not used promptly		9. Boil samples at 100°C for 8 min	
			10. Centrifuge at 15,000g for 2 min	
			11. Carefully remove and store supernatant at -20°C if the extract is not used promptly	

1.3 Isolement non invasif de l'ADN humain

Bien que les échantillons de sang soient utilisés comme source fiable d'ADN génomique, un tel échantillonnage est de nature invasive. De plus, il a souvent besoin de sources alternatives lorsque seul un protocole de collecte auto-administré est faisable sur le plan logistique ou économique, ou comme source d'ADN de secours dans les études qui collectent des échantillons de sang. La facilité d'échantillonnage et la nature non invasive de la méthode de collecte évitent les contraintes associées à l'échantillonnage à partir de sources traditionnelles.

1.3.1 Purification de l'ADN génomique des cheveux

Pour la purification simple de l'ADN génomique d'un cheveu avec la racine, une simple méthode de lyse alcaline peut être utilisée. Un cheveu avec racine est incubé à 95°C pendant 10 min dans du tampon NaOH, et le surnageant est soumis à une purification d'ADN après centrifugation.

Alternativement, une méthode de digestion chimique douce utilisant du dithiothréitol (DTT) peut être employée. Le DTT est un agent réducteur puissant avec une teneur en sel relativement élevée et également un détergent anionique. Un échantillon de cheveux peut être incubé 2h à 56°C avec un tampon contenant Tris-HCl, EDTA, NaCl, SDS, DTT et protéinase K, suivi d'un mélange doux et d'une incubation à 60°C pendant 2h ou jusqu'à ce que les cheveux soient complètement dissous. L'ADN génomique peut ensuite être extrait de la solution.

1.3.2 ADN génomique purifié à partir de la salive

Les cellules épithéliales buccales exfoliées et d'autres cellules trouvées dans la salive sont également une source alternative d'ADN très prometteuse car elles peuvent être obtenues à l'aide de techniques auto-administrées, non invasives et relativement peu coûteuses. Les écouvillonnages buccaux et les protocoles de bains de bouche sont les protocoles les plus couramment utilisés pour la collecte de cellules buccales. Différents types d'écouvillons buccaux, tels que des cotons-tiges ou des cyto-brosses, peuvent fournir des rendements d'ADN et des taux de réussite PCR similaires. Les échantillons d'écouvillons buccaux sont d'abord mis en suspension dans un tampon de lyse contenant du Tris, de l'EDTA, du SDS et de la protéinase K. L'échantillon est incubé 1 à 3 h à 56 °C jusqu'à ce que le tissu soit totalement dissous. L'ADN est ensuite extrait de la solution. Pour la méthode du rince-bouche, les échantillons de rinçages salins doivent être traités ou congelés immédiatement après le prélèvement. De la même manière, des cotons-tiges conservés pendant 4 jours à 37°C peuvent également être contaminés par de l'ADN d'origine majoritairement bactérienne. Un rince-bouche contenant de l'alcool peut être utilisé pour réduire la croissance bactérienne avant l'isolement.

1.3.3 Extraction d'ADN génomique à partir d'un échantillon d'urine

L'échantillon d'urine est inversé ou agité dans une coupelle d'échantillon pour créer une suspension homogène de cellules suivie d'une centrifugation. Le surnageant est retiré et un culot sec contenant des cellules est refroidi à 20 ° C pendant 15 minutes, suivi de l'ajout d'un tampon de lyse comprenant du Tris, de l'EDTA, du SDS et de la protéinase K. L'échantillon est incubé 2 h à 56 ° C. L'ADN est ensuite extrait de la solution.

1.3.4 Comparaison de différentes méthodes d'extraction

La comparaison des procédures d'extraction montre que l'ADN isolé des cellules buccales, de l'urine et des cheveux peut être utilisé avec succès pour effectuer des tests basés sur la PCR. Bien que l'ADN de tous les échantillons convenait à la PCR, les échantillons de sang et de cheveux ont fourni un ADN de bonne qualité pour l'analyse de restriction du produit de PCR par rapport aux échantillons d'écouvillon buccal et d'urine.

1.4 Préparation d'ADN à partir de micro-organismes

Les bactéries et les champignons ont des parois cellulaires qui doivent être brisées pour libérer l'acide nucléique. Habituellement, les méthodes chimiques, où la lyse cellulaire est provoquée par l'exposition à des agents chimiques qui affectent l'intégrité des barrières cellulaires, et les méthodes physiques, dans lesquelles les cellules sont perturbées par des forces mécaniques, sont utilisées pour briser les parois cellulaires. Les méthodes chimiques impliquent l'utilisation d'enzymes et de réactifs chimiques, et les méthodes physiques comprennent la congélation dans l'azote liquide, le broyage avec un pilon, la digestion enzymatique, le broyage de billes de verre et le micro-ondes.

1.4.1 Méthodes chimiques

Pour la méthode chimique, une combinaison d'enzymes et de réactifs chimiques est couramment utilisée. La lyse cellulaire par détergent est une alternative plus douce et plus facile à la perturbation physique des membranes cellulaires, bien qu'elle soit souvent utilisée en conjonction avec l'homogénéisation et le broyage mécanique. Les détergents brisent la barrière lipidique entourant les cellules en solubilisant les protéines et en perturbant les interactions lipide-lipide, protéine-protéine et protéine-lipide. Les détergents s'auto-associent et se lient aux surfaces hydrophobes. Ils sont composés d'un groupe de tête hydrophile polaire et d'une queue hydrophobe non polaire et sont classés selon la nature du groupe de tête comme ionique (cationique ou anionique), non ionique ou zwitterionique. Leur comportement dépend des propriétés du groupe de tête et de la queue.

En général, les détergents non ioniques et zwitterioniques sont plus doux et moins dénaturants que les détergents ioniques et sont utilisés pour solubiliser les protéines membranaires lorsqu'il est essentiel de maintenir la fonction des protéines et/ou de conserver les interactions protéine-protéine natives pour les dosages enzymatiques ou les dosages immunologiques. CHAPS, un détergent zwitterionique, et la série Triton-X de

détergents non ioniques sont couramment utilisés à ces fins. En revanche, les détergents ioniques tels que le SDS sont de puissants agents solubilisants et ont tendance à dénaturer les protéines, détruisant ainsi l'activité et la fonction des protéines.

Les cellules animales, les bactéries et les levures ont toutes des exigences différentes pour une lyse optimale en raison de la présence ou de l'absence d'une paroi cellulaire. En raison de la nature dense et complexe des tissus animaux, ils nécessitent à la fois une lyse détergente et mécanique. Pour les bactéries, le lysozyme est très efficace pour décomposer les parois cellulaires des bactéries Gram-positives car leurs parois cellulaires contiennent une forte proportion de peptidoglycane. Bien que les bactéries Gram-négatives soient moins sensibles au lysozyme, la combinaison de lysozyme et d'EDTA peut briser efficacement les parois cellulaires et les membranes. D'autre part, la méthode de perturbation mécanique peut être utilisée pour perturber efficacement les cellules bactériennes Gram-positives et les spores.

Les cellules fongiques sont difficiles à perturber car les parois cellulaires peuvent former des capsules ou des spores résistantes. La lyticase, également connue sous le nom de zymolyase, peut être utilisée pour digérer les parois cellulaires des cellules fongiques.

Les cellules peuvent également être brisées par des produits chimiques alcalins, des détergents et des xanthogénates. L'ADN génomique fongique peut également être extrait avec la méthode CTAB (bromure de cétyltriméthylammonium). Habituellement, le CTAB se mélange avec du Tris, de l'EDTA, un tampon à haute teneur en sel et la suspension de conidies suivie d'une incubation à 65°C pendant 1 h. Cela peut briser les parois des cellules fongiques très efficacement.

1.4.2 Méthode Chelex-100

Chelex-100, une résine chélatrice qui a une grande affinité pour les ions métalliques polyvalents, peut également être utilisée pour isoler rapidement et efficacement l'ADN des champignons. Il est fréquemment utilisé pour libérer l'ADN des cellules par un traitement d'ébullition, tout en protégeant l'ADN des effets d'ébullition avec des billes de résine. En règle générale, la suspension de conidies des échantillons est extraite en ajoutant 5 % de résine Chelex-100. Le mélange est incubé à 90°C pendant 30 min suivi d'une centrifugation. Les échantillons sont ensuite incubés à nouveau à 90°C pendant 15 minutes. Ensuite, le surnageant est récupéré après centrifugation.

La méthode d'extraction Chelex permet d'obtenir rapidement de l'ADN à partir de spores. Chelex-100 rend efficacement l'extraction de l'ADN des spores disponible pour une utilisation directe dans les analyses moléculaires. De plus, la quantité et la qualité de l'ADN extrait sont adéquates pour l'analyse par PCR. Comparativement, la qualité des échantillons d'ADN isolés à l'aide de la méthode Chelex permet d'obtenir de l'ADN de bonne qualité à partir de spores et offre une alternative à la méthode d'extraction à base de CTAB (Fig. 1.2). Le rendement est d'environ 28 ng et 17,9 ng par μL à partir d'une suspension de conidies de 200 μL qui contient 1 à 5105 spores/mL pour l'extraction de Chelex et l'extraction de CTAB, respectivement. Ainsi, la méthode Chelex est généralement recommandée pour l'extraction d'ADN génomique fongique compte tenu de sa simplicité et de sa rentabilité, en particulier dans le traitement de routine de grandes quantités d'échantillons.

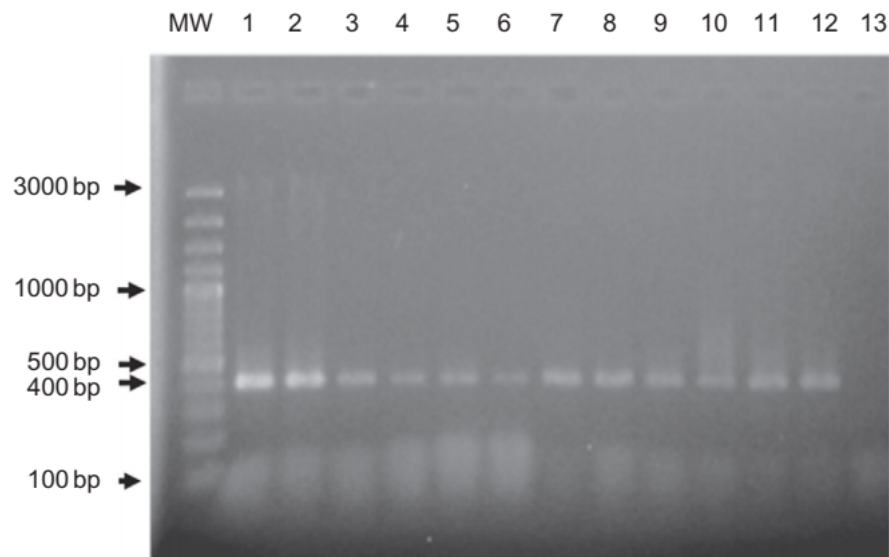


Figure 1.2 Un exemple de qualité d'ADN évaluée par amplification PCR.

Les ADN génomiques dérivés de deux méthodes d'extraction d'ADN sont comparés. Pistes 1 à 6 : échantillons d'ADN extraits avec Chelex-100. Pistes 7 à 12 : Les mêmes échantillons d'ADN extraits avec la méthode CTAB. MW, 100 bp plus échelle d'ADN (Fermentas Life Sciences), et Lane13 est le contrôle négatif (pas d'ADN).

1.5 Purification des acides nucléiques

La purification de l'ADN à partir d'un extrait cellulaire brut implique l'élimination des protéines, des glucides, des lipides et des débris cellulaires. Diverses techniques ont été développées, notamment des méthodes organiques, inorganiques et des colonnes de centrifugation.

1.5.1 Purification organique

Pour la méthode organique, le phénol et le chloroforme/alcool isoamylique (25:24:1) sont mélangés avec un volume égal d'échantillons au vortex. Bien que le phénol, un acide carbolique inflammable, corrosif et toxique, puisse dénaturer rapidement les protéines, il n'inhibe pas complètement l'activité de la ribonucléase (RNase). Un mélange de phénol:chloroforme:alcool isoamylique (25:24:1) est normalement utilisé pour inhiber l'activité RNase pendant le processus de purification. En outre, la protéinase K et la RNase peuvent être ajoutées à l'échantillon à cette étape pour éliminer les lipides et dégrader l'ARN, respectivement.

Après centrifugation, une émulsion biphasique se forme. La couche hydrophobe organique de l'émulsion est déposée au fond et la couche hydrophile aqueuse au-dessus. La couche hydrophile aqueuse supérieure contient de l'ADN et la couche organique inférieure contient les protéines précipitées (Fig. 1.3). La couche aqueuse supérieure est ensuite mélangée avec un volume égal de chloroforme, qui peut éliminer le phénol résiduel de la couche aqueuse. La couche aqueuse supérieure contenant l'ADN est maintenant pure et peut être davantage concentrée. Une concentration élevée de sel, comme l'acétate de sodium 0,3 M en présence de 2,5 volumes d'éthanol à 100 % ou 1 volume d'isopropanol à 20 °C ou moins peut précipiter l'ADN. Le précipité d'ADN est recueilli par centrifugation et l'excès de sel est rincé avec de l'éthanol à 70 % et centrifugé pour éliminer le surnageant d'éthanol. Le culot d'ADN est ensuite dissous avec une solution Tris-EDTA ou de l'eau distillée stérile pour un stockage à long terme.

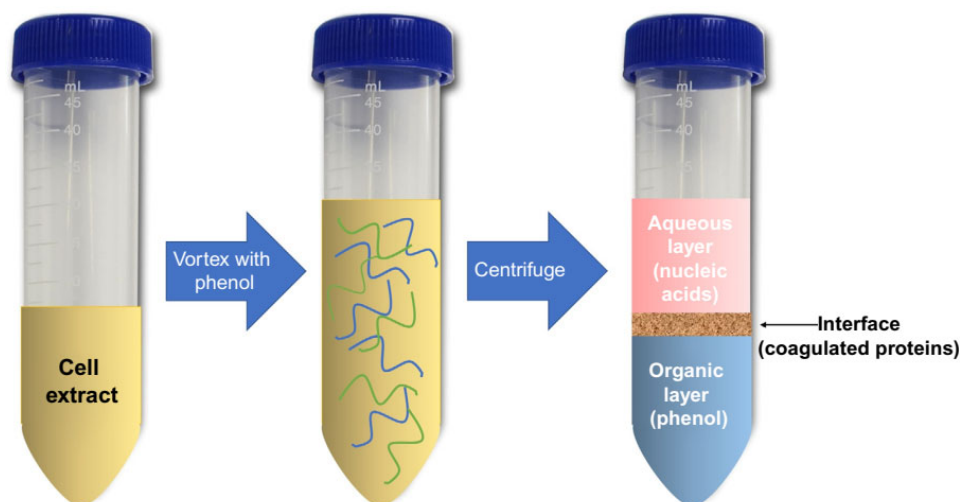


Figure 1.3 Extraction au phénol pour isoler les acides nucléiques sans contamination protéique.

Avec certains extraits cellulaires, la teneur en protéines est si importante qu'une seule extraction au phénol n'est pas suffisante pour purifier complètement les acides nucléiques. Ce problème peut être résolu en effectuant plusieurs extractions au phénol les unes après les autres, mais cela n'est pas souhaitable, car des mélanges et des centrifugations répétés peuvent casser les molécules d'ADN en petits fragments. L'autre alternative consiste à traiter l'extrait cellulaire avec une protéase telle que la protéinase K avant l'extraction au phénol. La protéinase peut décomposer les polypeptides en peptides plus petits, qui peuvent être facilement éliminés par le phénol.

1.5.2 Purification inorganique

La méthode inorganique implique l'incubation des noyaux avec uniquement la protéinase K à 65°C. Il a été montré que la protéinase K est plus active sur la protéine dénaturée et qu'après une incubation prolongée à 65°C elle peut être auto-inactivée. En conséquence, après une incubation de plus de 2h avec du Tris, de l'EDTA et de la protéinase K dans un tampon à faible teneur en sel, l'ADN extrait peut être utilisé directement pour l'analyse diagnostique sans aucune purification supplémentaire. Le rendement est supérieur à 90% de la théorie avec une taille moyenne supérieure à 300kb.

Le relargage est une autre méthode simple d'isolement de l'ADN inorganique. Dans cette procédure, du NaCl saturé est utilisé pour précipiter la protéine. Ensuite, l'ADN est purifié à partir du surnageant par addition de 1 à 2 volumes d'éthanol absolu refroidi à 20 °C (100 %). Dans ces approches, de l'ADN pur est obtenu et des substances non toxiques sont utilisées lors du traitement des échantillons. Cette technique est rapide et peu coûteuse pour une utilisation en laboratoire.

2 Hybridation moléculaire

2.1 Les différentes techniques d'hybridation

Les techniques d'hybridation reposent sur les propriétés physico-chimiques des acides nucléiques : (i) la complémentarité des bases constitutives de l'ADN (A/T, G/C) et de l'ARN (A/U, G/C) ; (ii) la réversibilité du processus de séparation des deux brins d'une molécule d'ADN (dénaturation) et de réassociation des deux brins (renaturation). Les dénaturations se traduisent par une dissociation des brins d'ADN (à 100 °C ou à pH alcalin) due à la rupture des liaisons hydrogène qui unissent les deux brins d'ADN. Lorsqu'il y a renaturation (reformation des liaisons hydrogène entre les deux brins), on parle d'hybridation :

- Si les deux brins s'hybrident présentent exactement les mêmes séquences (100 % de similarité), l'hybride sera stable dans des conditions de contrainte importante.
- Si les deux brins ne présentent pas de similarité dans leur séquence, l'hybridation n'aura pas lieu.
- En situation intermédiaire (similarité partielle entre les deux brins), l'hybride formé sera instable en conditions de contrainte forte, mais stable en conditions de contrainte faible. Une condition de contrainte forte est par exemple un milieu proche de l'eau (concentration en sel très faible) et une température élevée (supérieure à 65 °C). La température à laquelle les deux brins d'ADN se séparent est propre à chaque séquence (T_m). Plus la molécule étudiée est riche en bases G et C, plus la molécule bicaténaire sera stable, car l'appariement G/C est constitué de trois liaisons hydrogène, alors que l'appariement A/T n'en possède que deux.

2.1.1 Southern blotting ou hybridation ADN/AND

Le Southern blot permet la détection d'une séquence d'ADN donnée dans un mélange complexe de séquences d'ADN. Il peut être utilisé pour identifier des séquences homologues dans l'ADN génomique, ou pour faciliter la cartographie génétique par cartographie de restriction de gènes ou dans la détection de polymorphismes de longueur de fragment de restriction.

La première étape de l'hybridation Southern consiste à purifier l'ADN génomique de cellules eucaryotes ou de bactéries, suivie d'une ou de quelques digestions par des enzymes de restriction spécifiques. Les digestions enzymatiques produiront des milliers de fragments d'ADN génomique. Chaque enzyme spécifique produira des résultats de digestion différents. Par la suite, ces fragments d'ADN sont séparés par électrophorèse sur gel d'agarose, qui sépare les fragments en fonction de la taille, le plus petit ADN fragments migrant le plus loin dans le champ électrique. Si ces fragments sont visualisés par coloration, le gel doit montrer une série de bandes de frottis sans aucune bande discrète et distinguable (Fig. 2.1A). Après électrophorèse sur gel, les fragments d'ADN sont encore fragmentés à une taille inférieure à 1 kb et dénaturés par une solution alcaline douce suivie d'un transfert (Fig. 2.1B). Enfin, une sonde marquée est ajoutée pour s'hybrider avec les fragments sur la membrane pour la détection et l'identification.

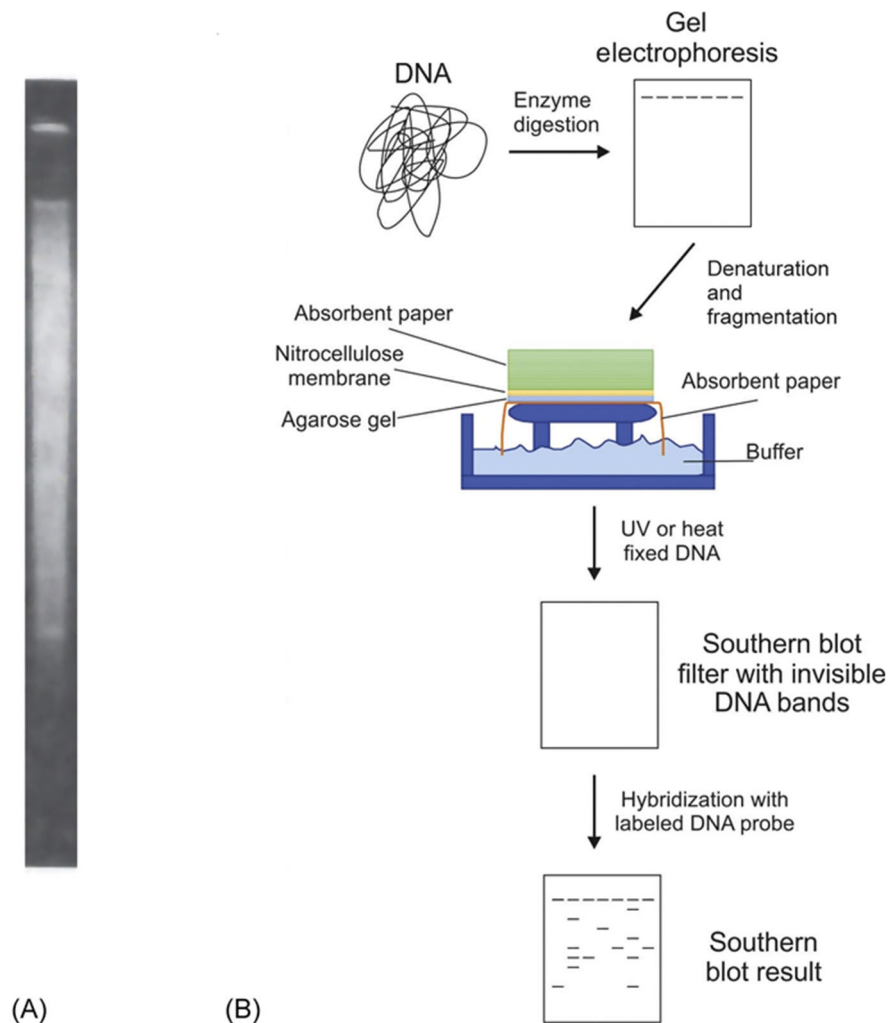


Figure 2.1 Hybridation Southern blot

(A) L'ADN génomique est digéré avec une ou quelques enzymes de restriction spécifiques suivies d'une électrophorèse sur gel. (B) Le processus schématique de l'hybridation Southern.

2.1.2 Northern blotting ou hybridation ARN/ADN

Cette technique, la sonde est constituée d'ADN, mais la cible est un ARN. Des ARN sont isolés d'un tissu et purifiés, puis séparés par électrophorèse sur gel d'agarose ou d'acrylamide. Les ARN complémentaires de la sonde sont identifiés après transfert sur une membrane de Nylon ou de nitrocellulose qui est ensuite soumise à l'hybridation. Ce type d'hybridation ADN/ARN permet d'évaluer la taille de l'ARN étudié ainsi que le taux d'accumulation d'une population d'ARN dans un tissu donné à un moment déterminé de son développement (différentes phases de l'embryogenèse) ou sous l'effet d'un stress biotique ou abiotique autoradiographiques dont les signaux sont révélés par numérisation.

2.1.3 Dot-blotting et Slot-blotting

La sonde est constituée d'ADN alors que la cible peut être soit de l'ADN monocaténaire, soit de l'ARN. À la différence des techniques de type Southern ou northern, les acides nucléiques cibles ne sont pas séparés par électrophorèse. Ils sont directement déposés sur une membrane d'hybridation à une concentration connue soit sous forme de point (dot), soit sous forme de trait (slot). Des appareils adaptés à ces techniques sont commercialisés. Après hybridation entre la sonde et la cible, l'intensité de la tache radioactive obtenue reflète la concentration de l'acide nucléique étudié (sonde) parmi les acides nucléiques cibles. Cette technique est plus rapide à mettre en œuvre que les expériences de type Southern ou northern. Cependant, il faut être certain que l'ADN sonde soit très spécifique de la séquence cible recherchée car l'absence de séparation des molécules cibles selon leur taille (par électrophorèse) empêche de déceler des hybridations non spécifiques.

2.1.4 L'hybridation in situ

Le principe de l'hybridation in situ des ARN consiste à déterminer la localisation de l'ARN issu d'un gène d'intérêt sur coupes de tissu, sur organe entier voire sur organisme entier selon la taille de l'échantillon. Cette technique repose sur une hybridation moléculaire entre deux molécules nucléotidiques complémentaires : une sonde nucléique marquée et une molécule d'ARN cible. La qualité de l'hybridation in situ varie en fonction de la sonde utilisée, de la température d'hybridation, de la concentration en sonde.

3 Amplification des acides nucléiques

3.1 La réaction en chaîne par polymérase

La PCR est une réaction enzymatique *in vitro* pour amplifier une région d'ADN définie. Une réaction PCR utilise normalement deux amorces oligonucléotidiques, qui sont hybridées aux bordures 5' et 3' de la séquence cible, et une ADN polymérase, qui peut étendre les amorces annelées en ajoutant des désoxyribonucléosides triphosphates (dNTP) pour générer des produits double brin (Fig. 3.1). Par la suite, le duplex d'ADN copié est dénaturé par la chaleur en un seul brin, et le cycle suivant commence avec deux fois plus de segments d'ADN à copier. Par conséquent, le nombre copié de segments d'ADN double idéalement après un cycle de traitement thermique. La PCR est utile comme test de point final pour examiner l'existence d'une séquence cible d'ADN dans un échantillon. N'importe quelle région de n'importe quelle molécule d'ADN peut être choisie. La seule exigence est que les séquences aux frontières de la région soient connues. Les oligonucléotides, qui agissent comme amorces pour les réactions de synthèse d'ADN, définissent la région amplifiée (Fig. 3.2). En augmentant et en abaissant la température du mélange réactionnel, les deux brins du produit d'ADN sont séparés et peuvent servir de matrices pour le cycle suivant d'hybridation et d'extension, et le processus est répété. L'amplification est généralement réalisée par la polymérase Taq, qui est une enzyme ADN polymérase I thermostable de *Thermus aquaticus*. La thermostabilité de la polymérase Taq est une exigence essentielle dans la PCR, car une réaction de PCR passera par différentes étapes de température qui peuvent dénaturer ou désactiver la polymérase non thermostable.

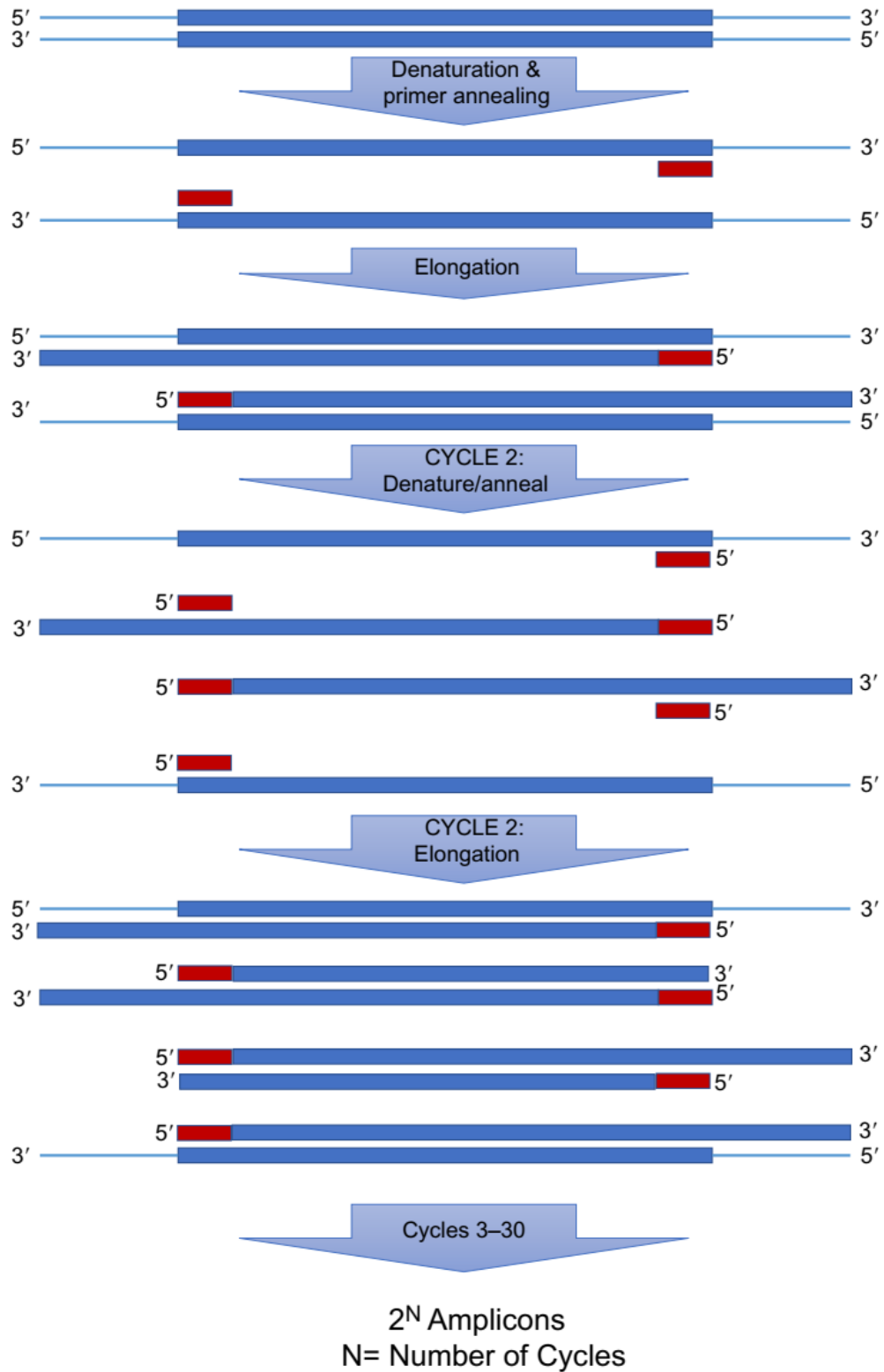


Figure 3.1 Amplification par PCR d'une région définie de l'ADN via plusieurs cycles de dénaturation, d'hybridation d'amorce et d'extension.

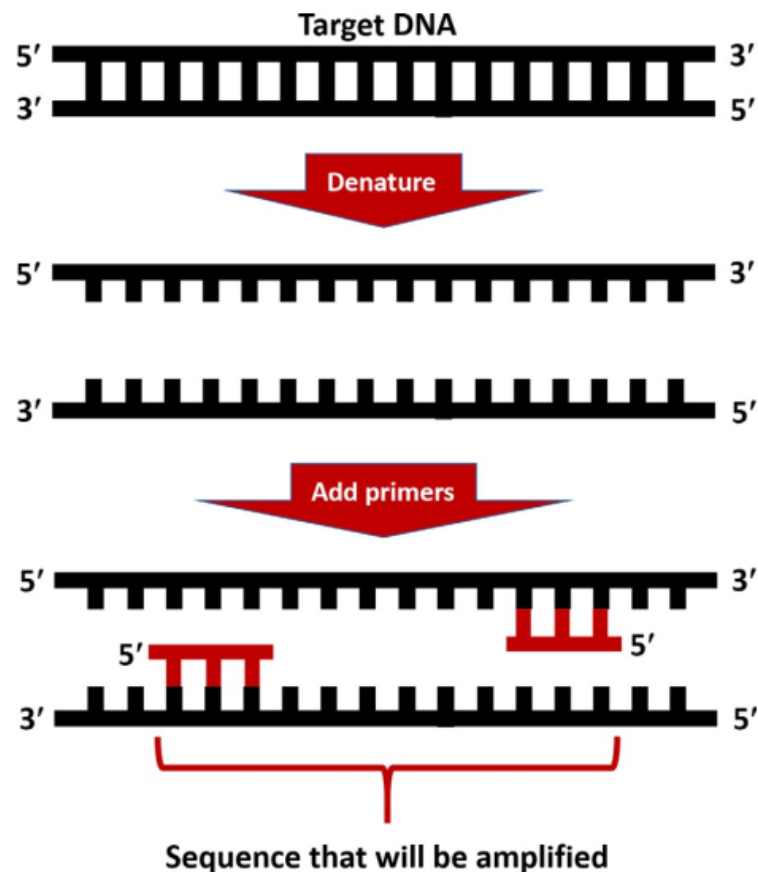


Figure 3.2 L'hybridation des amorces oligonucléotidiques au cours de l'étape d'anneau de la PCR définit la région d'amplification.

3.1.1 Les étapes d'une réaction PCR

Dans une expérience de PCR typique, l'ADN cible est mélangé avec la polymérase Taq, les deux amorces oligonucléotidiques et une réserve de dNTP. La quantité d'ADN cible peut être très faible car la PCR est extrêmement sensible et fonctionnera avec une seule molécule de départ. La réaction est démarrée en chauffant le mélange à 94°C. A cette température, les liaisons d'hydrogènes qui maintiennent ensemble les deux polynucléotides de la double hélice sont rompues et l'ADN cible se dénature en molécules simple brin (Fig. 3.3). La température est ensuite abaissée à 50-60°C, ce qui est juste en dessous de la température de recuit. Cela entraîne une certaine réunion des brins simples de l'ADN cible, mais permet également aux amorces de se fixer à leur séquence complémentaire. Normalement, la température est d'environ 2°C en dessous de la température de fusion. La synthèse d'ADN peut maintenant commencer lorsque la température est portée à 72-74°C, juste en dessous de l'optimum pour la Taq polymérase.

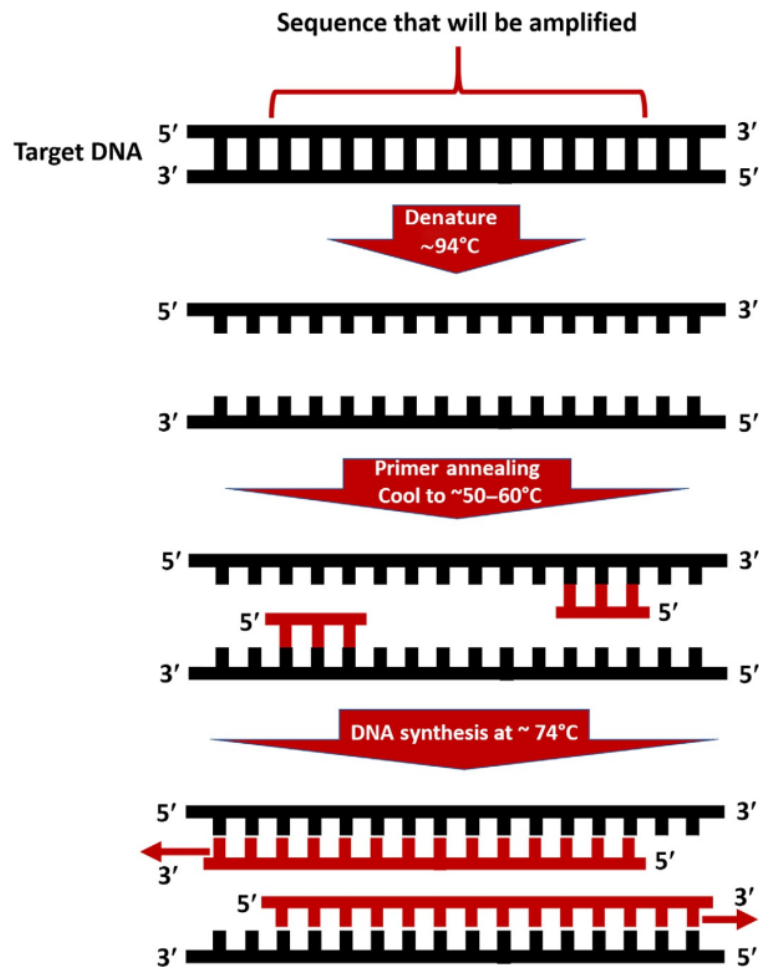


Figure 3.3 Une réaction PCR typique comporte trois étapes : dénaturation, hybridation et extension.

Dans ce premier cycle de PCR, un ensemble de "produits longs" est synthétisé à partir de chaque brin de l'ADN cible. Ces polynucléotides ont des extrémités 5' identiques mais des extrémités 3' aléatoires, ces dernières représentant des positions où la synthèse d'ADN se termine par hasard (Fig. 3.2). Le cycle de dénaturation-hybridation-synthèse est maintenant répété. Les produits longs se dénaturent et les quatre brins résultants sont copiés lors de l'étape de synthèse de l'ADN. Cela donne quatre molécules double brin, dont deux sont identiques aux produits longs du premier cycle et dont deux sont entièrement constituées d'ADN nouveau. Au cours du troisième cycle, les deux ADN nouvellement fabriqués servent de matrice pour produire les produits d'amplification corrects, dont les extrémités 5' et 3' sont toutes deux définies par les positions d'hybridation des amorces. Dans les cycles suivants, le nombre de produits d'amplification corrects s'accumule de façon exponentielle (doublement à chaque cycle) jusqu'à épuisement d'un des composants de la réaction ou jusqu'à la fin d'un cycle défini. Ainsi, après 30 cycles,

3.1.2 Analyse des produits PCR

Pour évaluer si une réaction PCR est réussie, il faut examiner les produits PCR. La procédure la plus simple et la plus simple consiste à analyser les produits de PCR par électrophorèse sur gel d'agarose. Normalement, suffisamment d'ADN ayant été produit à partir du fragment amplifié peut être visible sous la forme d'une bande discrète après coloration avec du bromure d'éthidium. Cela peut en soi fournir des informations utiles sur la région d'ADN qui a été amplifiée, ou si le rendement en ADN est faible, le produit peut être détecté par hybridation Southern. Si la bande attendue est présente, la réaction PCR est réussie. Si la bande attendue est absente, ou si des bandes supplémentaires sont présentes, quelque chose s'est mal passé et l'expérience doit être répétée. Alternativement, le produit de PCR peut être examiné par des techniques telles que le séquençage et le clonage d'ADN.

3.2 Techniques d'amplification des acides nucléiques

3.2.1 Réaction en chaîne par polymérase transcriptase inverse

La PCR de transcriptase inverse (RT) (RT-PCR) a été développée pour amplifier les cibles d'ARN. Dans ce processus, l'ADN complémentaire (ADNc) est d'abord produit à partir de cibles d'ARN par transcription inverse, puis l'ADNc est amplifié par PCR. La transcription inverse est réalisée par l'action de la transcriptase inverse, une enzyme isolée du virus à ARN. Inverse la transcriptase nécessite des amorces telles que des amorces oligo dT ou des hexamères aléatoires pour amorcer la synthèse du brin d'ADN initial (Fig. 3.4). Les amorces oligo dT sont des séquences poly dT monocaténares de 18 bases qui amorceront la synthèse d'ADNc à partir d'ARNm avec des queues poly A. Les hexamères ou décimères aléatoires sont des oligonucléotides simple brin de 6 ou 10 bases de séquences aléatoires. Ces amorces correspondront et s'hybrideront à des sites aléatoires dans l'ARN cible pour amorcer la synthèse d'ADNc. Ces amorces peuvent générer de l'ADNc à partir de tous les ARN de l'échantillon, et la dernière étape est l'amplification de l'ADNc par PCR.

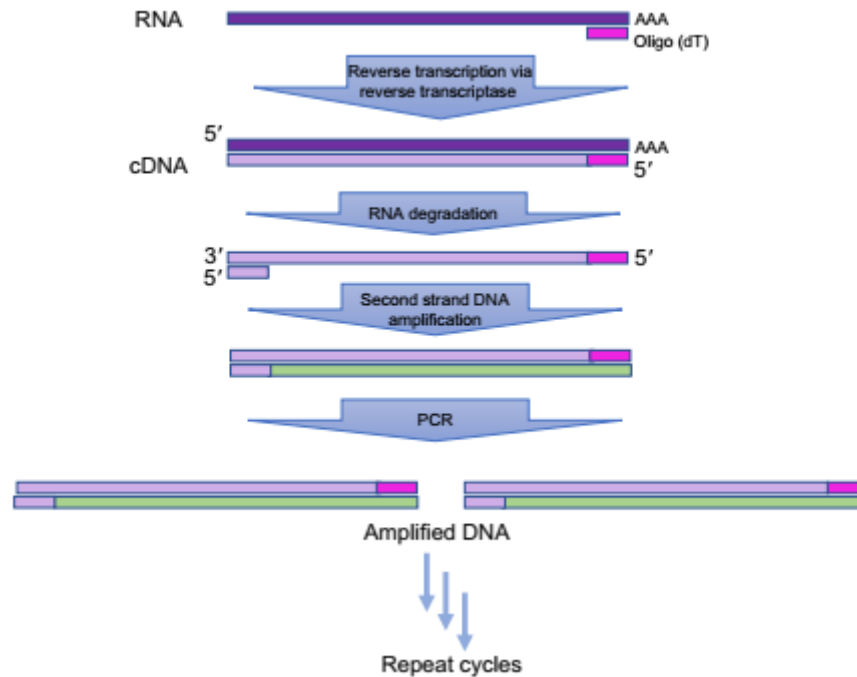


Figure 3.4 Présentation de la PCR de transcription inverse.

3.2.2 Réaction en chaîne par polymérase nichée

La PCR nichée a été développée pour augmenter à la fois la sensibilité et la spécificité de la PCR. Cette technique utilise deux paires d'amorces d'amplification et deux cycles de PCR (Fig. 3.5). En règle générale, une paire d'amorces est utilisée dans le premier tour de l'amplification de la PCR de 15 à 30 cycles. Les produits du premier cycle d'amplification sont ensuite soumis à un second cycle d'amplification en utilisant le second jeu d'amorces. Le second jeu d'amorces s'hybride à une séquence interne à la séquence amplifiée par le premier jeu d'amorces. La sensibilité accrue provient du nombre de cycles total élevé, et la spécificité accrue provient de l'anneau du deuxième jeu d'amorces aux séquences produites par le premier tour. La principale préoccupation de cette méthode est la contamination qui se produit lors du transfert du produit du premier cycle dans le deuxième tube pour le deuxième cycle d'amplification. Cependant, cela peut être évité en séparant physiquement les mélanges d'amplification du premier et du second tour avec une couche de cire ou d'huile.

3.2.3 Réaction en chaîne par polymérase multiplex

Dans la PCR multiplex, deux ensembles d'amorces ou plus conçus pour l'amplification de différentes cibles sont inclus dans la même réaction de PCR. En utilisant

cette technique, plus d'une séquence cible dans un échantillon clinique peut être amplifiée dans un seul tube. En tant qu'extension de l'utilisation pratique de la PCR, cette technique peut économiser du temps et des efforts. Les amorces utilisées dans les réactions multiplex doivent être sélectionnées avec soin pour avoir des températures de recuit similaires et ne doivent pas être complémentaires les uns des autres. Les tailles d'amplicon doivent être suffisamment différentes pour former des bandes distinctes lorsqu'elles sont visualisées par électrophorèse sur gel. La PCR multiplex peut être conçue soit dans une réaction PCR à matrice unique qui utilise plusieurs jeux d'amorces pour amplifier des régions spécifiques dans une matrice, soit dans une réaction PCR à matrices multiples, qui utilise plusieurs matrices et plusieurs jeux d'amorces dans le même tube de réaction (Fig. 3.6).

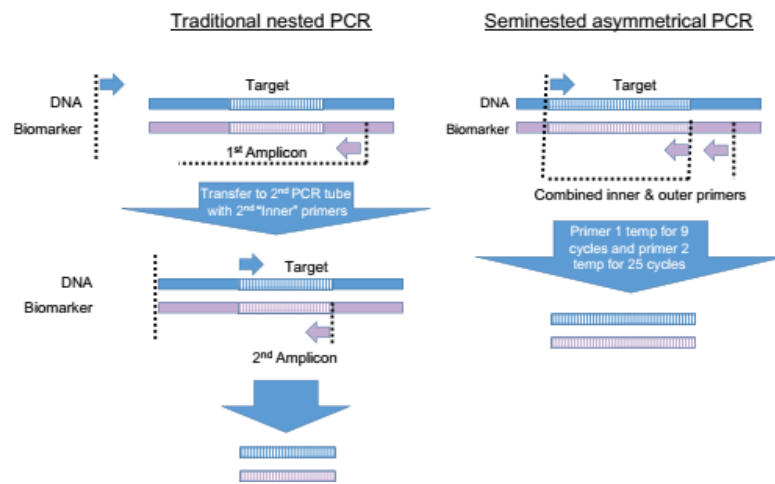


Figure 3.5 PCR nichée versus semi-nichée.

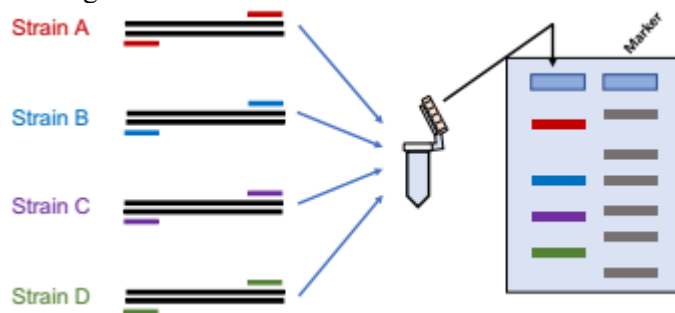


Figure 3.6 Présentation de la PCR multiplex.

3.3 Paramètres affectant la PCR

La longueur de l'amorce et la température d'anneau sont les facteurs les plus importants qui peuvent influencer de manière significative le résultat de la PCR.

3.3.1 La taille du produit PCR

La première étape pour concevoir l'amorce consiste à déterminer la région cible destinée à l'amplification par PCR. Normalement, les séquences cibles à amplifier doivent être maintenues en dessous de 3 kb. Des fragments jusqu'à 10 kb peuvent être amplifiés par des techniques de PCR standard, mais plus le fragment est long, moins l'amplification est efficace.

3.3.2 Les longueurs des amorces

Une fois la région cible déterminée, le problème suivant est la longueur des amorces. Si les amorces sont trop courtes, elles peuvent s'hybrider à des sites non cibles et donner des produits d'amplification indésirables. Bien que des amorces plus longues produisent un produit de PCR plus spécifique que des amorces plus courtes, il n'est pas possible d'avoir des amorces très longues. En effet, la longueur de l'amorce influence la vitesse à laquelle elle s'hybride à l'ADN matrice. Habituellement, les amorces longues s'hybrident à un rythme plus lent.

3.3.3 Température d'hybridation

Pour chaque cycle d'une réaction PCR, le mélange réactionnel PCR est incubé entre trois températures (Fig. 3.7):

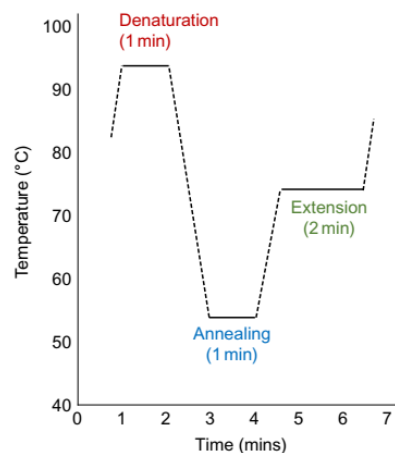


Figure 3.7 Profil de température pour les étapes de PCR.

Pour chaque cycle d'une réaction PCR, le mélange réactionnel PCR est incubé entre trois températures :

1. La température de dénaturation, généralement de 94 °C, qui rompt les liaisons hydrogène entre les paires de bases et libère l'ADN simple brin pour servir de matrice pour le cycle suivant de synthèse d'ADN ;
2. La température d'hybridation ou d'hybridation, généralement comprise entre 50°C et 60°C, à laquelle les amorces se lient aux matrices ;
3. La température d'extension, à laquelle se produit la synthèse d'ADN. Celle-ci est généralement fixée entre 72°C et 74°C, juste en dessous de l'optimum pour la polymérase Taq.

Les températures de dénaturation et d'extension sont généralement constantes dans toute réaction de PCR. Cependant, la température d'anneau dépend des séquences d'amorces et est le facteur critique pour une réaction PCR réussie. En effet, les séquences d'amorces déterminent la température d'anneau et cette température peut affecter la spécificité de la réaction PCR. Si la température de recuit est trop élevée, aucune hybridation n'a lieu, et il s'agit d'une condition de stringence élevée. Par conséquent, dans cette condition, les amorces et les matrices restent dissociées (Fig. 3.8). D'autre part, si la température de recuit est trop basse, des hybrides mal appariés sont produits. Dans cette condition de faible stringence, le nombre de sites d'hybridation potentiels pour chaque amorce est considérablement augmenté, et l'amplification est plus susceptible de se produire sur des sites non cibles dans la molécule matrice.

La température d'hybridation idéale doit être suffisamment basse pour permettre l'hybridation entre l'amorce et la matrice, mais suffisamment élevée pour empêcher la formation d'hybrides incompatibles. Parce que la température de recuit détermine l'hybridation

La T_m peut être déterminée expérimentalement mais est plus généralement calculée à partir de la formule

$$T_m = 81.5^\circ\text{C} + 16.6(\log M) + 0.41(\%GC) - 0.61(\%form) - 500/L$$

dans laquelle M est la concentration de chlorure de potassium dans la réaction, qui est d'environ 25 à 50 mM ; %GC est le pourcentage de nucléotides G et C dans la séquence de l'amorce ; %forme est le pourcentage de formaldéhyde dans la réaction ; et L est la longueur de la séquence d'amorce. Alternativement, le T_m peut être estimé par la formule simple

$$T_m = (4 \times [G + C]) + (2 \times [A + T])^\circ\text{C}$$

dans laquelle $[G + C]$ est le nombre de nucléotides G et C dans la séquence de l'amorce, et $[A + T]$ est le nombre de A et nucléotides T.

Étant donné que la T_m est déterminée par la séquence de l'amorce, il est normal que les deux amorces dans une réaction de PCR soient conçues pour avoir une T_m similaire. La teneur en GC d'un ADN a un effet significatif sur sa T_m : plus la teneur en GC d'un ADN est élevée, plus sa T_m est élevée.

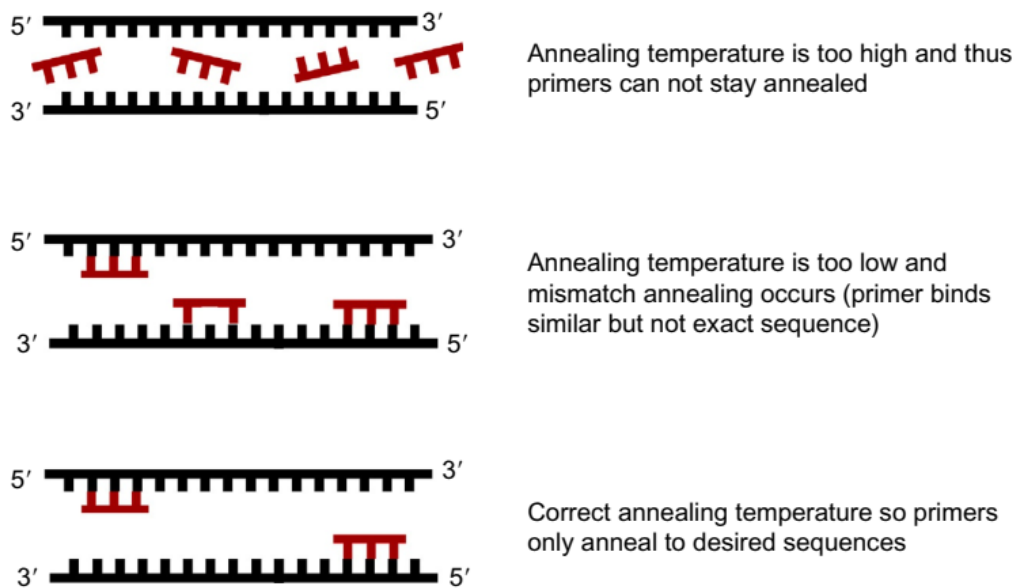


Figure 3.8 L'effet de la température sur l'hybridation de l'amorce 5 en PCR.

3.4 PCR quantitative

La PCR en temps réel ou PCR quantitative (qPCR) est la technologie de quantification des acides nucléiques. Cette méthode est un moyen de quantifier l'amplification d'un ADN au fur et à mesure qu'elle se produit, ce qui signifie qu'elle détecte l'accumulation d'amplicon après chaque cycle thermique en temps réel. Le « temps réel » implique que la collecte et l'analyse des données se produisent au fur et à mesure qu'une réaction se déroule. La concentration initiale de séquence cible peut être déterminée par une comparaison avec la cinétique d'amplification entre la cible et un contrôle interne dont le nombre de copies est connu. (Fig. 3.9). Ainsi, la qPCR peut être utilisée dans l'analyse de l'expression génique et dans la quantification de routine de l'ADN, et devient très populaire dans la recherche et le diagnostic de divers domaines. En raison de son importance, il remplace la PCR conventionnelle. Par exemple, plusieurs nouvelles chimies de détection qPCR ont été développées pour la quantification et le génotypage d'agents

pathogènes, l'expression de gènes, l'analyse d'ADN et de microARN méthylés, la validation de données de puces à ADN, la discrimination allélique et le génotypage (détection de mutations, analyse de polymorphismes mononucléotidiques (SNP) et microsatellites, identification des altérations chromosomiques), validation de l'efficacité de la pharmacothérapie, études médico-légales et quantification des organismes génétiquement modifiés.

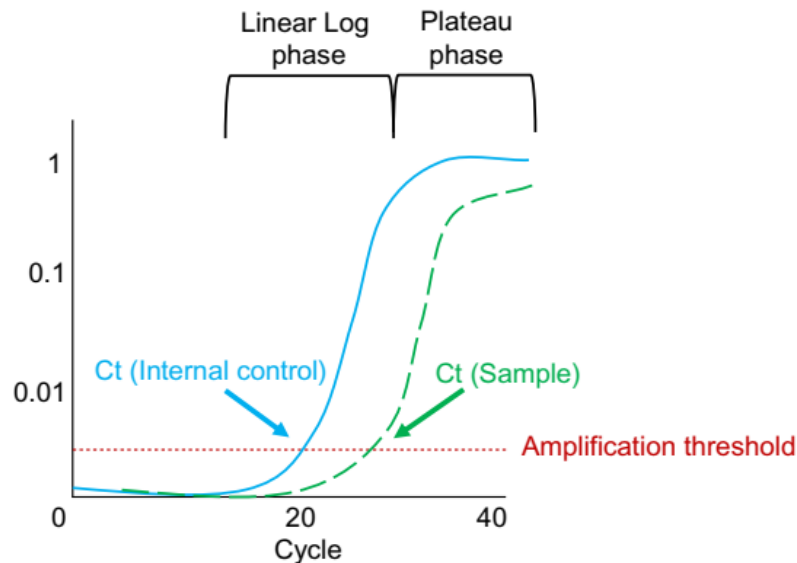


Figure 3.9 Un exemple de tracé d'amplification PCR, où le cycle de seuil (Ct) est le nombre de cycles auquel la fluorescence franchit le seuil d'amplification.

La concentration initiale de la séquence cible peut être déterminée par une comparaison avec le Ct entre la cible et un contrôle interne avec un nombre de copies connu. Le nombre de cycles de PCR nécessaires pour dépasser la fluorescence de fond est appelé le Ct.

4 Techniques de séquençage

4.1 Principe de la procédure didésoxynucléotide

Le séquençage de l'ADN par terminaison de chaîne, également appelé procédure didésoxynucléotide, est basé sur le principe selon lequel, lors de la synthèse de l'ADN, l'ajout d'un nucléotide triphosphate nécessite un groupe hydroxyle libre sur le carbone 3' du sucre du dernier nucléotide du brin d'ADN en croissance (Fig. 4.1). Cependant, si un didésoxynucléotide synthétique dépourvu de groupe hydroxyle au carbone 3' du fragment sucre est incorporé à la fin de la chaîne en croissance, la synthèse d'ADN s'arrête car une liaison phosphodiester ne peut pas être formée avec le prochain nucléotide entrant. En conséquence, la synthèse d'ADN est terminée et c'est la caractéristique déterminante de la méthode de séquençage d'ADN didésoxynucléotidique.

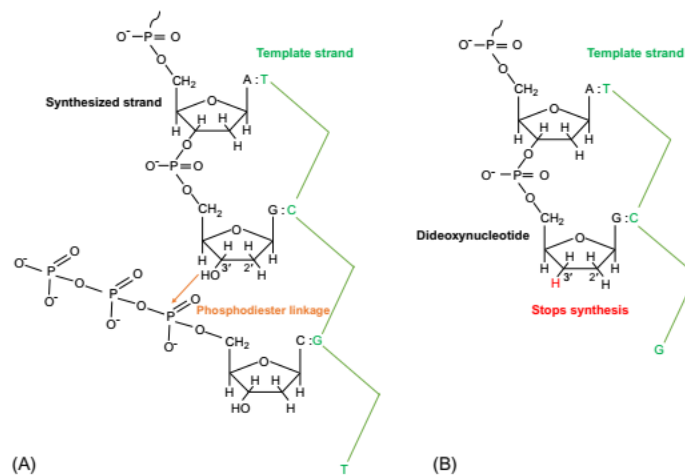


Figure 4.1 Le principe de la terminaison de chaîne de la synthèse de l'ADN. Le 3' OH (A) est remplacé par un H (B), ce qui entraîne l'arrêt de la synthèse d'ADN.

4.1.1 Procédure

Le matériau de départ pour une expérience de séquençage de terminaison de chaîne est une préparation de molécules d'ADN simple brin identiques. La première étape consiste à recuire un oligonucléotide court (17 à 24 nucléotides) à la même position sur chaque molécule. Cet oligonucléotide sert ensuite d'amorce pour la synthèse d'un nouveau brin d'ADN complémentaire de la matrice (Fig. 4.2A). La réaction de synthèse de brin est catalysée par le fragment de Klenow de l'ADN polymérase.

La méthode de Sanger effectue cette synthèse d'ADN dans quatre tubes séparés. Chaque tube nécessite une petite quantité de chacun des quatre didésoxynucléotides (ddNTP - ddATP, ddCTP, ddGTP et ddTTP) et des quatre désoxyribonucléotides triphosphates (dNTP - dATP, dCTP, dGTP et dTTP) comme substrats. L'enzyme de Klenow ne fait pas de distinction entre les désoxy- et les didésoxynucléotides, mais une fois incorporé, un didésoxynucléotide bloque l'élongation supplémentaire car il lui manque le groupe 3'-hydroxyle nécessaire pour former une connexion avec le nucléotide suivant. En tant que tel, un excès de désoxynucléotides normaux doit être utilisé, avec juste assez de didésoxynucléotide pour arrêter l'extension du brin d'ADN de temps en temps au hasard. A cause du hasard fin de la synthèse d'ADN, la réaction produira différentes longueurs de brins d'ADN. Le résultat est une série de fragments de longueurs différentes dans chacun. Par exemple, dans le tube 1 où l'on ajoute ddATP, tous les fragments se terminent par A ; dans le tube 2 où l'on ajoute ddGTP, tous les fragments se terminent par G ; dans le tube 3 où le ddCTP est ajouté, tous les fragments se terminent par C ; et dans le tube 4 où ddTTP est ajouté, tous les fragments se terminent par T (Fig. 4.2B). Par la suite, les quatre mélanges réactionnels sont soumis à une électrophorèse dans des conditions dénaturantes, de sorte que tous les ADN sont à simple brin. Étant donné que le dATP radioactif est inclus dans la réaction, une autoradiographie est effectuée pour visualiser les fragments d'ADN, qui apparaissent sous forme de bandes horizontales sur un film radiographique (Fig. 4.2C). En conséquence, la séquence d'ADN peut être lue à travers chaque bande apparue sur chaque tube.

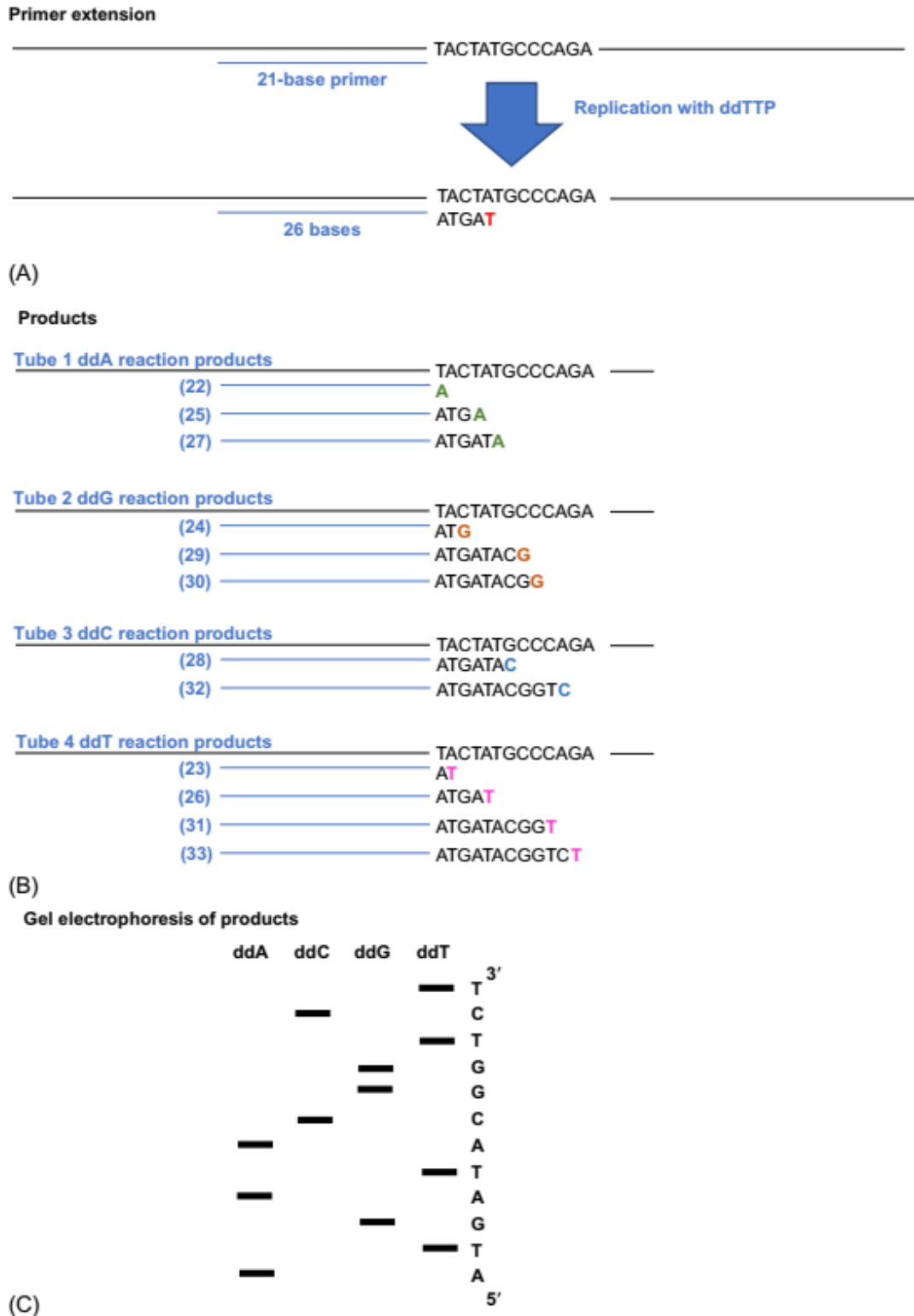


Figure 4.2 La procédure schématique du séquençage de terminaison de chaîne.

- (A) Un exemple typique du séquençage de terminaison de chaîne en ajoutant ddTTP pour arrêter la réaction. (B) Le séquençage de terminaison de chaîne s'effectue dans quatre tubes séparés et chaque tube contient l'un des quatre didésoxynucléotides et les quatre désoxyribonucléotides triphosphates. Les nombres devant chaque chaîne nucléotidique représentent la longueur de cette chaîne nucléotidique correspondante dans chaque réaction. (C) Électrophorèse sur gel du séquençage de terminaison de chaîne pour séparer toutes les chaînes de nucléotides dans chaque tube et révéler la séquence.

4.1.2 ADN polymérase utilisée dans le séquençage

Dans la méthode originale de séquençage de terminaison de chaîne, la polymérase de Klenow a été utilisée comme enzyme de séquençage. Parce que la polymérase de Klenow a une faible processivité, elle ne peut synthétiser qu'un brin d'ADN relativement court avant de se dissocier de la matrice. Cela limite la longueur de la séquence qui peut être obtenue à partir d'une seule expérience à environ 250 pb. Pour résoudre ce problème de synthèse courte, la plupart des séquençages utilisent aujourd'hui une enzyme plus spécialisée, telle que la Sequenase qui est une version modifiée de l'ADN polymérase codée par le bactériophage T7. La séquenase a une processivité élevée et aucune activité exonucléase et est donc idéale pour le séquençage de terminaison de chaîne, permettant d'obtenir des séquences allant jusqu'à 750 pb en une seule expérience.

4.1.3 Modèle de séquençement de terminaison de chaîne

La matrice d'une expérience de terminaison de chaîne est une version simple brin de la molécule d'ADN à séquencer. Une façon d'obtenir de l'ADN simple brin est d'utiliser un vecteur M13. Cependant, le vecteur M13 n'est bon à utiliser que pour les fragments d'ADN de moins de 3 kb. Les vecteurs plasmidiques sont les alternatives, mais certains moyens sont nécessaires pour convertir le plasmide double brin en une forme simple brin. Une façon de convertir les vecteurs plasmidiques en ADN simple brin consiste à utiliser une dénaturation alcaline ou bouillante. Il s'agit d'une méthode courante pour obtenir de l'ADN matrice pour le séquençage de l'ADN. Cependant, une lacune est qu'il peut être difficile de préparer de l'ADN plasmidique qui n'est pas contaminé par de petites quantités d'ADN et d'ARN bactériens, et ces matériaux contaminés peuvent agir comme de fausses matrices ou amorces dans l'expérience de séquençage d'ADN. Une autre approche consiste à cloner l'ADN dans un phagemide, un vecteur plasmidique qui contient une origine de réplication M13 et qui peut donc être obtenu sous forme d'ADN à la fois double et simple brin. Les phagemides peuvent être utilisés avec des fragments jusqu'à 10 kb ou plus.

4.1.4 Séquençage automatisé de l'ADN

Le séquençage automatisé a été développé pour séquencer une très grande quantité d'ADN. Cette procédure utilise le principe de la méthode de terminaison de chaîne de Sanger. Au lieu de marquer le dATP dans la méthode originale de Sanger, chacun des

didésoxynucléotides utilisés dans la réaction est marqué avec un marqueur fluorescent différent.

Une fois les réactions d'extension et la terminaison de chaîne terminées, pour élaborer la séquence d'ADN, le mélange est chargé dans un puits d'un gel de plaque de polyacrylamide ou dans un tube d'un système de gel capillaire, et une électrophorèse est effectuée pour séparer les molécules en fonction de leurs longueurs. Après séparation, les molécules passent devant un détecteur fluorescent capable de discriminer les marqueurs attachés aux didésoxynucléotides (Fig. 4.3). Le détecteur détermine donc si chaque molécule se termine par un A, un C, un G ou un T. Cette séquence peut être imprimée pour examen ou saisie directement dans un dispositif de stockage pour une analyse ultérieure. L'ensemble du processus de séquençage des didésoxynucléotides a été automatisé pour augmenter le taux d'acquisition des données de séquence d'ADN. Ceci est essentiel pour les projets de séquençage à grande échelle tels que ceux impliquant des génomes procaryotes ou eucaryotes entiers.

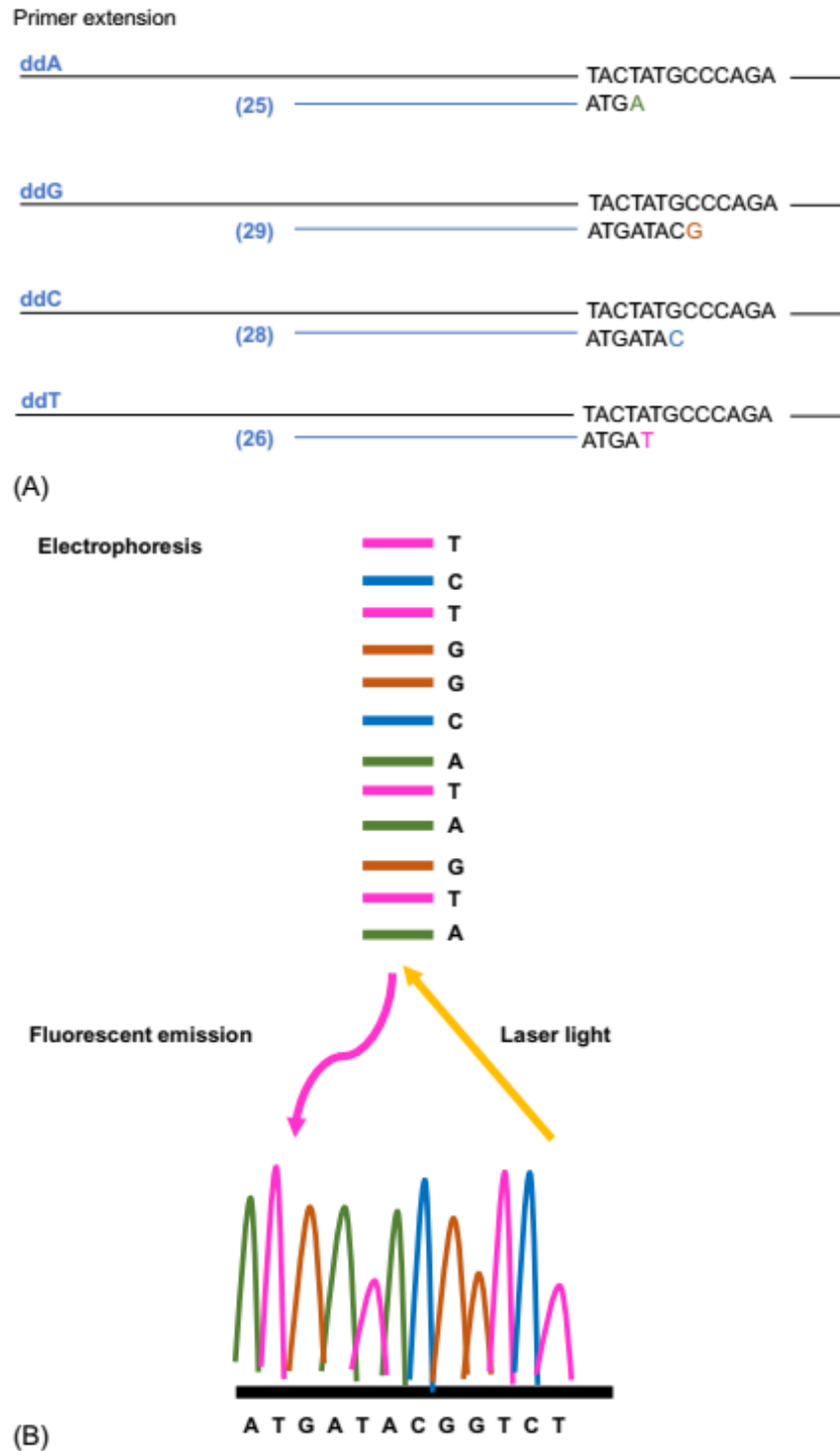


Figure 4.3 Le processus de séquençage automatisé.

La procédure est très similaire à la terminaison de chaîne de Sanger. (A) Chacun des didésoxynucléotides utilisés dans la réaction est marqué avec un marqueur fluorescent différent. Le séquençage de terminaison de chaîne s'effectue dans un tube et ce tube contient quatre didésoxynucléotides marqués par fluorescence.

Les nombres devant chaque chaîne nucléotidique représentent la longueur de cette chaîne nucléotidique correspondante dans chaque réaction. (B) Après les réactions d'extension, le mélange est chargé dans un puits d'un gel de dalle de polyacrylamide, ou dans un tube d'un système de gel capillaire, et une électrophorèse est effectuée pour séparer les molécules en fonction de leurs longueurs. Après séparation, les molécules passent devant un détecteur fluorescent capable de discriminer les marqueurs attachés aux didésoxynucléotides.

4.2 Séquençage à haut débit : Le pyroséquencage

Le pyroséquencage a été la première des technologies de séquençage d'ADN à haut débit (également appelée séquençage de nouvelle génération [NGS]) à être commercialisée et a contribué à la production rapide de grandes quantités de données de séquence par la communauté scientifique. L'avantage du pyroséquencage est qu'il peut être automatisé de manière massivement parallèle, ce qui permet d'obtenir des centaines de milliers de séquences à la fois, peut-être jusqu'à 1000 millions de paires de bases (Mb) en une seule passe. Une séquence est donc produite beaucoup plus rapidement que ne le permet la méthode de terminaison de chaîne.

4.2.1 Principe du pyroséquencage

La base du pyroséquencage est la détection du pyrophosphate qui est libéré lors de la synthèse de l'ADN. Il nécessite la préparation de molécules d'ADN simple brin identiques comme matériau de départ. Ceux-ci peuvent être obtenus par dénaturation alcaline de produits de PCR ou, plus rarement, de molécules plasmidiques recombinantes. Après fixation de l'amorce, la matrice est copiée par une ADN polymérase (généralement un fragment de Klenow) de manière directe sans didésoxynucléotides ajoutés. Lorsqu'un brin d'ADN est allongé par l'ADN polymérase, l' α -phosphate attaché au carbone 5' du sucre d'un désoxynucléotide triphosphate entrant forme une liaison phosphodiester avec le groupe hydroxyle 3' du dernier nucléotide du brin en croissance. Les β - et γ -phosphates terminaux du nucléotide entrant sont libérés sous la forme d'une unité connue sous le nom de pyrophosphate (Fig. 4.4A). La libération du pyrophosphate est en corrélation avec l'incorporation d'un nucléotide spécifique dans le brin d'ADN en croissance. De cette manière, l'ajout d'un désoxynucléotide à l'extrémité du brin en croissance est détectable par la libération d'une molécule de pyrophosphate en temps réel, qui peut être convertie par l'enzyme sulfurylase en un flash de chimiluminescence (Fig. 4.4B).

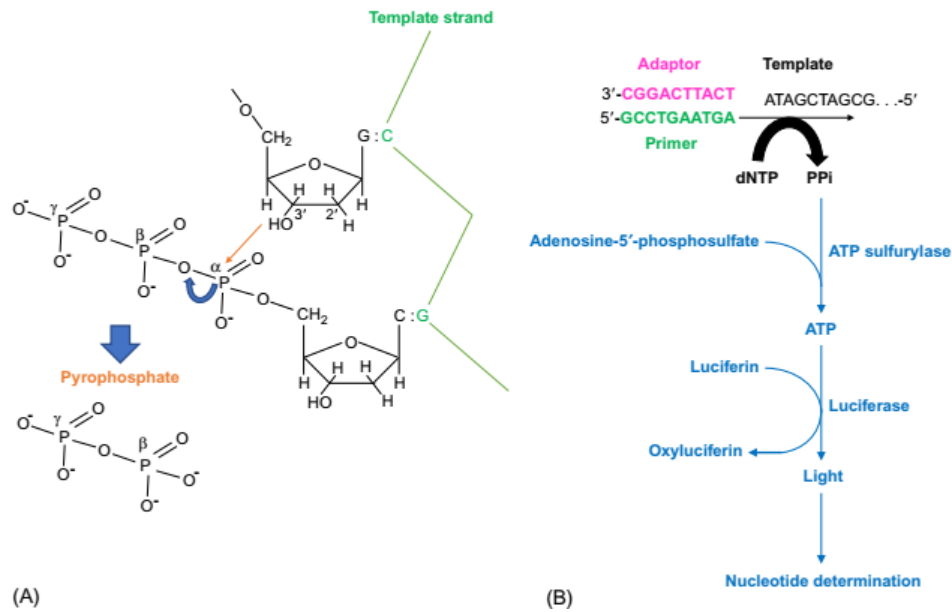


Figure 4.4 Principe du pyroséquençage.

(A) Lorsqu'un brin d'ADN est allongé par l'ADN polymérase, l' α -phosphate attaché au carbone 5' du sucre d'un désoxynucléotide triphosphate entrant forme une liaison phosphodiester avec le groupe hydroxyle 3' du dernier nucléotide du brin en croissance. Les β - et γ -phosphates terminaux du nucléotide ajouté sont clivés en tant qu'unité connue sous le nom de phosphate. (B) La molécule de pyrophosphate peut être convertie par l'enzyme sulfurylase en un éclair de chimiluminescence.

Le pyrophosphate, libéré de l'incorporation d'un nucléotide, se combine avec l'adénosine-5'-phosphosulfate en présence de l'enzyme ATP sulfurylase pour former l'adénosine triphosphate (ATP). À son tour, l'ATP entraîne la conversion de la luciférine en oxyluciférine par l'enzyme luciférase, une réaction qui génère de la lumière. La détection de lumière après chaque cycle d'addition de nucléotide et de réactions enzymatiques indique l'incorporation d'un nucléotide complémentaire. Ainsi, l'émission de lumière est proportionnelle à la quantité de pyrophosphate produite, qui est directement proportionnelle au nombre de nucléotides ajoutés. Qu'un dNTP distribué donné puisse être incorporé ou non, l'apyrase catalyse la dégradation des dNTP en excès avant que le dNTP suivant ne soit distribué. Ainsi, le pyroséquençage ne nécessite pas d'électrophorèse ou toute autre procédure de séparation des fragments et est donc plus rapide que le séquençage par terminaison de chaîne.

Dans le pyroséquençage, chaque désoxynucléotide est ajouté séparément, l'un après l'autre, avec une enzyme nucléotidase également présente dans le mélange réactionnel de sorte que si un désoxynucléotide n'est pas incorporé dans le polynucléotide, il est rapidement dégradé avant l'ajout du suivant (Fig. 4.5). Cette procédure permet de suivre l'ordre d'incorporation des désoxynucléotides dans le brin en croissance. Si les quatre désoxynucléotides étaient ajoutés en même temps, les éclairs de lumière seraient visibles

tout le temps et il serait difficile de déterminer quel désoxynucléotide est ajouté. Par conséquent, cette technique semble compliquée, mais elle nécessite simplement qu'une série répétitive d'ajouts soit faite au mélange réactionnel, précisément le type de procédure qui est facilement automatisée.

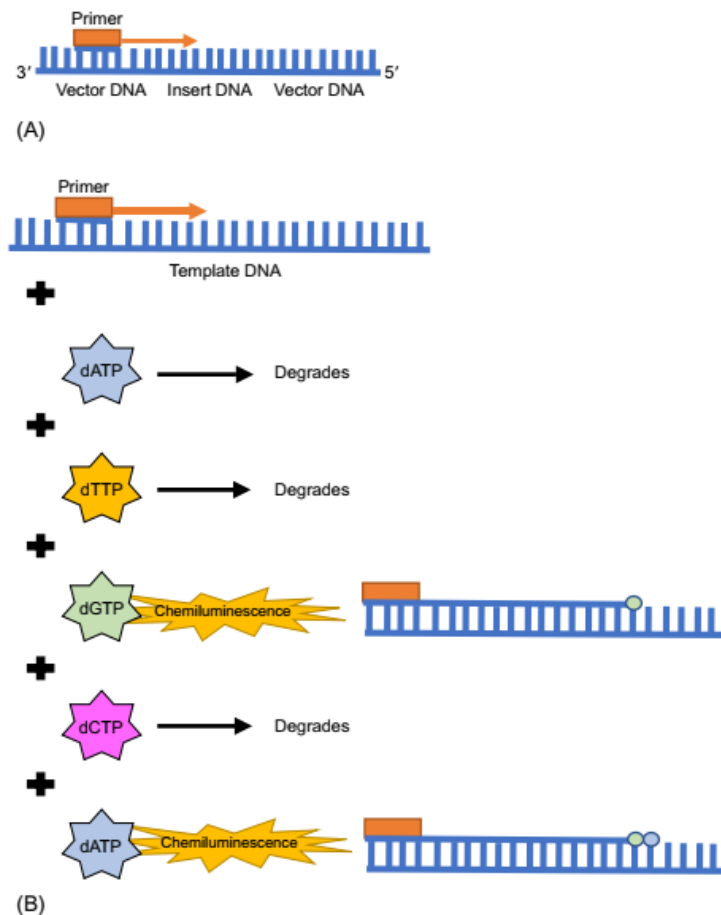


Figure 4.5 Procédure automatisée de pyroséquençage.

(A) Dans le pyroséquençage, si les quatre désoxynucléotides étaient ajoutés en même temps, des éclairs de lumière seraient alors visibles tout le temps et aucune information utile sur la séquence ne serait obtenue. (B) Chaque désoxynucléotide est donc ajouté séparément, l'un après l'autre, avec une enzyme nucléotidase également présente dans le mélange réactionnel de sorte que si un désoxynucléotide n'est pas incorporé dans le polynucléotide, alors il est rapidement dégradé avant que le suivant ne soit ajouté. Cette procédure permet de suivre l'ordre d'incorporation des désoxynucléotides dans le brin en croissance.

4.2.2 Avantage du pyroséquençage

Bien qu'il ne soit capable de générer que jusqu'à 150 pb en une seule expérience, l'avantage du pyroséquençage est qu'il peut être automatisé de manière massivement parallèle, ce qui permet d'obtenir des centaines de milliers de séquences à la fois, peut-être jusqu'à 1 000 Mo dans un seul passage. La séquence est donc produite beaucoup plus rapidement qu'il n'est possible par la méthode de terminaison de chaîne.

5 Les outils bio-informatiques

5.1 Introduction à la Bio-informatique

La bio-informatique a été principalement utilisée à partir de la fin des années 1980 pour désigner l'analyse informatique des données génomiques et décrit plus largement comme « l'étude des processus informatiques dans les systèmes biotiques ». Paulien Hogeweg et Ben Hesper ont utilisé pour la première fois le terme « bio-informatique » au début des années 1970, et le terme a été publié par lui en 1978. De nos jours, du fait des avancées des technologies à haut débit, en termes de fouille et de gestion de données, l'avènement du « Big Data » a posé de nouveaux problèmes, faisant appel à davantage de compétences informatiques dans le domaine. Le Big Data biologique a (et continue d'avoir) des conséquences importantes sur la capacité prédictive et la reproductibilité des résultats bio-informatiques couplés à un nombre toujours croissant d'outils bio-informatiques. Les complémentarités toujours croissantes entre l'informatique et la biologie ont conduit à des sous-disciplines récentes, telles que la biologie synthétique, la biologie des systèmes et la modélisation de cellules entières (Fig.5.1). Pour comprendre le code du vivant qu'est l'ADN, la bio-informatique a développé et contribué à l'essor de projets de recherche associés au séquençage d'ADN à haut débit et à d'autres domaines de la biologie. Son but ultime est de visualiser la richesse des connaissances biologiques cachées dans les séquences, les structures, la littérature et autres méga-données de la biologie.

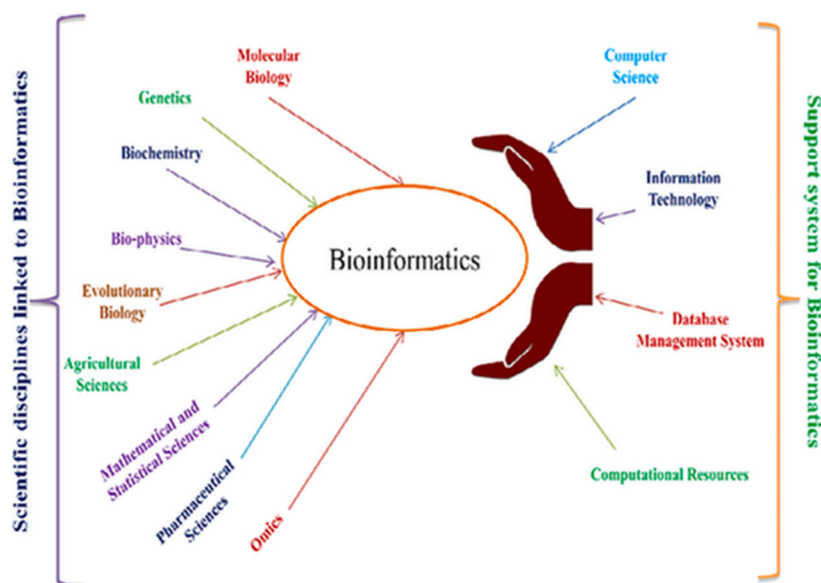


Figure.5.1 Disciplines scientifiques associées à la bio-informatique et à ses systèmes de soutien.

Les techniques bio-informatiques, telles que la simulation, l'amarrage, l'interaction protéine-protéine et l'analyse des données de séquençage de nouvelle génération (NGS), peuvent être utilisées pour étudier ou modifier la séquence afin de mieux adapter les gènes essentiels à une fonction particulière et pour étudier la fonction de ces gènes ou protéines au niveau du système (Tableau 5.1). Il a ensuite été possible d'utiliser ces informations génétiques, génomiques et protéomiques spécifiées pour cultiver des cultures résistantes, améliorées sur le plan nutritionnel et rentables, ainsi que pour découvrir des médicaments thérapeutiques.

Tableau 5.1 Liste des ressources de base de données importantes et populaires pour la bio-informatique.

S. no.	Database	Application	Availability	References
1.	National Centre for Biotechnology Information (NCBI)	It offers access to biomedical and genomic information to boost research activity.	https://www.ncbi.nlm.nih.gov/	Benson, Boguski, Lipman, and Ostell (1990)
2.	GenBank	It is a nucleotide database available at NCBI. It is used for the retrieval of nucleotide sequences.	https://www.ncbi.nlm.nih.gov/genbank/	Benson et al. (2012)
3.	European Nucleotide Archive	It is a nucleotide database available at EBI-EMBL. It provides a detailed record of nucleotide sequencing data, covering raw sequencing data, information about sequence assembly, and functional annotation.	https://www.ebi.ac.uk/ena/browser/home	Leinonen, Sugawara, and Shumway (2010)
4.	DDBJ	It is a nucleotide database available at NIG, Japan. It is used for the retrieval of nucleotide sequences.	https://www.ddbj.nig.ac.jp/index-e.html	Tateno et al. (2002)
5.	Protein Information Resource (PIR)	PIR is an integrated database to support research and scientific studies in genomics, proteomics, and systems biology.	https://proteininformationresource.org/	Wu et al. (2003)
6.	UniProt	UniProt aims to provide a complete, high-quality, and freely available protein sequence and functional information resource to the scientific community.	https://www.uniprot.org/	Apweiler et al. (2004)
7.	Pfam	It is a broad set of protein families database used for domain analysis.	https://pfam.xfam.org/	Bateman et al. (2002)
8.	CATH (Class, Architecture, Topology, and Homology)	The CATH database offers information on the evolutionary relationships of protein domains as a free, publicly accessible online database resource.	https://www.cathdb.info/	Orengo et al. (1997)
9.	SCOP	The purpose of the SCOP database is to provide a detailed and systematic explanation of the structural and evolutionary relationships between proteins deposited in the RCSB Protein Data Bank.	http://scop.mrc-lmb.cam.ac.uk/	Murzin, Brenner, Hubbard, and Chothia (1995)
10.	Protein Data Bank (PDB)	It is a structural database that contains the 3D structure of macromolecules. It is crucial for research in the area of structural bioinformatics, drug discovery and protein structure prediction, etc.	https://www.rcsb.org/	Berman et al. (2000)
11.	PubChem	It is the largest database of chemical information. Here, we search chemical molecules and retrieve their structure for molecular docking/virtual screening.	https://pubchem.ncbi.nlm.nih.gov/	Bolton, Wang, Thiessen, and Bryant (2008)
12.	ZINC	It is a database that contains commercially available molecules for computational screening. We can search and retrieve analogs for any molecule based on similarity. It plays a key role in the discovery of lead molecules for drug development.	https://zinc.docking.org/	Sterling and Irwin (2015)
13.	GEO	GEO is a database for functional genomics. It contains gene expression/microarray data. Here, we can download gene expression profiles submitted by the scientific community throughout the world for further investigation as per need.	https://www.ncbi.nlm.nih.gov/geo/	Edgar, Domrachev, and Lash (2002)

5.2 But de la bio-informatique

La bio-informatique est un besoin de temps à l'ère des omiques, elle est nécessaire pour trouver de nouvelles informations. L'objectif des projets liés à la bio-informatique est généralement :

- Documentation des informations utiles sur les plantes médicinales disponibles dans la littérature/domaines publics sous la forme d'une base de données.
- Documentation des informations disponibles sur les plantes et les ressources génétiques animales.
- Développement de bases de données utiles basées sur les organismes/espèces pour la documentation des données omiques.
- Amélioration du contenu et de l'utilité des bases de données déjà développées/disponibles.
- Développement d'outils basés sur une interface utilisateur graphique (GUI) améliorés et rapides pour l'intégration et l'analyse des données.
- Amélioration des outils/logiciels disponibles pour les biologistes/scientifiques de laboratoire humide/scientifiques non informaticiens pour leur utilisation sans effort.
- Développement de logiciels indépendants de la plate-forme pour l'analyse informatique des données biologiques.

5.3 Les défis de la bio-informatique

Différentes disciplines scientifiques et technologiques sont associées à la bio-informatique. Cette association est également un défi pour comprendre le sujet en raison de sa nature interdisciplinaire.

Les principaux défis dans le domaine de la bio-informatique auxquels sont confrontés les scientifiques issus de la biologie sont :

- Conception d'une petite molécule de type médicament,
- Développement d'un modèle quantitatif de la voie de transduction du signal,
- Prédiction précise de la structure secondaire et tertiaire des protéines à l'aide de la séquence d'acides aminés,
- Comprendre l'évolution des protéines,
- Comprendre l'événement de spéciation,
- Comprendre l'ontologie des gènes,

En outre, les informaticiens travaillant dans le domaine de la bio-informatique sont confrontés à de nombreux défis, tels que :

- Gestion du big data en biologie,
- Gestion de l'information,
- Améliorer la précision des logiciels, et
- Programmabilité.

5.4 Application de la bio-informatique

La bio-informatique a un grand potentiel pour résoudre des problèmes complexes en biologie et dans les disciplines associées. Avec le démarrage du projet du génome humain, il devient un outil essentiel pour les scientifiques et joue un rôle clé dans le domaine des sciences de la vie, des sciences chimiques, des sciences physiques, des sciences agricoles et des sciences médicales (Fig. 5.2). Certaines applications importantes de la bio-informatique sont discutées dans les sections suivantes.

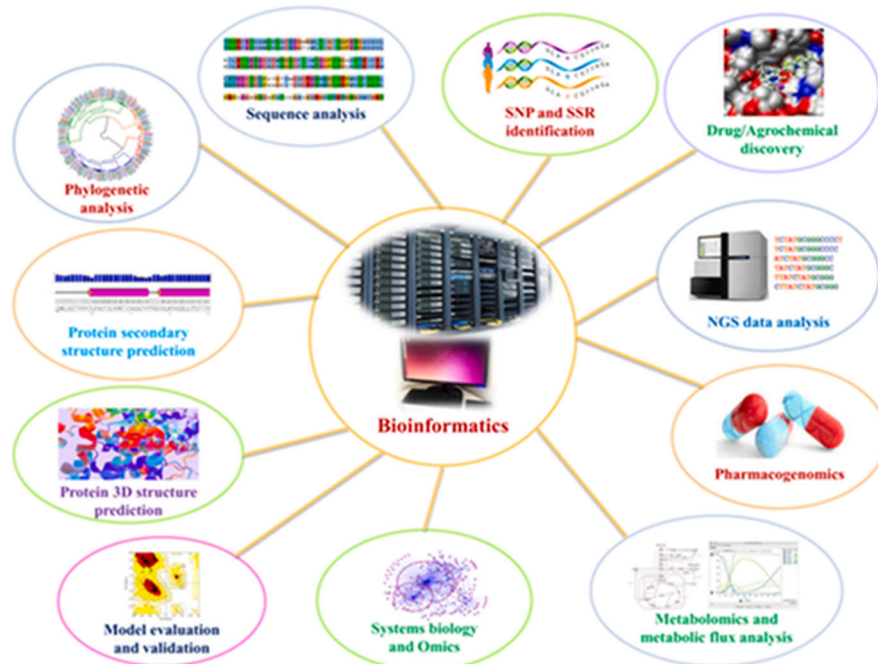


Figure 5.2 Application de la bio-informatique dans différents domaines des sciences biologiques, agricoles et médicales.

5.4.1 Analyse de séquence

L'analyse de séquence est l'une des principales applications de la bio-informatique avec le développement du programme Basic Local Alignment Search Tool (BLAST) en 1990 et est devenue populaire. Le domaine de l'analyse de séquences est très large ; ici, nous analysons la séquence nucléotidique ou protéique de tout organisme à plusieurs fins. Ici, nous pouvons analyser des séquences uniques ou multiples pour découvrir la similitude et l'identité entre elles via plusieurs outils d'alignement de séquences, tels que BLAST et FASTA Clustal. Il est également utilisé dans l'annotation de séquences nouvellement découvertes, pour découvrir des régions conservées et d'autres régions régulatrices parmi elles. Par ailleurs, la prédiction des propriétés physicochimiques des séquences relève également de l'analyse de séquences.

5.4.2 Analyse phylogénétique

L'analyse phylogénétique est un autre domaine de recherche important en bio-informatique. Ici, nous pouvons visualiser l'événement évolutif et construire une relation entre les organismes ou les séquences. Il est largement utilisé en biologie évolutive dans le but de déterminer l'événement évolutif via un alignement de séquences multiples suivi d'une construction d'arbres. Il prend également en charge l'identification des régions clés dans les séquences et joue un rôle important dans les programmes de conception de vaccins et de médicaments, etc.

5.4.3 Évaluation et validation du modèle de protéines prédites

Diverses méthodes de calcul sont développées par des scientifiques en bioinformatique pour l'analyse de la qualité et la validation du modèle 3D de protéine prédit. Le programme le plus populaire utilisé pour l'évaluation et la validation des modèles est Swiss PDB viewer, Structural Analysis and Verification Server (SAVES), Rampage et PROCHECK. Nous pouvons évaluer et affiner notre modèle prédit à l'aide des outils ci-dessus pour une enquête plus approfondie.

5.4.4 Découverte et conception de petites molécules menant au développement de médicaments/agrochimiques

La bio-informatique joue un rôle essentiel dans la découverte de molécules de plomb. Des techniques telles que l'amarrage moléculaire, le criblage virtuel et la simulation de dynamique moléculaire sont largement utilisées à cette fin. Grâce à l'amarrage moléculaire et au criblage virtuel, nous pouvons identifier de nouvelles molécules pour le traitement et la prévention de maladies chez l'homme, les animaux et les plantes. De plus, il peut être validé en termes de comportement conformationnel et de stabilité lors de l'interaction cible-ligand par rapport au temps à l'aide d'une étude de simulation de dynamique moléculaire. Cette étude aide à l'identification rapide du composé principal à l'aide d'outils informatiques et réduit le temps et le coût expérimental, le processus de découverte rapide de médicaments pour la validation en aval.

5.4.5 Analyse des données de séquençage de nouvelle génération

En raison des progrès et du développement récent de plusieurs plateformes omiques et de diverses technologies de séquençage, une grande quantité de données est générée

chaque jour et nécessite la bio-informatique pour leur analyse et leur gestion. Ces dernières années, un ensemble d'outils et de bases de données ont été développés pour la gestion, l'intégration et l'analyse de ces mégadonnées. Ces outils sont utiles dans l'identification de gènes exprimés de manière différentielle via l'analyse de données de puces à ADN et de RNAseq/transcriptome, l'assemblage et l'annotation du génome et du transcriptome, la découverte de SNP, l'étude d'association à l'échelle du génome, l'identification et la caractérisation de familles de gènes, la découverte de nouveaux gènes, etc. Les informations obtenues à partir de l'analyse des mégadonnées aideront dans différents programmes de recherche associés à la sélection végétale et animale, au développement de marqueurs, à l'identification de cibles médicamenteuses.

5.4.6 Identification SNP et SSR

Les progrès en NGS et en bio-informatique ont permis la découverte à grande échelle de marqueurs SNP et SSR. Il a révolutionné la recherche liée à la génomique et aide au programme de sélection moléculaire, et il aidera également à analyser la diversité génétique et la structure de la population, à concevoir des cartes génétiques de haute densité et à fournir des génotypes pour la recherche d'association à l'échelle du génome.

5.4.7 Pharmacogénomique

À l'ère postgénomique, la pharmacogénomique est reconnue comme l'un des domaines de recherche clés et très demandés pour étudier la réponse d'un médicament basé sur des variants génétiques pour le développement de la médecine personnalisée. Comme nous le savons, certains médicaments agissent sur une population et/ou un groupe de personnes en particulier, mais aucune réponse souhaitée de ces médicaments n'a été signalée dans d'autres populations ou pays en raison de changements infimes dans les régions génomiques. Ainsi, le concept de pharmacogénomique a été introduit pour le développement d'un médicament spécifique à une personne ou à une population ou à une région afin de réduire le risque d'effets secondaires et d'augmenter leur puissance. Par conséquent, les outils bioinformatiques sont rapidement utilisés pour l'analyse des données génomiques à haut débit pour la détection de la façon dont les gènes modifient la réponse d'un médicament, cette enquête mènera au développement de la médecine personnalisée.

5.4.8 Métabolomique et analyse des flux métaboliques

Une compréhension quantitative et qualitative des taux ou flux métaboliques sur un réseau métabolique et en particulier des compartiments cellulaires offre un aperçu de la régulation métabolique et aide à comprendre la contribution des changements métaboliques à la pathologie. La bioinformatique aide les chercheurs à analyser les données métabolomiques et à développer un modèle métabolique pour l'analyse de l'équilibre des flux. Cela aidera le programme d'ingénierie métabolique à diverses fins.

5.4.9 Biologie des systèmes et intégration de données omiques

De nos jours, la biologie des systèmes est l'un des domaines essentiels de la bioinformatique et a des contributions exceptionnelles dans le décodage de la complexité des systèmes biologiques via la modélisation, la simulation et l'analyse de réseau. Cela nous aide à intégrer différentes données omiques, c'est-à-dire la génomique, la protéomique, la métabolomique, la transcriptomique et d'autres omiques, pour le développement d'un modèle qui décrira le comportement des systèmes biologiques dans différentes conditions de manière holistique par rapport au temps.

Au cours des dernières années, les scientifiques travaillaient sur un seul gène ou protéine, mais ces dernières années, le paradigme a été le passage d'une approche réductionniste à une approche holistique en raison de l'augmentation de la technologie omique et de la biologie des systèmes. Ainsi, avec l'aide de la biologie des systèmes, nous pouvons prédire le comportement de systèmes entiers et identifier les composants clés impliqués dans différents processus biologiques grâce à la modélisation et à l'analyse des réseaux.

1. Aberg, K.A., Xie, L.Y., Nerella, S., Copeland, W.E., Costello, J., van den Oord, E.J.C.G., 2013. High quality methylome-wide investigations through next-generation sequencing of DNA from a single archived dry blood spot. *Epigenetics* 8 (5), 542–547.
2. Ali, I., Aboul-Enein, H.Y., Singh, P., Singh, R., Sharma, B., 2010. Separation of biological proteins by liquid chromatography. *Saudi Pharm. J.* 18, 59–73.
3. Allison, L.A., 2007. *Fundamental Molecular Biology*, first ed. Wiley-Blackwell. ISBN-10: 1405103795; 13: 978-1405103794.
4. Aranda, P.S., LaJoie, D.M., Jorcyk, C.L., 2012. Bleach gel: a simple agarose gel for analyzing RNA quality. *Electrophoresis* 33 (2), 366–369. Asiello, P.J., Baeumner, A.J., 2011. Miniaturized isothermal nucleic acid amplification, a review. *Lab Chip* 11, 1420–1430.
5. Asim, A., Kumar, A., Muthuswamy, S., Jain, S., Agarwal, S., 2015. Down syndrome: an insight of the disease. *J. Biomed. Sci.* 22 (1), 41.
6. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031.
7. Boesenberg-Smith, K.A., Pessarakli, M.M., Wolk, D.M., 2012. Assessment of DNA yield and purity: an overlooked detail of PCR troubleshooting. *Clin. Microbiol. Newsl.* 34, 1.
8. Brown, T.A., 2016. *Gene Cloning and DNA Analysis: An Introduction*, seventh ed. Wiley-Blackwell. ISBN: 978-1-119-07256-0.
9. Butterfield, D.A., Gu, L., Di Domenico, F., Robinson, R.A., 2014. Mass spectrometry and redox proteomics: applications in disease. *Mass Spectrom. Rev.* 33 (4), 277–301.
10. Cole, K.D., Tellez, C.M., 2002. Separation of large circular DNA by electrophoresis in agarose gels. *Biotechnol. Prog.* 18, 82–87.
11. Feist, P., Hummon, A.B., 2015. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int. J. Mol. Sci.* 16, 3537–3563.
12. Gauthier, J., Vincent, A.T., Charette, S.J., Derome, N., 2019. A brief history of bioinformatics. *Briefings in Bioinformatics*, 20(6),1981–1996.
13. Gullett, J.C., Nolte, F.S., 2015. Quantitative nucleic acid amplification methods for viral infections. *Clin. Chem.* 61 (1), 72–78.
14. Gustafsson, O.J.R., Arentz, G., Hoffmann, P., 2015. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim. Biophys. Acta: Proteins Proteomics* 1854 (6), 559–580.
15. Harrington, C.T., Lin, E.I., Olson, M.T., Eshleman, J.R., 2013. Fundamentals of pyrosequencing. *Arch. Pathol. Lab. Med.* 137, 1296–1303.
16. Holden, M.J., Haynes, R.J., Rabb, S.A., Satija, N., Yang, K., Blasic Jr., J.R., 2009. Factors affecting quantification of total DNA by UV spectroscopy and PicoGreen fluorescence. *J. Agric. Food Chem.* 57, 7221–7226.
17. Javadi, A., Shamaei, M., Mohammadi Ziazi, L., Pourabdollah, M., Dorudinia, A., Seyedmehdi, S.M., Karimi, S., 2014. Qualification study of two genomic DNA extraction methods in different clinical samples. *Tanaffos* 13 (4), 41–47.
18. Jiang, Z., Wang, H., Michal, J.J., Zhou, X., Liu, B., Woods, L.C., Fuchs, R.A., 2016. Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. *Int. J. Biol. Sci.* 12, 100–108.
19. Kunkel, T.A., 2004. DNA replication fidelity. *J. Biol. Chem.* 279 (17), 16895–16898.
20. Kurien, B.T., Hal Scofield, R. (Eds.), 2012. *Protein Electrophoresis: Methods and Protocols*, *Methods in Molecular Biology*. vol. 869. https://doi.org/10.1007/978-1-61779-821-4_3.
21. Magdeldin, S., Enany, S., Yoshida, Y., Xu, B., Zhang, Y., Zureena, Z., Lokamani, I., Yaoita, E., Yamamoto, T., 2014. Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis. *Clin. Proteomics* 11 (1), 16.
22. Miller, J.R., Andre, R., 2014. Quantitative polymerase chain reaction. *Br. J. Hosp. Med.* 75 (12), c188.
23. Pegadaraju, V., Nipper, R., Hulke, B., Qi, L., Schultz, Q., 2013. De novo sequencing of sunflower genome for SNP discovery using RAD (restriction site associated DNA) approach. *BMC Genomics* 14, 556.
24. Posch, A., 2014. Sample preparation guidelines for two dimensional electrophoresis. *Arch. Physiol. Biochem.* 120 (5), 192–197. Powers-Fletcher, M.V., Hanson, K.E., 2016. Molecular diagnostic testing for *Aspergillus*. *J. Clin. Microbiol.* 54, 2655–2660.

25. Raynal, B., Lenormand, P., Baron, B., Hoos, S., England, P., 2014. Quality assessment and optimization of purified protein samples: why and how? *Microb. Cell Fact.* 13, 180.
26. Rhoads, A., Au, K.F., 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13, 278–289.
27. Turan, C., Nanni, I.M., Brunelli, A., Collina, M., 2015. New rapid DNA extraction method with Chelex from *Venturia inaequalis* spores. *J. Microbiol. Methods* 115, 139–143.
28. Wang, Y., Hayatsu, M., Fujii, T., 2012. Extraction of bacterial RNA from soil: challenges and solutions. *Microbes Environ.* 27 (2), 111–121.
29. Weaver, R.F., 2011. *Molecular Biology*, fifth ed. McGraw-Hill Education. ISBN-10: 0073525324; 13: 978-0073525327.
30. Wisdom, G.B., 1994. Protein blotting. *Methods Mol. Biol.* 32, 207–213.
31. Xu, Z., Geng, Q., Luo, F., Xu, F., Li, P., Xie, J., 2014. Multiplex ligation-dependent probe amplification and array comparative genomic hybridization analyses for prenatal diagnosis of cytogenomic abnormalities. *Mol. Cytogenet.* 7, 84.
32. Zhang, M., Liu, K., Hu, Y., Lin, Y., Li, Y., Zhong, P., Jin, X., Zhu, X., Zhang, C., 2017. A novel quantitative PCR mediated by high-fidelity DNA polymerase. *Sci. Rep.* 7 (1), 10365.
33. Zhu, W., Zhang, X.Y., Marjani, S.L., Zhang, J., Zhang, W., Wu, S., Pan, X., 2017 Mar. Next-generation molecular diagnosis: single-cell sequencing from bench to bedside. *Cell. Mol. Life Sci.* 74 (5), 869–880.