



**HAL**  
open science

# Technologie $\mu$ LAS pour l'analyse et la purification d'ADN de haut poids moléculaire

Nicolas Milon

► **To cite this version:**

Nicolas Milon. Technologie  $\mu$ LAS pour l'analyse et la purification d'ADN de haut poids moléculaire. Biochimie [q-bio.BM]. INSA de Toulouse, 2019. Français. NNT : 2019ISAT0005 . tel-04544743

**HAL Id: tel-04544743**

**<https://theses.hal.science/tel-04544743>**

Submitted on 13 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE

## En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de  
Toulouse

---

Présentée et soutenue par

**Nicolas MILON**

Le 19 mars 2019

**Technologie  $\mu$ LAS pour l'analyse et la purification d'ADN de  
haut poids moléculaire**

---

Ecole doctorale : **SDM - SCIENCES DE LA MATIERE - Toulouse**

Spécialité : **Chimie-Biologie-Santé**

Unité de recherche :

**LAAS - Laboratoire d'Analyse et d'Architecture des Systèmes**

Thèse dirigée par  
**Aurélien BANCAUD**

Jury

M. Hervé COTTET, Rapporteur  
M. Vincent DION, Rapporteur  
M. François COUDERC, Examineur  
Mme Adriana ALBERTI, Examineur  
Mme Hélène BERGÈS, Examineur  
M. Aurélien BANCAUD, Directeur de thèse



# Remerciements

---

Une grande partie de cette thèse a été effectuée dans l'entreprise Adelis (ex Picometrics) dirigée successivement par Jean-Charles Garcia et Philippe Pourquoier, que je remercie pour m'avoir accueilli au sein de leur société.

Je remercie Hervé Cottet et Vincent Dion d'avoir accepté d'être les rapporteurs de cette thèse et pour leur retour constructif. Merci aux membres du jury pour leur présence et leurs commentaires à la soutenance : Hélène Bergès, Adriana Alberti et François Couderc.

Si ces trois années de thèse se sont aussi bien passées, c'est principalement parce que j'ai eu la chance de travailler dans un excellent environnement, à la fois du côté industriel et du côté académique, et de bénéficier d'un encadrement de grande qualité.

Une mention particulière à mon directeur de thèse, Aurélien Bancaud. Merci d'avoir encadré cette thèse avec brio, pour m'avoir aiguillé tout en me laissant une certaine liberté sur l'orientation du sujet. Merci pour ton enthousiasme et ton optimisme à toute épreuve, cela rend le travail bien plus motivant et permet d'éviter les coups de mou qui accompagnent généralement une thèse. Merci d'avoir toujours eu la porte ouverte, pour des discussions scientifiques ou pas... et d'avoir fait semblant de t'intéresser au tennis durant plus de 4 ans !! Si c'est 3 années de thèse furent aussi agréables c'est en grande partie grâce à toi.

Un grand merci également à Frédéric Ginot pour avoir encadré cette thèse du côté industriel. Merci pour ta disponibilité et ton professionnalisme. Merci d'avoir partagé tes connaissances et ton expérience, jamais à court d'idées ou hypothèses pour expliquer les résultats, parfois surprenant, obtenus durant cette thèse.

Les résultats présentés dans ce manuscrit n'auraient pu voir le jour sans l'aide de nos collaborateurs. Un grand merci à Hélène Bergès et au CNRGV pour cette collaboration fructueuse. Je voudrais remercier plus particulièrement Céline Chantry-Darmon pour sa disponibilité et sa forte implication dans ce projet, ce fut un plaisir de travailler avec toi. Merci également à Margaux Fustier pour l'analyse des données de séquençage.

Je remercie la plate-forme de séquençage de Toulouse (GeT-PlaGe) pour avoir participé avec enthousiasme au développement de notre instrument et pour avoir fourni les échantillons sur lesquels nous avons travaillé. Merci à Olivier Bouchez et Cécile Donnadiou pour cette collaboration, à Adrien Castinel pour les expériences sur le Chromium, ainsi qu'à Sophie et Adeline pour leur expertise sur le BluePippin.

Je souhaite remercier particulièrement toute l'équipe de Picometrics pour leur participation à ces travaux et pour la bonne humeur qui règne dans l'entreprise. Arnaud, sans qui Picometrics aurait du mal à fonctionner, pour son professionnalisme et son sérieux à toute épreuve. Jacques pour sa bonne humeur quotidienne. Louis pour avoir confectionné avec soin les dispositifs utilisés durant cette thèse

et pour ses vastes discussions aux pauses café. Audrey pour sa disponibilité et pour avoir partagé ses connaissances sans limite en électrophorèse capillaire. Laure pour avoir encadrer le début de cette thèse. Marion, Vincent, My-Dung, Christine et les employés de Sycomores pour les bons moments partagés autour d'une table ou sur un terrain de foot.

Je remercie chaleureusement l'équipe MILE du LAAS pour m'avoir accueilli durant ces 3 années. Bien que ma présence au laboratoire fût limitée, j'ai pris un réel plaisir à travailler avec vous. Merci donc à Bayan, Jeffrey, Inga, Anaïs, Marianne, Jean, Marion, Léo, Gabriel, Amani, Lyne, Pierre, Morgan... Je remercie également Rémi pour ces batailles épiques au ping-pong, en espérant que tu ne me tiennes pas trop rigueur de toutes ces défaites infligées... Merci également aux différents stagiaires qui ont participé à ce projet, Juan en particulier pour son excellent travail. Le LAAS m'a également permis d'évoluer dans un excellent environnement et de bénéficier d'un soutien technique de grande qualité. Merci donc à Charline pour sa disponibilité en salle bio et à Xavier pour la fabrication de nos dispositifs.

Le LAAS c'est aussi une vie associative très riche. Je remercie le capitaine de l'équipe de tennis, Philippe Menini, pour m'avoir permis de défendre haut les couleurs du LAAS au cours de ces campagnes (parfois) victorieuses. Merci à toute l'équipe : Rémi, Laurent, Alex, Matthieu, Philippe et Christian pour ces exploits sportifs partagés.

Je n'oublie pas les personnes qui rendent la vie si agréable en dehors du labo. Merci aux copains de l'INSA, du lycée ou même d'avant, toujours disponibles pour un verre, un tennis ou pour partir à l'aventure. Merci, Charlotte, de partager ma vie et pour tout ce que tu m'apportes au quotidien. Merci enfin à ma famille et particulièrement mes parents, pour leur soutien, pour m'avoir poussé dans cette voie et m'avoir permis d'évoluer sereinement durant toutes ces années. Merci pour tous les bons moments passés (et à venir) ensemble.

# Table des matières

---

<b>Table des matières .....</b>	<b>5</b>
<b>Lexique .....</b>	<b>9</b>
<b>Introduction.....</b>	<b>11</b>
<i>Contexte .....</i>	11
<i>Organisation du manuscrit.....</i>	12
<b>1. Technologies pour la génomique de 3ème génération.....</b>	<b>13</b>
1.1. Les 1 <sup>ère</sup> et 2 <sup>ème</sup> générations de séquençage .....	13
1.1.1 Le séquençage de Sanger .....	13
1.1.2 Le séquençage NGS .....	14
1.1.3 Pourquoi la nécessité d’une 3 <sup>ème</sup> génération ? .....	14
1.2 Séquençage SMRT – PacBio.....	18
1.2.1 Principe de fonctionnement.....	18
1.2.2 Limitations .....	20
1.2.3 Préparation et caractéristiques de l’échantillon à séquencer.....	21
1.3 Séquençage Nanopore – ONT .....	22
1.3.1 Principe de fonctionnement.....	22
1.3.2 Limitations .....	24
1.3.3 Préparation et caractéristiques de l’échantillon à séquencer.....	25
1.4 Technologie Chromium – 10X Genomics .....	26
1.4.1 Principe de fonctionnement.....	26
1.4.2 Préparation et caractéristiques de l’échantillon à séquencer.....	27
1.5 Cartographie optique .....	29
1.5.1 Principe de fonctionnement.....	29
1.5.2 Préparation et caractéristiques de l’échantillon .....	31

1.6	Conclusion du chapitre 1 .....	32
<b>2.</b>	<b>Préparation et contrôle qualité d'ADN de haut poids moléculaire .....</b>	<b>35</b>
2.1	Extraction d'ADN de haut poids moléculaire .....	36
2.1.1	Purification liquide et chimique .....	37
2.1.1.1	Phénol-Chloroforme .....	37
2.1.1.2	Gradient de Chlorure de Césium (CsCl) .....	38
2.1.2	Immobilisation réversible sur phase solide (SPRI) .....	38
2.1.2.1	Purification sur colonne.....	38
2.1.2.2	Billes magnétiques.....	39
2.1.2.3	Place de la microfluidique en SPRI .....	41
2.1.3	Extraction en « plugs » d'agarose .....	42
2.1.4	Purification électrophorétique .....	43
2.1.4.1	SCODA.....	43
2.1.4.2	Isotachophorèse .....	44
2.1.5	Comparaison des méthodes d'extraction .....	45
2.2	Sélection en taille d'ADN de haut poids moléculaire .....	46
2.2.1	Billes magnétiques.....	46
2.2.2	Gels d'agarose .....	46
2.2.2.1	Nimbus.....	47
2.2.2.2	BluePippin.....	47
2.3	Contrôle qualité d'ADN de haut poids moléculaire .....	50
2.3.1	Pureté de l'échantillon – Mesure d'absorbance .....	50
2.3.2	Concentration de l'échantillon – mesures fluorimétriques .....	51
2.3.3	Distribution en taille – Techniques de séparation d'ADN .....	52
2.3.3.1	Les gels d'agarose.....	52
2.3.3.2	L'électrophorèse capillaire .....	53
2.3.3.3	Séparation d'ADN en systèmes microfluidiques .....	54

2.3.3.4	Méthodes chromatographiques.....	56
2.4	Conclusion du Chapitre 2 .....	58
<b>3.</b>	<b>μLAS pour la manipulation d'ADN de haut poids moléculaire .....</b>	<b>61</b>
3.1	Principe de fonctionnement de μLAS.....	62
3.1.1	La séparation sans matrice solide .....	62
3.1.2	Migration transverse viscoélastique .....	63
3.1.3	La séparation par μLAS.....	65
3.1.4	Caractéristiques du tampon .....	69
3.1.5	Mise en œuvre de la séparation par μLAS .....	69
3.1.5.1	Format capillaire.....	69
3.1.5.2	Format puce microfluidique .....	70
3.2	Séparation d'ADN de haut poids moléculaire par μLAS.....	72
3.2.1	Optimisation des conditions opératoires en capillaire .....	72
3.2.2	Vers des séparations rapides en puces microfluidiques .....	74
3.2.2.1	Optimisation de la géométrie des puces.....	74
3.2.2.2	Une limite de séparation vers 100 kb.....	76
3.2.2.3	Anomalies de migration en puces microfluidiques .....	77
3.2.3	Comparaison des formats puce et capillaire.....	78
3.3	La concentration par μLAS.....	80
3.3.1	Principe de fonctionnement.....	80
3.3.2	Dimensions du dispositif .....	81
3.3.3	Analyse haute sensibilité de l'ADN de haut poids moléculaire .....	82
3.3.4	Comparaison des performances avec l'état de l'art .....	84
3.4	Conclusion du Chapitre 3 .....	88



<b>4. Capture spécifique de longs fragments d'ADN via un couplage entre la technique CRISPR/Cas9 et la technologie <math>\mu</math>LAS .....</b>	<b>89</b>
4.1 Conclusion du Chapitre 4 .....	114
<b>5. Développement d'une vanne accordable pour la sélection en taille d'ADN .....</b>	<b>115</b>
5.1 Conclusion du Chapitre 5 .....	135
<b>Conclusion générale et perspectives .....</b>	<b>137</b>
<i>Conclusion</i> .....	137
<i>Perspectives</i> .....	138
<b>Références .....</b>	<b>143</b>
<b>ANNEXES.....</b>	<b>161</b>
A. Fabrication des puces microfluidiques.....	161
A.1. Photolithographie optique .....	161
A.2. Gravure du silicium.....	161
A.3. Perçage des puces .....	161
A.4. Oxydation des puces .....	162
A.5. Encapsulation des puces .....	162
B. Calcul des résistances hydraulique et électrique du dispositif capillaire.....	163
B.1. Vitesse d'écoulement dans le dispositif capillaire.....	163
B.2. Champ électrique dans le dispositif capillaire.....	163
C. Protocole d'extraction en phase liquide .....	165

# Lexique

---

Adaptateur : Courte séquence nucléotidique capable de réaliser le pontage entre deux fragments d'ADN terminés par des séquences non complémentaires.

Allèle : Séquence alternative au locus d'un gène. Chez les organismes diploïdes, chaque allèle d'un même gène provient du père ou de la mère.

Amorce : Oligonucléotide dont la séquence est complémentaire d'une matrice et utilisé pour initier l'action d'une polymérase.

Amplification : Production de copies supplémentaires d'une séquence d'ADN.

Assemblage : Etape qui consiste à aligner et relier les reads obtenus par les séquenceurs en contig.

BAC pour chromosome bactérien artificiel : Vecteur bactérien conçu pour la cartographie et l'analyse des génomes complexes et utilisé pour cloner des fragments d'ADN de 100 à 300 kb dans une bactérie.

Carte physique : Permet de localiser, à l'aide de mesures physiques (exprimées en nombre de bases) les distances séparant des repères chromosomiques : sites de restriction, gènes, marqueurs.

Coefficient de variation (CV) : Rapport entre l'écart-type et la moyenne exprimé en pourcentage. C'est une mesure relative de la dispersion des données autour de la moyenne.

Contig : Ensemble de reads chevauchants.

Couverture/profondeur : Longueur totale des fragments séquencés divisé par la taille du génome séquencé. Exprimée en X, soit le nombre de fois qu'un génome est couvert.

Délétion : Suppression d'une base/séquence d'ADN dans un génome.

Dénaturation d'acide nucléique : Passage de l'état double brin à l'état simple brin pour un acide nucléique.

Duplication : Dédoublement d'un chromosome ou d'un gène.

Fidélité (d'une mesure) : La fidélité est la faculté d'un instrument à donner des mesures répétables. Généralement exprimée avec le coefficient de variation (CV).

Gène chimère : Gène formé de fragments d'ADN d'origines diverses.

Haplotype : Groupe d'allèles adjacentes sur le génome (provenant généralement d'un seul parent).

Insertion : Addition d'une Base/séquence d'ADN dans un génome.

Justesse (d'une mesure) : La justesse est la faculté d'un instrument à donner des mesures dont la moyenne est proche de la valeur vraie.

Ladder ADN : Marqueur de poids moléculaire d'ADN. Contient plusieurs fragments d'ADN de taille calibrée servant de référence pour analyser un échantillon inconnu.

Librairie de séquençage : Ensemble des fragments qui vont être séquencés.

N50 : Définit la qualité d'un assemblage en termes de continuité. Nombre de contigs nécessaires pour couvrir 50% du génome.

PCR (polymerase chain reaction) : Réaction enzymatique in vitro consistant à répliquer la chaîne des petits segments d'ADN. Amplification exponentielle.

Plug d'agarose : Petit bloc d'agarose permettant d'encapsuler des cellules ou de l'ADN.

Polymorphisme : Coexistence dans une population de plusieurs allèles pour un locus donné.

Read : Séquence brute déterminée par un séquenceur pour un fragment d'ADN.

Résolution : Définit la qualité d'une séparation entre deux pics. Elle est donnée par :

$R = 2\left(\frac{t_1 - t_2}{w_1 + w_2}\right)$ . Avec t le temps de migration et w la largeur à la base du pic.

Scaffold : Ensemble de contigs chevauchants.

Site abasique : Emplacement de l'ADN où la base (purine ou pyrimidine) est manquante.

SNPs (Single Nucleotide Polymorphisms) : Polymorphisme portant sur un seul nucléotide.

Variant : Désigne n'importe quel changement intervenu dans la séquence ADN

Variation structurale : Variant portant sur une grande région génomique (délétion, duplication, translocation...).

# Introduction

---

## Contexte

Les techniques de séquençage ont connu un extraordinaire développement depuis plus de 40 ans et ont permis une véritable révolution dans le domaine de l'analyse biologique avec l'entrée dans l'ère de la génomique. La première méthode de séquençage à être adoptée fut celle développée par Sanger dans les années 1970 qui permit d'obtenir la première séquence d'un organisme, celle du bactériophage  $\phi$ X174 en 1977<sup>1</sup>. Le développement de cette méthode a ensuite connu un essor considérable, avec comme point d'orgue l'achèvement du séquençage du premier génome humain en 2001. Le « Projet Génome Humain » s'étalant sur 15 ans et pour un budget de 3 milliards de dollars a permis d'innover en matière de séquençage mais également au niveau des techniques analytiques et bio-informatiques associées. On peut notamment citer la standardisation de l'électrophorèse capillaire pour l'analyse ADN à haut débit<sup>2</sup> ainsi que des méthodes d'extraction d'ADN. Le séquençage a ensuite évolué vers les techniques dites de « next generation sequencing (NGS) » permettant la lecture en parallèle de millions de petits fragments (reads) de quelques centaines de paires de base (bp). Cela a permis de diminuer drastiquement le temps et le prix du séquençage, de quelques milliers de dollars à moins d'un dollar par million de paires de base (Mbp)<sup>3</sup>.

Malgré la puissance de ces outils NGS et leur adoption massive dans les laboratoires, certains défis justifient encore le développement de nouvelles technologies de séquençage dites de 3<sup>ème</sup> génération. En effet, le séquençage de grands génomes de plusieurs milliards de paires de base (Gbp) reste très laborieux avec les techniques NGS du fait de la grande quantité de séquences à assembler d'un point de vue bio-informatique. Les génomes complexes, contenant de nombreuses régions répétées pouvant faire plusieurs milliers de paires de base (kb), et les grandes variations structurales peuvent également être difficiles à analyser en NGS. L'intérêt des récents séquenceurs de 3<sup>ème</sup> génération réside dans leur capacité à lire des longues molécules individuelles de plusieurs kb et ainsi de considérablement faciliter le travail bio-informatique<sup>4</sup>. Tel un puzzle, plus les fragments d'ADN décodés sont grands et plus l'assemblage du génome devient aisé.

Ces techniques de séquençage impliquent cependant de nouvelles contraintes sur la préparation et le contrôle qualité de l'échantillon. En effet, l'obtention de grandes molécules d'ADN lors des étapes d'extraction est loin d'être garantie selon les organismes étudiés. La génomique de 3<sup>ème</sup> génération a ainsi renouvelé l'intérêt de la communauté pour les techniques de manipulation d'ADN de haut poids moléculaire. Ces techniques ont été développées essentiellement dans les années 80, avec notamment l'apparition de l'électrophorèse en champs pulsés, mais ont ensuite subi un manque de développement<sup>5</sup>. Il y a donc aujourd'hui un besoin technologique important à combler lié aux

grands fragments, d'un point de vue préparatif et analytique, pour exploiter au mieux les capacités de ces séquenceurs. C'est dans cette optique que s'inscrivent ces travaux de thèse dont l'objectif est le développement de nouvelles méthodes d'analyse et de préparation de longs fragments d'ADN (10 – 200 kb) pour la génomique de 3<sup>ème</sup> génération.

Ces travaux de thèse sont basés sur la technologie  $\mu$ LAS (micro laboratoire pour l'analyse et la séparation de l'ADN) permettant la manipulation d'ADN dans un fluide viscoélastique sous l'application bidirectionnelle d'un flux laminaire et d'une électrophorèse<sup>6</sup>. Cette technique, inventée au LAAS-CNRS, a fait l'objet de plusieurs brevets concernant la séparation et la concentration d'ADN qui ont ensuite été licenciés à la société Picometrics Technologies pour un développement industriel<sup>7,8</sup>. D'un point de vue académique la technologie est utilisée sous un format de puces microfluidiques et Rémi Malbec, durant sa thèse au LAAS, a montré l'intérêt de  $\mu$ LAS pour l'analyse de l'ADN résiduel et la détection de séquences<sup>9</sup>. En parallèle, Picometrics a développé industriellement une méthode d'analyse ultrasensible (10 fg/ $\mu$ L) de l'ADN de bas poids moléculaire (100 – 1500 bp) dans l'optique d'analyser l'ADN circulant en tant que biomarqueur du cancer<sup>10</sup>. L'objectif de cette thèse CIFRE, menée entre le LAAS et Picometrics, est d'étendre le champ d'application de cette technique  $\mu$ LAS.

## ***Organisation du manuscrit***

Dans un premier chapitre, nous nous intéresserons aux techniques de génomique de 3<sup>ème</sup> génération en discutant leurs avantages et inconvénients par rapports aux générations antérieures. Nous nous focaliserons sur les caractéristiques des échantillons d'ADN utilisés par ces instruments.

Dans un deuxième chapitre, nous décrirons comment obtenir et caractériser l'ADN pour le séquençage. Nous aborderons plus particulièrement les instruments analytiques permettant d'effectuer un contrôle qualité de ces échantillons.

Dans un troisième chapitre, nous détaillerons les principes de fonctionnement de la technique  $\mu$ LAS sous les formats puce et capillaire. Nous verrons les différents développements permettant de travailler avec de l'ADN de haut poids moléculaire et discuterons des performances et de la pertinence de ce nouvel outil pour la génomique de 3<sup>ème</sup> génération.

Dans un quatrième chapitre, nous utiliserons  $\mu$ LAS pour capturer spécifiquement des grands fragments d'ADN de 30 kb et les caractériser avec des séquenceurs 3<sup>ème</sup> génération.

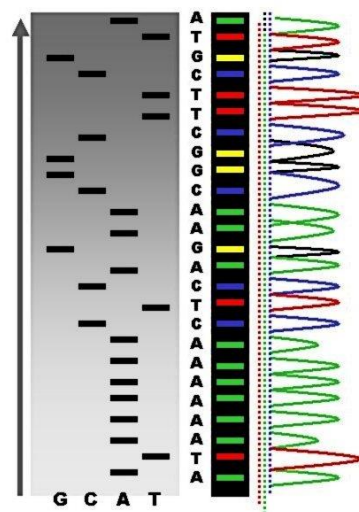
Enfin dans un cinquième chapitre, nous explorerons l'intérêt de la technologie  $\mu$ LAS pour la préparation d'échantillons d'ADN de haut poids moléculaire pour les techniques de séquençage 3<sup>ème</sup> génération avec le développement d'une vanne de sélection d'ADN accordable adaptée à l'enrichissement sélectif d'ADN selon sa taille.

# 1. Technologies pour la génomique de 3<sup>ème</sup> génération

## 1.1. Les 1<sup>ère</sup> et 2<sup>ème</sup> générations de séquençage

### 1.1.1 Le séquençage de Sanger

La méthode de Sanger, développée dans les années 70, fut la première à être massivement adoptée par les laboratoires de biologie moléculaire. Elle consiste à copier un fragment d'ADN simple brin en utilisant des nucléotides normaux mélangés avec des nucléotides marqués en fluorescence et modifiés afin de stopper la réaction d'élongation. L'échantillon à séquencer est réparti en quatre fractions, contenant des nucléotides normaux, une polymérase et seulement un nucléotide modifié en faible quantité. De ce fait, chaque réaction aboutit à des fragments de différentes tailles dont le dernier nucléotide est marqué, permettant ainsi de « lire » la nature chimique d'une base par sa couleur et sa position en mesurant la taille du fragment par électrophorèse (Figure 1).



**Figure 1 : Séquençage de Sanger.** Après une étape d'élongation des fragments à séquencer avec des nucléotides marqués, l'échantillon est séparé par électrophorèse afin d'identifier la position de chaque nucléotide.

Cette méthode possède des atouts indéniables comme un taux d'erreur très faible de 0,001% et la possibilité d'analyser des fragments d'ADN jusqu'à 1000 paires de bases (bp) environ. Elle reste d'ailleurs largement utilisée dans les laboratoires car elle sert de référence pour valider les variants identifiés en NGS<sup>11</sup> et pour le séquençage ponctuel de plasmides du fait de la rapidité d'un cycle de lecture 1 (Tableau 1). Cependant, les étapes de PCR et d'électrophorèse limitent le débit d'analyse et cette méthode est très chronophage pour de grands génomes. En effet, les instruments les plus

performants permettent la séparation en parallèle de 96 échantillons en 20 min soit un débit de 1,4 Mb/jour, 1000 fois moins performants que les systèmes NGS actuels (Tableau 1). Après l'obtention du premier génome humain en 2001, il est devenu évident que de nouvelles technologies, plus rapides et moins chères étaient nécessaires à l'analyse génomique. En 2004, le National Human Genome Research Institute a ainsi lancé le financement d'un programme visant à séquencer un génome humain pour moins de 1000 \$ dans les 10 ans, ce qui a conduit au développement des méthodes NGS<sup>12</sup>.

### 1.1.2 Le séquençage NGS

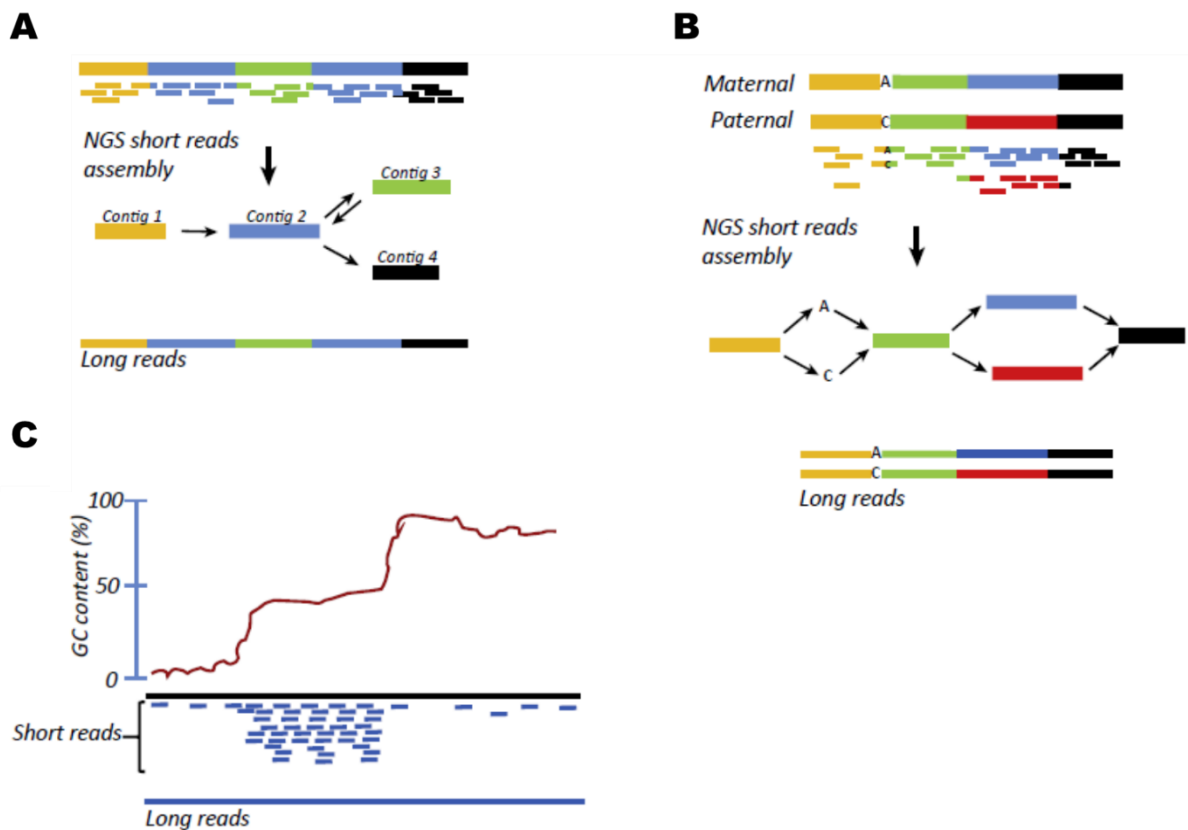
L'apparition des techniques NGS en 2007 a permis de diminuer drastiquement les coûts et le temps de séquençage grâce au multiplexage et à la miniaturisation des réactions (Tableau 1). Le séquençage est alors passé de centaines de réactions en parallèle à plusieurs millions. De plus, les méthodes NGS permettent la détection de la base durant l'étape de copie du fragment ce qui supprime l'étape de séparation en taille, limitante en termes de débit<sup>13</sup>. Bien qu'elles produisent plus d'erreurs que les techniques de 1<sup>ère</sup> génération (Tableau 1), la grande quantité de données générée permet une forte profondeur de lecture et ainsi de gommer ces erreurs. Ces technologies NGS sont ainsi fortement implantées dans toutes les plateformes de séquençage haut débit.

De manière générale, des fragments d'ADN de 100 à 400 bp sont immobilisés sur une puce où ils vont être amplifiés. La détection des bases est faite lors de l'incorporation de nucléotides par chimioluminescence, fluorescence et changement du pH pour respectivement, Roche (454), Illumina et IonTorrent<sup>14</sup>. Aujourd'hui le marché du séquençage 2<sup>nd</sup> génération est fortement dominé par Illumina avec 75% d'un marché estimé à 3,5 milliards de dollars en 2016<sup>15</sup>. Cette technologie permet à l'heure actuelle d'avoir le plus fort débit par cycle pour le plus faible prix à la Gb parmi les techniques NGS grâce à son dernier instrument, le NovaSeq 6000, permettant de produire jusqu'à 3 Tb en moins de 2 jours<sup>16</sup>.

### 1.1.3 Pourquoi la nécessité d'une 3<sup>ème</sup> génération ?

Les techniques NGS sont particulièrement performantes pour produire de grandes quantités de données à faible coût et avec une bonne justesse, cependant, ces dernières conservent certaines faiblesses qui ont poussé au développement de nouveaux séquenceurs (Tableau 1). Parmi ces limitations on peut citer notamment le besoin d'une étape d'amplification de l'ADN qui peut induire des erreurs lors de la réplication, des biais dépendant de la séquence à amplifier<sup>17</sup>, ie séquences riches en GC (Figure 2, C), ou encore la perte de l'information épigénétique. La seconde limitation des techniques NGS est la longueur des séquences pouvant être décodées, typiquement de quelques centaines de bases. Or l'assemblage de tous ces petits fragments d'ADN en un seul et même grand

fragment (contig) peut s'avérer particulièrement complexe et demande des ressources bio-informatique colossales<sup>18</sup>. De plus, la plupart des génomes contient des régions répétées plus longues que les reads NGS. Cela ne permet pas d'assembler correctement les séquences et conduit à des trous dans la séquence finale qui est donc divisée en de nombreux contigs (Figure 2, A). Les grandes variations structurales, impliquées dans de nombreuses maladies<sup>19</sup> sont également très complexes à résoudre avec les reads courts des séquenceurs de 2<sup>nd</sup> génération (Figure 2, B).



**Figure 2 : Avantages des techniques de séquençage de 3<sup>ème</sup> génération. A : Une plus grande longueur des reads facilite l'assemblage des régions répétées par rapport aux techniques NGS. B : Les reads longs facilitent le phasage des haplotypes. Les séquenceurs NGS ne permettent pas de couvrir en un seul read des variations structurales entre les allèles paternelles et maternelles ce qui conduit à des trous dans les contigs. C : Les méthodes de 3<sup>ème</sup> génération ne reposant pas sur la PCR ne sont pas sensibles aux biais d'amplifications, notamment dans les régions riches ou pauvres en GC au contraire des techniques NGS. Tiré de Vandijk et al.<sup>20</sup>.**

Au milieu des années 2000 a donc débuté le développement des séquenceurs de 3<sup>ème</sup> génération ayant pour caractéristique principale de ne pas avoir recours à l'amplification PCR et d'obtenir des reads beaucoup plus longs, entre 10 et plusieurs centaines de kb<sup>21</sup>. Ces techniques de 3<sup>ème</sup> génération permettent d'analyser les génomes avec une résolution inégalée grâce à leur capacité à lire des molécules d'une très grande taille.



Dans ce manuscrit nous nous attacherons à décrire le fonctionnement de quatre technologies de génomique de 3<sup>ème</sup> génération qui sont les plus implantées actuellement dans le marché : la technologie « Single Molecule Real Time (SMRT) sequencing » développée par PacificBiosciences (PacBio), le séquençage nanopore dominé par Oxford Nanopore Technologies (ONT), la technologie Chromium de 10X Genomics qui utilise des séquenceurs de 2<sup>nd</sup> génération mais permet d’avoir une information sur de longues molécules et enfin les techniques de cartographie optique. A partir de revues de références<sup>20,22,23</sup>, nous détaillerons le fonctionnement de ces différentes plateformes en nous focalisant en particulier sur les caractéristiques de l’échantillon ADN nécessaire à leur bon fonctionnement. Nous nous attacherons uniquement à décrire la préparation des bibliothèques de longs fragments supérieurs à 10 kb. L’objectif étant d’identifier des besoins en matière d’analyse et de préparation d’ADN qui seraient mal ou partiellement pourvus par les outils du marché.

**Tableau 1 : Comparaison des principales caractéristiques des 3 générations de séquenceurs avec leurs avantages et inconvénients**

Génération	Plateforme	Instrument	Longueur de read	Débit par cycle	Temps par cycle	Taux d'erreur	Prix Instrument (US\$)	Coût par Gb (\$)	Avantages	Inconvénients
1	Applied biosystem	3730 Series Genetic Analyzer	900 bp	1,38 Mb/jour	20 min	0.001%	268000 (en 2005)	1000000	Justesse, prix et rapidité par cycle	Débit, prix par Gb
2	Illumina	MiSeq v3	75 ou 300 bp	3,8 ou 15 Gb	21 ou 56 h	0.10%	99000	250 ou 110	Haut débit, Prix/Gb, Technique la plus utilisée	Chargement de l'échantillon, reads courts, durée cycle
2		HiSeq 2500	125 bp	900 - 1000 Gb	6 jours	0.10%	690000	30		
2	Ion torrent	Ion proton	200 bp	10 Gb	2-4 h	1%	224000	80	Rapidité cycle, pas de détection optique	Taux d'erreur
3	PacBio	Sequel	20 kb	10 Gb	4 h	13% et <1% en CCS	350000	85	Taux d'erreur en CCS, reads longs, sans PCR, épigénétique	Prix, Débit, Quantité d'ADN de départ
3	ONT	MinIon	Variable, jusqu'à 1 Mb	40 Gb	48 h	12%	1000	24	Débit, portabilité, très longs reads, prix, sans PCR, épigénétique	Taux d'erreur + systématiques, Quantité d'ADN de départ
3		PromethIon		Jusqu'à 15 Tb	64 h	12%	160000	5		
2 et 3	10X genomics	Chromium + Illumina	~100 kb	Voir Illumina			Voir Illumina + 125000 \$	Voir Illumina + 500 \$ par échantillon	Avantages Illumina, Peu d'ADN de départ	Pas de vrais reads longs, PCR obligatoire

## 1.2 Séquençage SMRT – PacBio

### 1.2.1 Principe de fonctionnement

PacBio fut la 1<sup>ère</sup> entreprise à commercialiser un séquenceur de 3<sup>ème</sup> génération, en 2011, basé sur la technologie SMRT. Le séquençage s'effectue à partir d'ADN double brin auquel est attaché aux extrémités deux adaptateurs en épingle à cheveux afin de former un fragment circulaire appelé SMRTbell (Figure 3, B). Une polymérase et une amorce sont également attachés à chaque SMRTbell lors de la préparation de la librairie. Celle-ci est ensuite chargée sur une puce contenant 150000 puits d'environ 70 nm de diamètre, appelés « Zero Mode Waveguide » (ZMW)<sup>24</sup>, pour le système RSII et 1 million pour le système plus récent Sequel (Figure 3, C). A l'intérieur de chaque puits une polymérase unique se retrouve immobilisée et réplique le fragment d'ADN à séquencer. Les nucléotide A, T, G, C, sont marqués d'un fluorophore différent de sorte que l'incorporation d'un nucléotide donne un signal fluorescent caractéristique de chaque base (Figure 3, D). Les puits sont imagés en temps réel ce qui permet d'identifier la base incorporée sur chaque molécule individuelle d'ADN mais également de discriminer les bases méthylées ou non en fonction de leur temps d'incorporation<sup>25</sup> (Figure 3, E).

Le taux d'erreur lors de l'identification des bases est assez important, de l'ordre de 13% (Tableau 1); ces erreurs sont aléatoires et sont donc corrigibles avec une couverture suffisante<sup>26</sup> (Figure 3, F, G, H, Long insert library). La justesse du séquençage peut également être améliorée via l'utilisation de librairies avec de plus courts fragments (~1,5 kb) dans les SMRTbell ; une fois le fragment d'intérêt séquencé une première fois (réplication complète d'un brin), la polymérase continue son chemin via les adaptateurs et séquence l'autre brin du fragment. De nombreux cycles peuvent ainsi être effectués pour former un « continuous long read » (CLR) et ainsi augmenter considérablement la justesse de séquençage, le taux d'erreur pouvant atteindre 0.001% à partir de 25 passages (Figure 3, F, G, H, Short insert library). Cependant, la durée de vie de la polymérase étant limitée, plus un fragment est long et moins il subira de cycles dans le CLR et plus faible sera la justesse du séquençage<sup>27</sup>. Le protocole utilisant des librairies avec des longs fragments produit des reads de 20 kb en moyenne avec un maximum de 60 kb<sup>28</sup>.

L'absence d'amplification PCR lors de la préparation de la librairie rend cette technologie particulièrement intéressante pour l'analyse de longues régions répétées et riches en GC<sup>29</sup>. La longueur des reads permet également de combler des trous dans des génomes pourtant déjà étudiés en profondeur comme le génome humain<sup>30</sup>.

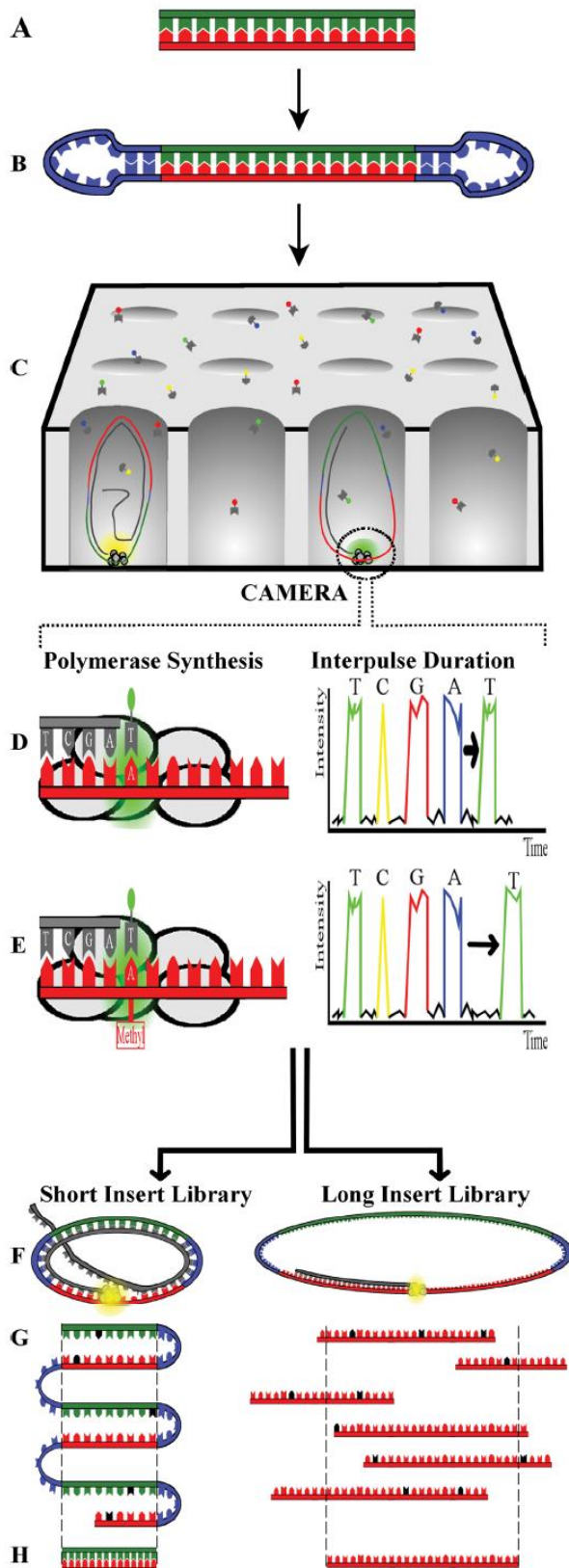


Figure 3 : Principe de fonctionnement du séquençage SMRT. A : La librairie est créée à partir d'ADN double brin. B : Adaptateurs, polymérase et amorces sont attachés sur l'ADN durant la création de la librairie afin de produire les SMRTbell. C : Une polymérase unique est introduite dans les ZMWs et incorpore les nucléotides marqués. D : Lors de l'incorporation des nucléotides, un fluorophore spécifique de chaque base est relâché et émet un signal fluorescent caractéristique. E : Chaque ZMW est observé en temps réel ce qui permet de connaître le temps d'incorporation de chaque nucléotide. On peut ainsi obtenir une information épigénétique, les bases méthylées étant plus longues à incorporer. F-H, panneau de gauche : La forme circulaire des SMRTBell permet à la polymérase de décoder plusieurs fois les deux brins d'un petit fragment d'ADN. Cela conduit à un read long continu qui permet d'avoir une très bonne justesse de séquençage. F-H, panneau de droite : Pour les longs fragments, le taux d'erreur est plus important (nucléotides noirs) à cause de la limite de processivité de la polymérase. Une forte couverture permet de gommer les erreurs et d'obtenir une séquence de qualité. Tiré de Ardui et al.<sup>31</sup>.

## 1.2.2 Limitations

Malgré les avantages associés aux reads longs, la technologie PacBio souffre de quelques désavantages notamment liés au rendement du séquençage. En effet, leur dernier instrument, le Sequel, reste bien moins productif que les techniques les plus performantes en termes de Gb/jour comme Illumina ou Oxford Nanopore Technologies (Tableau 1). Le faible rendement de cette technologie peut s'expliquer par le faible taux de micro puits fonctionnels au sein d'une cellule, i.e. ne contenant qu'une seule et unique polymérase. En effet, le chargement d'une polymérase par puits suit une distribution de Poisson et implique un maximum de 37% de puits fonctionnels<sup>32</sup> soit environ 365000 puits pour le Sequel. Cela impacte également le prix par Gb, plus élevé chez PacBio, avec de surcroît un coût lié à l'instrument non négligeable (Tableau 1).

La technologie PacBio est en outre exigeante en termes de quantité d'ADN requise pour un séquençage. Il est en effet recommandé d'utiliser 5 µg d'ADN pour le séquençage d'un génome humain. Or une couverture de 60X, soit celle nécessaire pour identifier insertions et délétions avec une bonne précision<sup>33</sup>, correspond à la lecture de 400 pg d'ADN environ soit 10000 fois moins que la quantité requise. Une partie de l'ADN initial est en fait consommé lors de la préparation de la librairie (voir 1.2.3). Une autre raison de cette exigence en quantité d'ADN tient à la diffusion de l'ADN dans les ZMWs quand on charge l'échantillon dans la puce. En effet, le coefficient de diffusion de l'ADN est décrit selon :

$$D = \frac{kbT}{6\pi\eta Rh} \quad (1)$$

Pour une molécule d'ADN de 48 kb (phage lambda) on trouve un coefficient de diffusion D de 0,47 µm<sup>2</sup>/s<sup>34</sup>. On peut, à partir de ce coefficient de diffusion, évaluer le temps nécessaire à une molécule d'ADN pour explorer par pure diffusion une distance d selon la relation :

$$t = \frac{1.37d^2}{\pi^2 D} \quad (2)$$

La puce de l'instrument Sequel fait quelques millimètres de hauteur. On peut ainsi évaluer à 300 h environ le temps nécessaire à un fragment de 48 kb pour diffuser à travers la puce jusqu'au ZMW où il sera séquencé, d'où la nécessité de charger une forte quantité d'ADN sur la puce. De plus, le chargement des ZMW par diffusion induit le séquençage préférentiel des petits fragments par rapport aux grands, ceux-ci diffusant plus rapidement.

Un autre élément limitant le séquençage des grands fragments tient au confinement des molécules à l'intérieur des ZMW. En effet ils ne dépassent pas quelques dizaines de nm alors que le rayon de gyration d'une molécule de 50 kb est d'environ 700 nm<sup>34</sup>. Afin d'entrer dans le ZMW, l'ADN doit se déformer, ce qui crée une barrière entropique<sup>35</sup> d'autant plus élevée que la molécule est longue. PacBio a ainsi développé une méthode de chargement plus homogène pour les grands fragments à l'aide de billes paramagnétiques, les « MagBead ». Cela consiste à ancrer les SMRTbell sur des billes, plus grandes que les ZMWs, qui vont ensuite être chargées sur la puce. Ainsi, les courts fragments ou les dimères d'adaptateurs, trop petits pour atteindre le fond des puits, restent accrochés aux billes alors que les molécules > 1 kb peuvent s'y ancrer. PacBio développe également une technologie utilisant un champ électrique pour le chargement de l'ADN dans la puce permettant une mobilité similaire entre tous les fragments et ainsi d'éviter les biais liés à la taille des molécules. Cette technique a permis d'augmenter le chargement d'un facteur 100000 et contribue au développement de nouveaux protocoles utilisant de faibles quantités de matière<sup>36</sup>. Malgré ces améliorations, une sélection en taille de l'échantillon est recommandée afin de retirer les petits fragments et ainsi améliorer la qualité du séquençage<sup>31</sup>.

### 1.2.3 Préparation et caractéristiques de l'échantillon à séquencer

PacBio ne préconise pas explicitement les protocoles d'extraction et de préparation de l'échantillon à séquencer, seulement les caractéristiques finales de celui-ci. Il doit ainsi être composé d'ADN de grande taille préférentiellement supérieur à 10 kb et de bonne qualité : ne contenant pas de coupures simple brin, de sites abasiques ou autres modifications ni d'intercalant. La pureté doit également être évaluée afin d'éviter les contaminations par de l'ARN ou de l'ADN simple brin, des sels ou des protéines pouvant interférer dans la quantification. Ces procédures de contrôle qualité de l'échantillon seront décrites plus tard dans ce manuscrit (voir 2.3).

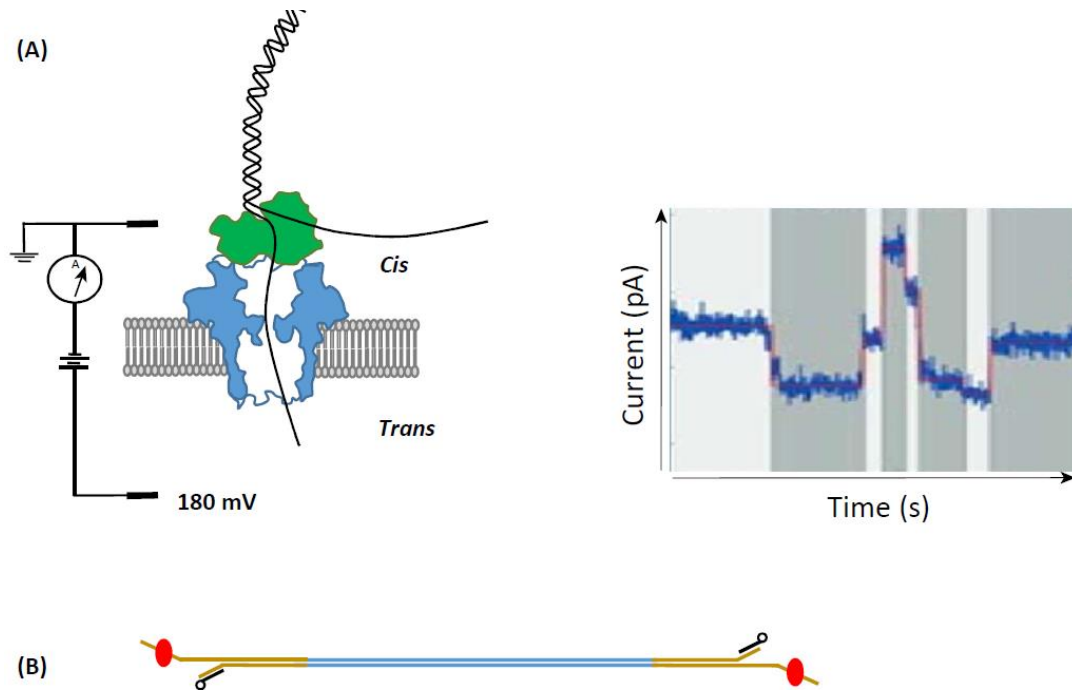
La préparation de la librairie dure environ 3 jours et utilise 5 µg d'ADN. Afin d'avoir des fragments d'ADN d'une taille homogène, l'échantillon peut être fragmenté puis sélectionné en taille. Après traitement avec diverses enzymes pour réparer l'ADN, lier les adaptateurs, digérer l'ADN simple brin, puis la purification, il reste environ 10% de l'échantillon de départ. Il est caractérisé en taille et en concentration avant l'ajout des amorces et de la polymérase pour finalement déposer environ 5 ng soit 0,1% de l'ADN de départ pour une cellule RSII. Chaque nouvelle étape ajoutée (fragmentation, sélection en taille, dosage) contribue au faible rendement de la préparation de la librairie.

## 1.3 Séquençage Nanopore – ONT

### 1.3.1 Principe de fonctionnement

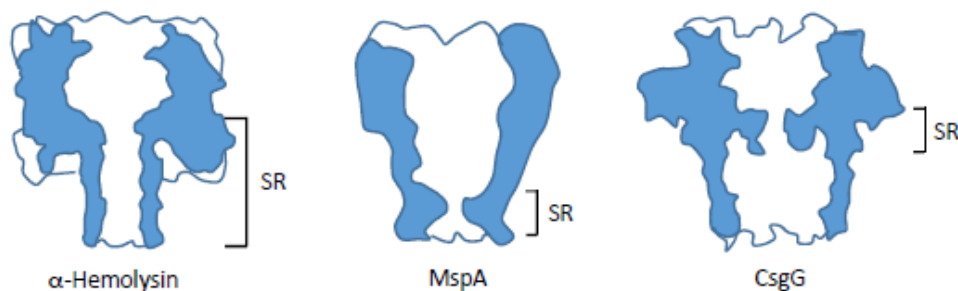
L'idée de séquencer de l'ADN à travers un nanopore est assez ancienne, initiée dans les laboratoires au début des années 90<sup>37</sup>. En 2005 a été fondée la société Oxford Nanopore Technologies (ONT) qui fut la 1<sup>ère</sup> à commercialiser un instrument pour le séquençage en nanopore, le Minlon en 2014. Cette technologie repose sur l'utilisation d'un pore de quelques nanomètres de diamètre, inséré dans une membrane en polymère séparant deux solutions ioniques. Sous l'application d'un champ électrique et à l'aide d'une protéine motrice, l'ADN passe à travers le nanopore (Figure 4, A, panneau de gauche). L'ADN obstrue partiellement le pore lors de son passage et induit une modification du courant de quelques dizaines de pico ampère<sup>38</sup>. Chaque nucléotide qui passe modifie le courant de manière caractéristique, ce qui permet l'identification des 4 bases de l'ADN ainsi que de l'ARN et des méthylations en temps réel (Figure 4, A, panneau de droite). Cette méthode ne nécessite donc aucun fluorophore ou système optique complexe, réduisant fortement les coûts de séquençage.

Afin d'être séquencé l'échantillon doit être attaché à des adaptateurs contenant la protéine motrice (polymérase ou hélicase) qui facilite la capture et le passage du brin d'ADN à travers le nanopore (Figure 4, B). L'ADN double brin est ensuite déroulé de sorte qu'un seul brin passe à travers le nanopore.



**Figure 4 : Représentation schématique de la technologie ONT. A : Deux solutions ioniques sont séparées par une membrane contenant un nanopore (bleu). Une protéine motrice (vert) facilite l'entrée d'un fragment simple brin d'ADN dans le nanopore par électrophorèse. Le passage de l'ADN à travers le pore induit une modification du courant caractéristique de chaque base. Le courant est enregistré en temps réel et permet d'identifier quelle base est transloquée à travers le nanopore. La modification du courant lors du passage à travers le pore est également caractéristique de l'information épigénétique. B : Des adaptateurs sont liés à chaque extrémité (jaune) ainsi qu'une protéine motrice (rouge). A l'autre extrémité du brin d'ADN, un nucléotide spécifique (noir) est hybridé afin de faciliter l'entrée du second brin d'ADN dans le pore (protocole 1D<sup>2</sup>). Tiré de Van Dijk et al.<sup>20</sup>.**

Cette technologie a été historiquement développée en utilisant la protéine  $\alpha$ -Hémolysine en tant qu'entité physique de nanopore. Cependant, avec cette protéine, l'identification des bases n'était pas suffisamment précise et de nouvelles protéines ont été testées avec des constriction plus courtes permettant d'augmenter la sensibilité de la détection (Figure 5). La protéine la plus performante est un mutant du pore CsgG d'*E.Coli*<sup>39</sup>.



**Figure 5 : Différents types de nanopores utilisés par ONT. L' $\alpha$ -Hémolysine est la protéine historiquement utilisée pour le séquençage nanopore. Les protéines MspA et CscG ont une meilleure justesse lors du séquençage grâce à une région de reconnaissance (SR) plus courte. La protéine CsgG est celle utilisée dans le dernier kit commercialisé par ONT. Tiré de Van Dijk et al.<sup>20</sup>.**



Environ 450 bases par secondes peuvent être analysées à travers cette dernière génération de nanopore. De plus, le multiplexage est aisé via l'augmentation de la taille ou du nombre des puces contenant les nanopores. Ainsi trois instruments sont actuellement disponibles selon les besoins en termes de portabilité et de débit : du Minlon, pas plus gros qu'une clé USB et contenant 512 nanopores au Promethlon pouvant contenir 144000 nanopores et pouvant produire jusqu'à 15 Tb par jour (Tableau 1). Cette modularité permet à ONT de rendre le séquençage accessible à tous (le Minlon est disponible à partir de 1000 \$), mobile et rapide. On peut notamment citer son utilisation pour le séquençage d'échantillons sur le terrain pour la reconnaissance de pathogènes, comme lors de l'épidémie Ebola en 2015 en Guinée<sup>40</sup> ou lors de séquençages effectués dans l'espace<sup>41</sup>. D'autre part, la commercialisation du Promethlon a également permis d'infiltrer le marché du séquençage haut-débit avec la possibilité de produire de grandes quantités de données pour un coût similaire aux technologies 2<sup>nd</sup> génération les plus performantes (Tableau 1). ONT envisage également la possibilité de séquencer des polypeptides selon le même principe de fonctionnement<sup>42,43</sup>.

Contrairement à la technologie SMRT de PacBio, il n'y a pas de limite intrinsèque à la longueur des reads. Ainsi une préparation méticuleuse de l'échantillon ADN permettra d'obtenir de très longs reads, jusqu'à 1 Mb<sup>44</sup>. Ces reads ultra-longs permettent de caractériser des régions inexplorées, particulièrement complexes, comme les séquences centromériques<sup>45</sup> et télomériques<sup>46</sup>.

### 1.3.2 Limitations

Comme la technologie SMRT développée par PacBio, le séquençage nanopore souffre de la comparaison avec le NGS sur le taux d'erreur, d'environ 15% pour le protocole 1D lors de la lecture d'un seul brin de la molécule à séquencer<sup>39</sup>. Mais contrairement à PacBio, les erreurs de séquences en nanopore ne sont pas aléatoires<sup>22</sup>, ainsi une grande couverture ne permettra pas d'améliorer la justesse. Il n'est pas non plus possible de séquencer le même fragment à de multiples reprises comme cela peut être fait à l'aide des SMRTbell. Afin d'améliorer la justesse, ONT a tout de même développé un protocole 2D utilisant un principe similaire. Un adaptateur en forme d'épingle à cheveux a été ajouté à une extrémité des fragments à séquencer de sorte qu'une fois le premier brin lu, le second passe à son tour à travers le pore pour être décodé et ainsi augmenter la justesse à 91% environ<sup>47</sup>. Cependant, cette méthode de ligation avec une épingle à cheveux est protégée par un brevet de PacBio, ce qui a empêché son utilisation<sup>48</sup>. Un nouveau protocole appelé 1D<sup>2</sup> a donc été développé et repose sur l'utilisation de séquences spécifiques attachées aux extrémités du fragment qui augmentent la probabilité que le second brin passe à travers le nanopore après la lecture du premier. D'après ONT, jusqu'à 60% des fragments ont ainsi leurs deux brins séquencés et cela résulte en une

justesse de 97%<sup>49</sup>. Cela diminue cependant le débit de séquençage puisque les deux brins doivent être lus.

Tout comme PacBio, la technologie requiert une grande quantité d'ADN de départ, de l'ordre de quelques µg. Le champ électrique est quasiment nul lorsque l'ADN est situé loin du pore, ainsi le passage des molécules dans son voisinage se fait également par diffusion. Comme discuté pour le SMRT, la diffusion de longues molécules jusqu'aux pores peut être longue, ainsi le chargement d'une grande quantité de matière augmente le débit de l'instrument.

Malgré la possibilité d'effectuer des séquençages rapidement avec des protocoles simplifiés, il faut noter que l'obtention d'un séquençage de qualité, avec des reads de plusieurs centaines de kb, n'est accessible que par une préparation de l'échantillon méticuleuse. Ainsi les instruments « périphériques » pour cette préparation doivent être pris en compte dans les coûts du séquençage.

### 1.3.3 Préparation et caractéristiques de l'échantillon à séquencer

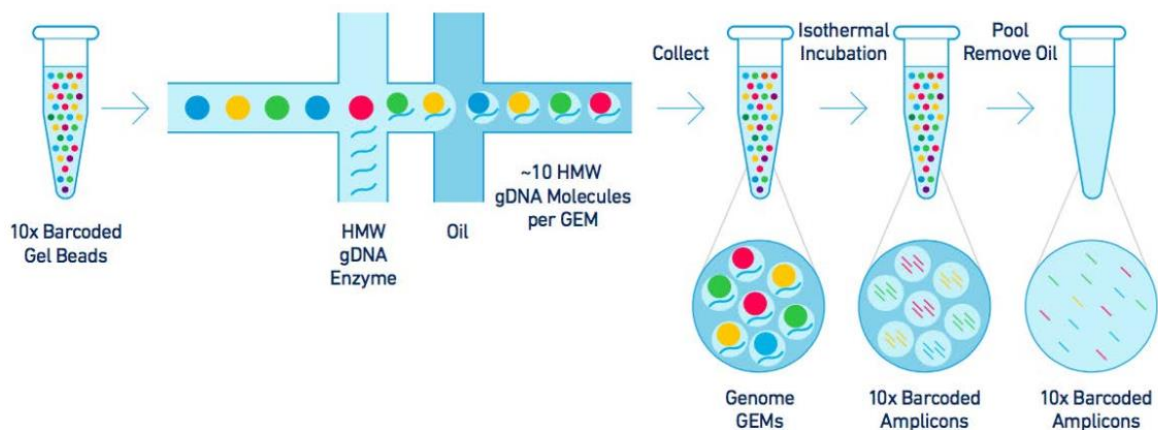
Tout comme PacBio, ONT ne propose pas de protocole spécifique pour l'extraction d'ADN. Leurs protocoles commencent par la préparation de la librairie à partir de l'ADN purifié. Deux protocoles sont disponibles : un premier très simplifié, faisable dans n'importe quel environnement, et un second plus complexe pour maximiser la qualité du séquençage. On distingue ainsi un « rapid sequencing kit » et un « ligation sequencing kit ». Le premier est très rapide et ne nécessite pas de grandes précautions en termes de qualité d'ADN. Le second utilise environ 1 µg d'ADN de départ et une sélection en taille de l'échantillon peut être effectuée par l'utilisateur si besoin, la longueur des reads correspondant à la longueur des molécules introduites dans la cellule. Le rendement du kit est d'environ 50%, ainsi ~500 ng d'ADN sont chargées sur une cellule du Minlon qui fournira jusqu'à 8 Gb en 48h. Cela donne un rendement de séquençage de  $1,8 \cdot 10^{-5}$ .

ONT a également développé un instrument pour effectuer la préparation de la librairie de manière automatisée, le VoltRAX. L'instrument, de la taille d'une clé USB, permet la lyse et la purification de l'échantillon avec des billes magnétiques (voir 2.1.2.2). Il permet de manipuler et déplacer des gouttes d'échantillon ou de réactif par électro mouillage<sup>50</sup>. Cet instrument est cependant dédié à des extractions rapides, loin des laboratoires, et n'est pas forcément adapté à la production de longs fragments.

## 1.4 Technologie Chromium – 10X Genomics

### 1.4.1 Principe de fonctionnement

La technologie Chromium, commercialisée par 10X Genomics en 2016, n'est pas une technique de séquençage à proprement parler mais permet d'obtenir de l'information sur de longues molécules d'ADN à partir d'un séquençage de 2<sup>nd</sup> génération grâce à une préparation astucieuse des bibliothèques. En effet cette technologie repose sur l'intégration de longues molécules d'ADN à l'intérieur de gouttelettes (GEMs) créées par une émulsion. Brièvement, des billes contenant chacune un code-barre génétique unique sont introduites dans un canal central d'une puce microfluidique. Les molécules d'ADN à séquencer et des enzymes sont ajoutées aux billes via un canal latéral. Un second canal latéral contenant de l'huile permet la production des gouttelettes de sorte qu'une seule bille est encapsulée par gouttelette (Figure 6)<sup>51</sup>. Chaque GEMs contient ainsi le matériel nécessaire à une amplification du fragment ADN mais également un code-barre spécifique à chaque gouttelette (Figure 6). L'ADN est fragmenté, amplifié et le code-barre est intégré à chaque fragment d'ADN à l'intérieur des gouttelettes de sorte que chaque fragment issu d'une même molécule possède le même code-barre. Les gouttelettes sont ensuite réunies afin d'effectuer une préparation de bibliothèque et un séquençage NGS classique (préférentiellement Illumina).



**Figure 6 : Principe de fonctionnement de la technologie Chromium de 10X Genomics. L'ADN de haut poids moléculaire est injecté avec des billes de gel contenant des codes-barres génétiques et le matériel nécessaire à une amplification. A l'aide d'une émulsion huile-eau des gouttelettes sont créées, chacune contenant une bille de gel et son code ainsi qu'une longue molécule d'ADN. Les gouttelettes sont ensuite rassemblées afin d'effectuer une amplification de l'ADN à l'intérieur des gouttelettes et d'intégrer le code-barre. L'ADN peut ensuite être séquençé classiquement avec des techniques NGS. Tiré du site internet de 10X Genomics.**

Ce système de code-barre facilite l'étape d'assemblage. En effet chaque read court issu d'une même gouttelette est associé au sein d'un grand « linked read » via son code-barre<sup>52</sup>. Jusqu'à 8 échantillons peuvent être chargés simultanément sur la puce permettant de générer la librairie. Plus de 4 millions de code-barres sont disponibles ce qui permet d'analyser de grandes quantités d'ADN contrairement aux précédents systèmes à code-barres en plaques micropuits comme le SLR d'Illumina<sup>53</sup>. On séquence ainsi des molécules d'une centaine de kb tout en gardant la grande justesse d'un séquençage 2<sup>nd</sup> génération qui fait actuellement fortement défaut aux deux autres technologies de séquençage 3<sup>ème</sup> génération. Cette justesse rend la technologie particulièrement adaptée à l'identification de différents haplotypes<sup>54,55</sup>.

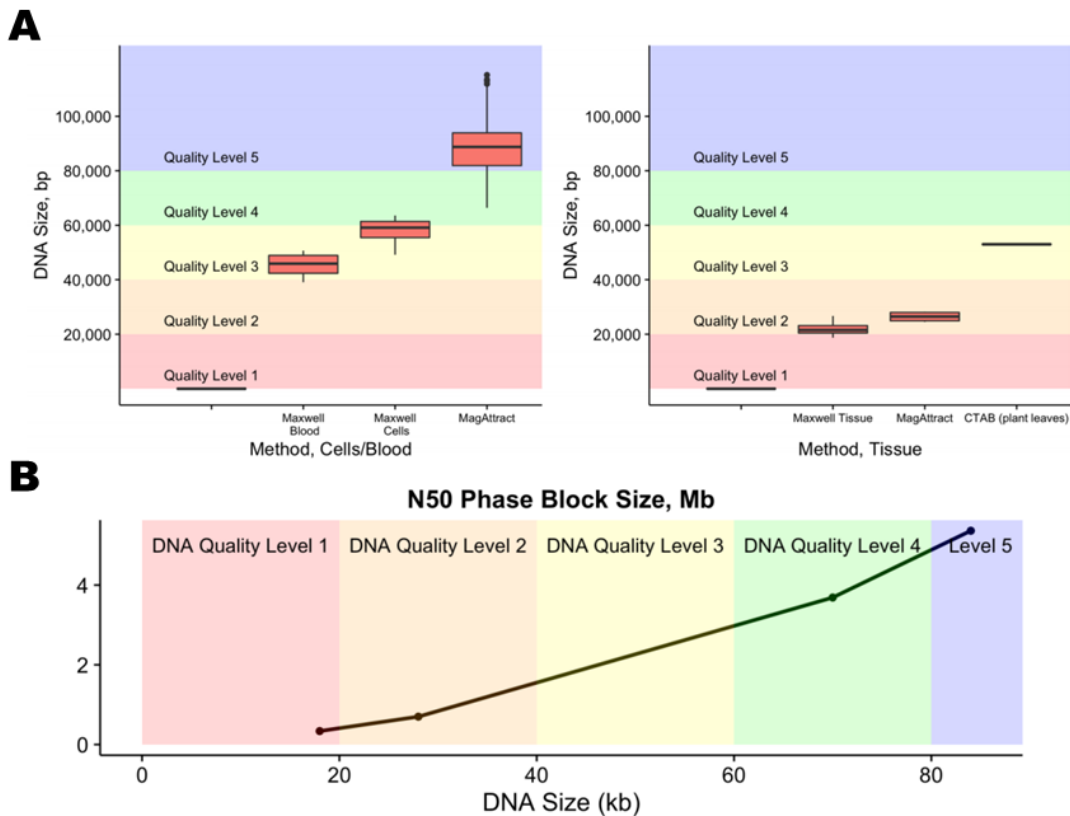
L'amplification PCR confère avantages et inconvénients. Elle permet d'un côté d'effectuer un séquençage avec de faibles quantité d'ADN, 1 ng, mais empêche d'obtenir directement l'information épigénétique et conserve les biais d'amplifications dans les régions riches en GC. Ainsi la technologie 10X Genomics ne semble pas être destinée à résoudre des régions complexes, riches en GC et répétées du génome.

Cette technologie est principalement utilisée pour des séquençage de novo<sup>56</sup>, du phasage d'haplotype<sup>57</sup> ou encore pour identifier des variants structuraux<sup>54</sup>. La possibilité de phaser des haplotypes est particulièrement intéressante pour l'étude des plantes car elles sont souvent polyploïdes et jusqu'à hexaploïdes. Elle est également utilisée pour du séquençage sur cellule unique via l'encapsulation des cellules directement à l'intérieur des GEMs<sup>58</sup>. La cellule est ensuite lysée et son ADN marqué avec le code-barre de la même manière.

#### 1.4.2 Préparation et caractéristiques de l'échantillon à séquencer

Un séquençage avec la technologie Chromium de 10X Genomics ne nécessite que 1 ng d'ADN, l'échantillon devant être utilisé à 1 ng/ $\mu$ L. En pratique, les procédures de contrôle qualité consomment également 15-20 ng d'ADN. Il est en effet essentiel que la quantité d'ADN par gouttelette soit maîtrisée. Une trop faible quantité engendrera une baisse de rendement de séquençage ainsi qu'une augmentation des duplications PCR<sup>59</sup> alors qu'un trop fort chargement augmentera les chances que deux fragments couvrant le même locus soient issus de deux haplotypes différents. Il est ainsi recommandé d'insérer entre 1 et 10 molécules par gouttelette selon la taille des fragments d'ADN (environ 500 kb/gouttelette), ce nombre étant maîtrisé par la concentration de l'échantillon de départ. 10X Genomics définit 5 niveaux de qualité d'ADN en fonction de sa taille, de 1 pour les molécules inférieures à 20 kb jusqu'à 5 pour l'ADN supérieur à 80 kb (Figure 7). Pour obtenir un séquençage optimal ils recommandent de l'ADN de niveau 3 à 5 selon les applications. Afin d'effectuer une extraction de qualité satisfaisante, 10X genomics recommande l'utilisation du kit MagAttract HMW

DNA de Qiagen (Voir 2.1.2.2) qui permet théoriquement d'éliminer les petits fragments de dégradation créés lors de l'extraction et de récupérer des fragments entre 100 et 200 kb (Figure 7). Le kit fonctionne sur des tissus, du sang, ou des bactéries et il est recommandé de faire l'extraction à partir de 5 µg d'équivalent ADN. Si l'ADN n'est pas d'une qualité suffisante, une sélection en taille de l'échantillon est recommandée à l'aide du BluePippin de SageSciences (2.2.2.2) ce qui améliore la qualité des résultats<sup>60</sup>.



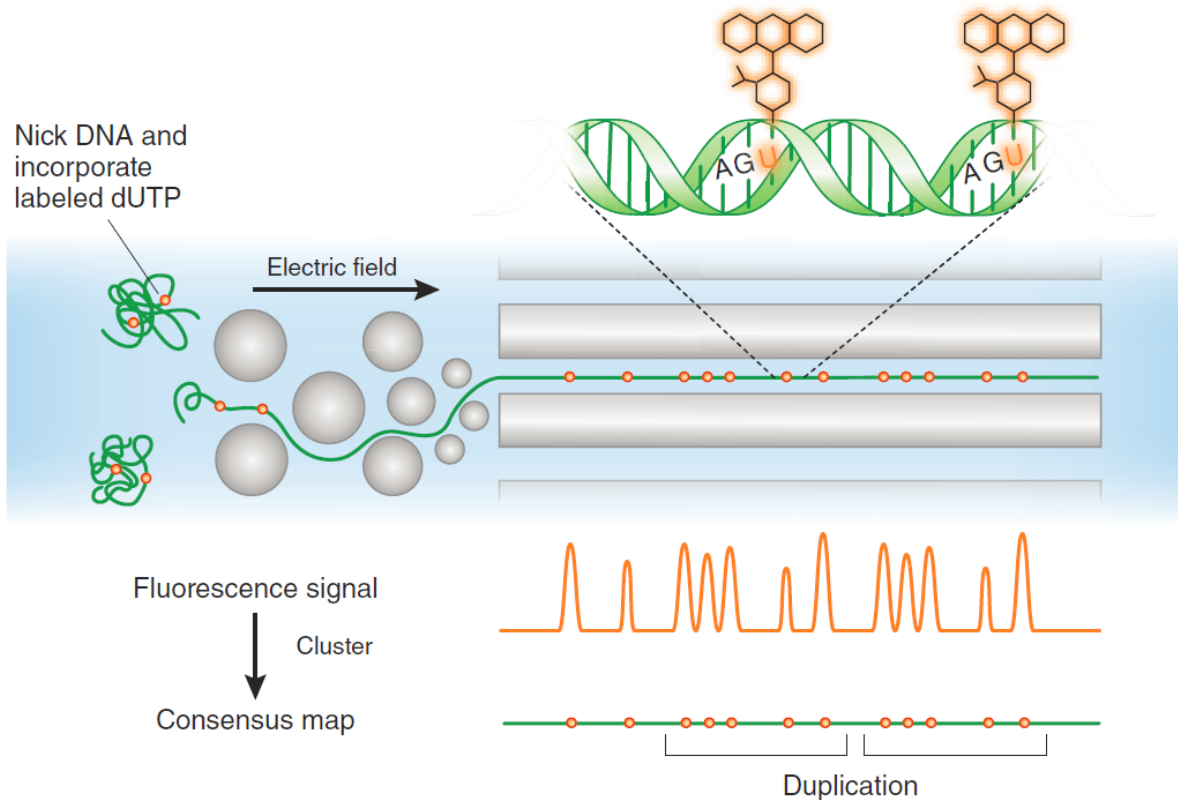
**Figure 7 : Qualité de l'ADN selon 10X Genomics. A : Comparaison de différentes méthodes d'extraction d'ADN à partir de cellules/sang (panneau de gauche) ou de tissu (panneau de droite) ainsi que la qualité d'ADN associée. B : Longueur du N50 obtenu en fonction de la qualité de l'ADN. Tiré d'une note d'application de 10X Genomics.**

La qualité du séquençage repose aussi sur les dommages causés à l'ADN lors des procédés d'extraction et de sélection en taille. En effet les molécules étant injectées en simple brin, toute coupure entrainera la diminution de la taille des « linked reads ». Il faut ainsi éviter l'utilisation de solvants organiques, le chauffage ou le pipetage excessif de l'échantillon qui peuvent conduire à une détérioration de l'ADN. La qualité de l'ADN reste cependant un facteur difficilement mesurable.

## 1.5 Cartographie optique

### 1.5.1 Principe de fonctionnement

La cartographie optique est une technique introduite en 1993 par Schwartz et al. et qui permet d'obtenir de l'information génétique longue portée, pour déterminer l'architecture du génome étudié<sup>61</sup>. La technique consiste à étirer de longues molécules d'ADN, jusqu'à plusieurs Mb, et d'effectuer un marquage fluorescent spécifique de certaines séquences. La molécule est ensuite observée en microscopie par fluorescence et la distance physique entre plusieurs balises fluorescentes le long de la molécule étirée est mesurée. Chaque molécule possède un profil caractéristique qui permet d'assembler toutes les molécules d'ADN en une grande carte physique consensus du génome via le chevauchement de ces profils. Quatre entreprises commercialisent des instruments de cartographie optique : Bionanogenomics, GenomicVision, Pathogenetix et OptGen. La plateforme de Bionanogenomics, probablement la plus utilisée actuellement, repose sur l'utilisation d'une puce contenant des micro et nanostructures qui vont progressivement dérouler l'ADN<sup>62</sup> et le guider dans des nanocanaux de 45 nm de diamètre à l'intérieur desquels une seule molécule d'ADN avance dans un état étiré<sup>63</sup> (Figure 8). L'ADN migre par électrophorèse à l'intérieur de la puce et les molécules peuvent être étirées jusqu'à 95% de leur longueur de contour<sup>64</sup>. Cette technologie utilise différentes nickases pour créer des coupures simple brin sur certaines séquences particulières. Des nucléotides marqués sont alors introduits au niveau des coupures ce qui permet de générer des balises ensuite repérées par microscopie<sup>65</sup> (Figure 8). OptGen utilise la méthode historiquement développée par Schwartz et al. utilisant un écoulement pour étirer l'ADN puis des enzymes de restrictions pour l'identification de séquence<sup>66</sup>. Cependant, cette méthode possède un débit d'analyse faible, l'étirement de l'ADN n'est pas uniforme ce qui engendre des imprécisions de mesures et un taux d'erreur important<sup>67</sup>. Pathogenetix utilise également des enzymes de restriction mais un dispositif microfluidique avec une constriction à travers laquelle l'ADN s'étire<sup>68</sup>. L'utilisation d'un flux laminaire pour l'étirement implique que les molécules sont en mouvement ce qui limite la résolution lors de l'imagerie<sup>67</sup>. Enfin GenomicVision utilise le peignage moléculaire pour étirer l'ADN sur une surface hydrophobe puis des sondes type « FISH » pour identifier des séquences ou des gènes d'intérêt<sup>69</sup>.



**Figure 8 : Principe de fonctionnement de la plateforme BioNanogenomics. L'ADN ayant subi des coupures simple brin et l'incorporation de fluorophores dans ces coupures est étiré à travers un réseau de micro et nano pylônes par électrophorèse. Il entre sous une forme étirée dans un nanocanal où il est observé en microscopie par fluorescence afin de mesurer la distance physique entre les balises fluorescentes. Chaque molécule d'ADN possède ainsi un profil caractéristique qui permet de créer une carte physique du génome. Tiré de Machaeli et al.<sup>70</sup>.**

L'information obtenue en cartographie optique n'est pas seulement génétique, on peut également identifier de l'information épigénétique grâce à un marquage des bases méthylées<sup>71</sup> ou encore protéique via des marqueurs spécifiques aux histones<sup>72</sup>. La capacité de la cartographie optique à analyser des très longues molécules la rend particulièrement utile pour la résolution de grandes variations structurales, identifier des contigs incorrectement assemblés, faciliter l'assemblage *de novo*<sup>73</sup> ou encore pour le typage de pathogènes.

Le nombre de nickases disponibles est une des limitations de la technologie. Bionano genomics travaille actuellement au développement de nouvelles nickases grâce à la technologie CRISPR/Cas9 pour la reconnaissance de séquence<sup>74,75</sup>.

## 1.5.2 Préparation et caractéristiques de l'échantillon

L'objectif étant d'obtenir de l'information sur plusieurs centaines de kb, il est indispensable d'avoir un ADN d'excellente qualité. Plus la molécule introduite est longue, plus les profils seront spécifiques à une portion du génome et plus l'assemblage sera univoque et fiable. Le système de Bionanogenomics élimine d'ailleurs informatiquement les molécules inférieures à 150 kb afin qu'elles ne soient pas analysées. De plus, lors de l'utilisation de nickases pour la reconnaissance de séquences, il est important de ne pas avoir de coupures simple brin dans l'ADN de départ, générées par exemple durant l'extraction. Ils seraient alors également marqués par fluorescence et induiraient du bruit dans l'information.

Afin de s'assurer de la bonne qualité des échantillons, BioNanogenomics fournit des protocoles spécifiques à l'extraction d'ADN de haut poids moléculaire. Ils utilisent des plugs d'agarose (voir 2.1.3) afin de préserver l'ADN du cisaillement. Le protocole recommande l'encapsulation d'environ 6 µg d'ADN dans l'agarose pour une lyse et une purification qui dure 3 jours. Cependant, la préparation d'une puce Bionanogenomics Saphyr ne requiert que 300 ng d'ADN purifié et supérieur à 150 kb. Il est recommandé d'utiliser un échantillon entre 40 et 200 ng/µL pour effectuer les étapes de coupures simple brin, d'intégration des fluorophores dans ces coupures, et du marquage du reste de la molécule. Ensuite, environ 100 ng d'ADN sont chargés sur une puce qui produit en moyenne 640 Gb soit 200 fois le génome humain sachant que l'objectif est généralement d'obtenir une couverture de 100X.

Une des limitations à l'adoption massive de ces technologies est la complexité de la préparation d'échantillon. En effet les protocoles d'extraction d'ADN en plug d'agarose sont laborieux et nécessitent un manipulateur expérimenté. Le développement de nouvelles méthodes d'extraction permettant une qualité d'ADN équivalente mais avec des protocoles simplifiés et automatisés est attendu.



## 1.6 Conclusion du chapitre 1

Dans ce premier chapitre, nous avons vu que le développement des technologies utilisant de l'ADN de haut poids moléculaire est foisonnant. Les instruments génèrent de plus en plus de données et permettent d'obtenir de l'information sur de longues molécules ce qui facilite l'analyse des séquences. De plus, certaines technologies, en évitant l'étape d'amplification PCR, donnent directement accès à de l'information épigénétique et permettent une couverture plus homogène du génome en supprimant les biais d'amplification. Chaque instrument possède cependant des limitations et il semble qu'un séquençage de grande qualité soit obtenu par la combinaison de plusieurs techniques. Ainsi, les nombreuses erreurs d'un séquençage nanopore peuvent être gommées par un séquençage illumina tout en permettant à l'utilisateur de bénéficier de l'information des reads longs. Peut-être que de futurs développements permettront à ONT de réduire le nombre d'erreurs afin de concurrencer le séquençage NGS et Illumina. La technologie PacBio a une meilleure justesse sur la lecture des séquences mais le faible débit de ses instruments, malgré les récents progrès du Sequel, risque de devenir un frein dans cette course technologique. Les techniques de cartographie optique semblent être prometteuses en complément des méthodes de séquençage afin de faciliter l'assemblage des reads et d'augmenter la taille des contigs, parfois à l'échelle d'un chromosome<sup>76</sup>.

Nous avons également identifié les besoins de chaque instrument en termes de quantité et de qualité d'ADN requise ainsi que certaines contraintes de préparation de l'échantillon afin de définir des orientations pour le développement d'un instrument de préparation dédié. Ainsi le tableau ci-dessous récapitule les besoins de chaque technologie en termes de quantité et qualité mais aussi la complexité des protocoles et l'intervention des étapes de sélection en taille (Tableau 2). On peut retenir que les technologies PacBio et ONT sont particulièrement gourmandes en ADN, et que la plupart des techniques font intervenir une étape de sélection de grands fragments. Les techniques de 10X Genomics et de Bionanogenomics sont celles qui requiert le moins d'ADN et la plus grande taille de fragment, le protocole de Bionanogenomics étant particulièrement laborieux.

**Tableau 2 : Récapitulatif des besoins en ADN de chaque technologie de 3<sup>ème</sup> génération**

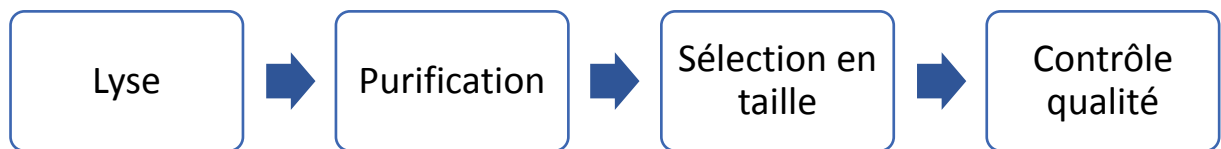
Système	10X Chromium	PacBio	Bionanogenomics	ONT
Nature analyse Couverture	Séquençage Illumina Couverture 60X	Séquençage Couverture 60X	Cartographie Couverture 100X	Séquençage Couverture 10X
Quantité Initiale	5 µg	Non proposé	3 µg	Non proposé
Durée extraction de l'échantillon	70 minutes (MagAttract)		3 jours	
Taille ADN pour la librairie	≥ 50 kb	> 20 kb	> 150 kb	> 30 kb
Quantité mini pour la librairie	20 ng	5 µg	300 ng	1 µg
Concentration pour la librairie	1 ng/µL	40 ng/µL	40 ng/µL	20 ng/µL
Sélection en taille	MagAttract + BluePippin si nécessaire	Oui	Non	Oui
Qualité de l'ADN	Sensible	Sensible	Très sensible	Selon application



## 2. Préparation et contrôle qualité d'ADN de haut poids moléculaire

---

La préparation des échantillons d'ADN est aujourd'hui une étape cruciale à l'obtention de données de séquençage de qualité. Cette préparation comprend la lyse des cellules pour libérer l'ADN, suivie d'une purification, d'une sélection en taille des fragments si nécessaire, et des procédures de contrôle qualité de l'échantillon (Figure 9).



**Figure 9 : Etapes successives de préparation d'un échantillon avant un séquençage.**

Toutes ces étapes mènent à la création des bibliothèques génomiques, lesquelles vont ensuite constituer le matériel de départ pour le séquençage. Cependant cette préparation d'ADN de haut poids moléculaire reste une étape complexe avec les techniques de biologie moléculaire actuelles. La communauté est donc en attente de réelles innovations concernant la manipulation, l'analyse et la préparation de grands fragments d'ADN remplissant le cahier des charges des séquenceurs de 3<sup>ème</sup> génération afin d'utiliser leur plein potentiel.

Au cours de ce chapitre nous aborderons d'abord les différentes méthodes permettant d'extraire de l'ADN de haut poids moléculaire. Ensuite nous nous intéresserons aux techniques de sélection en taille de l'ADN, puis aux méthodes analytiques de contrôle qualité des échantillons avant séquençage.

## 2.1 Extraction d'ADN de haut poids moléculaire

L'extraction d'ADN comprend généralement deux phases :

- La lyse des cellules ou tissus pour libérer l'ADN par un actionnement chimique (digestion enzymatique, utilisation de détergent, choc osmotique) ou physique (ultrasonication, pression). La méthode de lyse varie énormément d'un organisme à un autre et influence la qualité de l'ADN obtenu. Il est par exemple nettement plus complexe d'extraire un ADN de qualité à partir de plantes qu'à partir de bactéries Gram négatif ou de cellules eukaryotes<sup>77</sup>. Dans cette partie nous n'aborderons pas en détails les différents processus de lyse cellulaire. Pour plus d'informations sur le sujet, la revue suivante donne une bonne vue d'ensemble des méthodes de lyse macro et microfluidiques<sup>78</sup>.
- La purification de l'ADN via l'élimination des différents contaminants cellulaires comme les lipides, glucides et protéines, notamment les nucléases et les histones afin de protéger et libérer l'ADN respectivement.

Dans cette partie nous discuterons les méthodes les plus couramment utilisées pour isoler de l'ADN de haut poids moléculaire ainsi que les caractéristiques de l'échantillon obtenu en termes de pureté, taille et rendement. Nous avons identifié quatre principales familles de méthodes exploitant la densité de l'ADN, sa charge, sa solubilité ou sa taille, présentées en 2.1.1 à 2.1.4 (Figure 10).

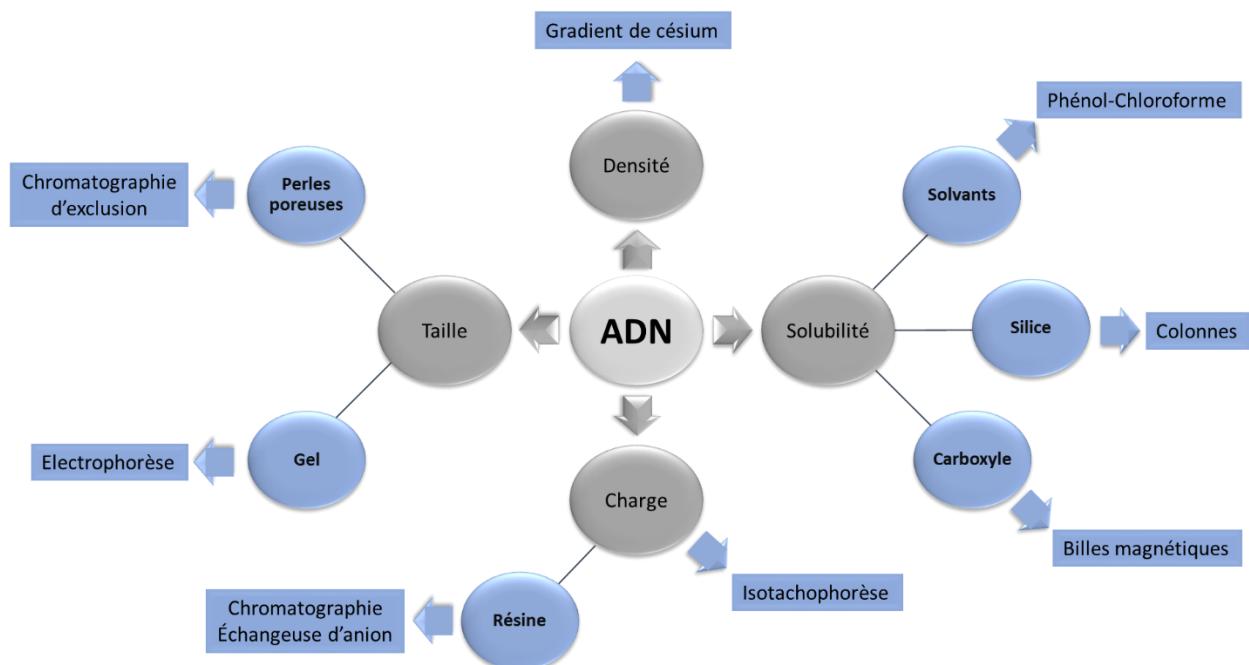


Figure 10 : Différentes caractéristiques de la molécule d'ADN utilisées pour sa purification. Inspiré de <http://purigenbio.com/the-solution/>.

## 2.1.1 Purification liquide et chimique

Il s'agit des méthodes historiques d'isolation d'ADN, la plupart étant assez laborieuses, nécessitant du savoir-faire et pouvant utiliser des produits chimiques relativement toxiques. Cependant, dans l'optique de produire de l'ADN de haut poids moléculaire pour la 3<sup>ème</sup> génération, certains protocoles ont été remis au goût du jour notamment pour leur capacité à produire de l'ADN de grande taille et de manière peu coûteuse. Ces méthodes se fondent sur des différences de solubilités entre les composants cellulaires et l'ADN dans différents tampons et solvants.

### 2.1.1.1 Phénol-Chloroforme

C'est la première méthode utilisée pour la purification d'ADN, développée en 1956 par Kirby<sup>79</sup>. Les protéines dénaturées par le phénol, les lipides et d'autres contaminants cellulaires ayant une solubilité importante dans le phénol sont solubilisés dans la phase organique du mélange. Au contraire, l'ADN qui n'est pas soluble dans le phénol reste dans la phase aqueuse. Bien que légèrement soluble dans l'eau, le phénol peut être séparé par centrifugation par différence de densité. L'ajout de chloroforme, dans lequel le phénol est soluble, permet d'améliorer la séparation de phase. Plusieurs lavages sont généralement nécessaires pour obtenir une pureté suffisante. Cette méthode permet de récupérer de grandes quantités d'ADN d'environ 150 kb lorsqu'elle est faite par un utilisateur expérimenté<sup>80</sup>. A noter qu'en jouant sur les conditions de pH et via l'ajout de Thyocyanate de guanidium, il est possible de séparer l'ADN de l'ARN<sup>81</sup>. Cette méthode a également l'avantage d'éliminer efficacement les nucléases, dénaturée par le thyocyanate de guanidium. Une fois l'ADN séparé des protéines, il peut être purifié et concentré par précipitation à l'éthanol ou l'isopropanol puis solubilisé dans un faible volume de tampon. Cette méthode a progressivement été abandonnée à mesure que les techniques de purification sur colonne se sont développées (voir 2.1.2.1), en particulier à cause de la toxicité des solvants organiques utilisés. Elle connaît un regain d'intérêt pour sa capacité à produire des grands fragments d'ADN de bonne qualité pour le séquençage 3<sup>ème</sup> génération. Cette méthode a notamment été utilisée avec succès pour du séquençages Nanopore<sup>44,82</sup>, PacBio<sup>83,84</sup> et pour de la cartographie optique<sup>84,85</sup>.

L'extraction d'ADN de plantes, plus complexe à cause de la forte proportion de polysaccharides, peut également être réalisée à l'aide de solvants organiques par l'ajout de CTAB, un détergent qui permet de précipiter les polysaccharides<sup>86</sup>.

### 2.1.1.2 *Gradient de Chlorure de Césium (CsCl)*

L'ADN génomique peut être isolé de l'ADN plasmidique et des débris cellulaires via une purification par gradient de Chlorure de Césium. La méthode consiste à incorporer du Bromure d'Ethidium (BET) dans l'ADN puis d'effectuer plusieurs heures d'ultracentrifugation dans un gradient de densité de CsCl. Le BET s'incorpore moins dans les ADN circulaires qui sont ainsi situés dans une zone de faible densité du gradient alors que l'ADN génomique se trouve dans la partie où la densité du gradient est élevée. L'utilisation de solvant permet ensuite d'éliminer le BET et l'ADN peut être précipité et concentré par l'éthanol. Cette méthode permet d'obtenir de l'ADN pure de grande taille mais nécessite de longues étapes d'ultracentrifugation et fait intervenir le BET, un intercalant cancérigène. De plus, les intercalants peuvent provoquer des dégâts et des cassures sur l'ADN<sup>87</sup>. Cette méthode est surtout utilisée pour la purification d'ADN plasmidique<sup>88</sup>.

## 2.1.2 Immobilisation réversible sur phase solide (SPRI)

La SPRI est la méthode la plus couramment rencontrée dans les laboratoires pour l'extraction d'acides nucléiques, très souvent vendue en kits commerciaux<sup>89</sup>. Ces méthodes sont pratiques, faciles à mettre en œuvre et permettent de purifier l'ADN avec de bons rendements. Elles comprennent généralement quatre étapes : lyse des cellules, adsorption de l'ADN sur une phase solide, lavages puis élution de l'ADN.

Une fois l'ADN libéré des cellules, il peut s'adsorber sur une phase solide dans certaines conditions de pH et de salinité<sup>90</sup>. Parmi les supports les plus utilisés pour fixer l'ADN, on trouve les colonnes, les membranes ou encore les billes magnétiques dont nous détaillerons le fonctionnement ci-après. Ces supports sont fonctionnalisés avec de la silice, des surfaces échangeuses d'anions ou des groupements carboxyles afin d'adsorber l'ADN. Des lavages successifs à l'éthanol ou à l'isopropanol permettent ensuite d'éliminer les contaminants cellulaires et les sels. Enfin l'ADN peut être élué grâce à un tampon approprié dans lequel il aura une solubilité importante. L'inconvénient de ces kits est qu'ils peuvent être difficiles à automatiser et demandent l'intervention d'un utilisateur dont l'expertise peut jouer sur le rendement.

### 2.1.2.1 *Purification sur colonne*

Les purifications sur colonnes sont effectuées via la fixation de l'ADN sur une membrane en silice ou sur des résines échangeuses d'anion<sup>91</sup>. L'adsorption de l'ADN sur une surface de silice n'est pas encore complètement comprise et plusieurs effets lui sont attribués. En effet, un tampon Tris-EDTA (TE) de pH ~8 est utilisé dans la plupart des conditions expérimentales de manipulation d'ADN. Or, à ce pH, l'ADN et les groupements silanols de la silice sont chargés négativement, conduisant à une

forte répulsion électrostatique. Afin de favoriser l'interaction ADN-silice sont utilisés des tampons de faible pH afin de réduire la densité de charge négatives de la silice et la répulsion électrostatique avec l'ADN. De même, l'ajout de sels comprime la double couche électrique ce qui facilite également l'adsorption de l'ADN. L'ajout d'agents chaotropes comme le chlorure de guanidium permet également la fixation de l'ADN sur la silice<sup>92</sup>. Cette molécule perturbe la couche d'hydratation de l'ADN et ainsi facilite sa fixation sur la membrane via des interactions hydrophobes. Elle joue aussi un rôle de dénaturant protéique qui inactive les DNases. Ensuite, des lavages successifs sont effectués avec de l'éthanol qui permet d'éliminer les contaminants (protéines, lipides, carbohydrates) à travers la membrane pendant que l'ADN reste fixé sur la phase solide. L'ADN peut ensuite être élué à l'aide d'un tampon de faible force ionique et avec un pH alcalin comme le TE. A chaque étape, une centrifugation est effectuée afin d'éliminer les différents contaminants et d'éluer un maximum d'ADN. Malheureusement cette étape induit un fort cisaillement et conduit à une fragmentation de l'ADN jusqu'à 50 kb environ avec la plupart des kits. Cette méthode n'est donc pas bien adaptée à la purification d'ADN de haut poids moléculaire. La plupart des colonnes peuvent purifier de 5 à 10 µg d'ADN en environ 1 h pour un utilisateur expérimenté. La parallélisation est assez aisée même si ce genre de kit est très manuel et difficilement automatisable.

Afin de limiter le cisaillement de l'ADN, certaines colonnes fonctionnent uniquement par gravité, ce qui permet d'obtenir un ADN de meilleur qualité, d'environ 100 kb mais avec des temps de préparation beaucoup plus longs<sup>93</sup>.

#### 2.1.2.2 Billes magnétiques

Les billes magnétiques sont généralement composées de polystyrène et de magnétite avec un coating de groupements carboxyle ou en silice<sup>94</sup>. Sur le même principe que la purification sur colonne, en présence de PEG ou d'agents chaotropes, l'ADN se fixe sur les groupements carboxyle ou la silice respectivement, ce qui permet d'effectuer les lavages<sup>95</sup>. L'utilisation d'un aimant permet d'agglomérer les billes sur une paroi du tube et d'éliminer facilement le surnageant lors des lavages. L'utilisation de billes offre aussi un bon rapport surface/volume et donc une bonne capacité de capture. Cette méthode permet d'éviter les étapes de centrifugation qui dégradent l'ADN via des lavages plus doux tout en conservant une purification assez rapide.

Qiagen a récemment commercialisé le kit MagAttract pour la préparation d'ADN de haut poids moléculaire. Ce kit permet d'obtenir entre 4 et 8 µg d'ADN d'environ 150 kb à partir de 200 µL de sang en 70 min. Il peut de plus être automatisé avec des robots<sup>95</sup> (NucliSENS® de BioMérieux ou MagNA Pure de Roche). Ce kit est d'ailleurs recommandé par 10X Genomics dans l'optique d'effectuer un séquençage Chromium à partir de sang, de bactéries ou de tissus. Il n'est cependant pas adapté à



tous les types cellules et organismes, ainsi il ne permet pas d'extraire efficacement de l'ADN issu de plantes. Les rendements du kit sont de l'ordre de 80% mais, comme avec la plupart de ces kits commerciaux, la qualité de la purification et son rendement peuvent varier fortement d'un utilisateur expérimenté à un débutant. Ce kit a d'ailleurs été testé durant cette thèse pour l'extraction d'ADN génomique issu de cellules SH-S5Y5. Les rendements étaient très faibles, de l'ordre de 10% pour des fragments d'une taille comprise entre 50 et 150 kb.

Ce kit a tout de même été utilisé avec succès pour des séquençages 10X Genomics<sup>96</sup> et Nanopore<sup>97</sup>. Il est assez coûteux, de l'ordre de 5 euros par échantillon, ce qui a poussé certains groupes à développer leurs propres protocoles utilisant des billes magnétiques afin d'obtenir un ADN de qualité, issu de différents organismes et pour un coup moindre que les kits commerciaux<sup>77</sup>.

Des optimisations ont également été faites sur la structure des billes en utilisant notamment des disques composés de strates successives de lamelles de silices de quelques nanomètres d'épaisseur (Figure 11, A). Ils permettent d'augmenter la ratio surface/volume de capture d'ADN tout en limitant le cisaillement<sup>98</sup>. Cette innovation technologique a conduit à la commercialisation d'un kit vendu par Circulomics appelé Nanobind, spécialement dédié à la génomique de 3<sup>ème</sup> génération. L'entreprise met en avant de meilleures performances que les méthodes classiques d'extraction tout en gardant leur simplicité. Il permet la purification de quelques milligrammes d'ADN entre 50 et 300 kb, sans coupure simple brin ou sites abasiques en 1 h environ. Ces très bonnes performances ont un coût d'environ 20 \$/extraction<sup>99</sup>. On note cependant la présence minoritaire de fragments inférieurs à 50 kb après la purification pouvant limiter les performances de séquençage (Figure 11, B). Cette méthode d'extraction a été utilisée pour des séquençages PacBio<sup>100</sup> et Nanopore<sup>101</sup>, une sélection en taille ayant été réalisée avant le séquençage Nanopore pour éliminer les petits fragments.

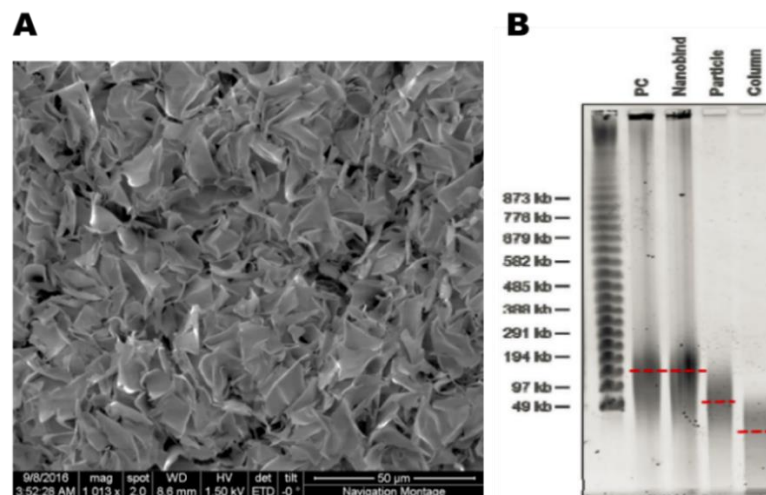


Figure 11 : A : Image MEB de la surface de silice nanostructurée Nanobind. B : Comparaison de la qualité de l'ADN extrait selon différentes méthodes. (PC) Purification phénol-chloroforme, (Nanobind) Purification avec le kit développé par Circulomics, (Particle) Purification avec des billes magnétiques (Column), Purification sur colonne avec centrifugation. Adapté Zhang et al.<sup>98</sup>.

### 2.1.2.3 Place de la microfluidique en SPRI

Certaines équipes ont travaillé à utiliser la microfluidique pour la préparation d'échantillon en amont d'analyse génomique. L'objectif de la miniaturisation est de limiter les contaminations et l'utilisation de réactifs, d'accélérer les protocoles d'extraction, ainsi que de limiter l'intervention de l'utilisateur afin de les utiliser sur le terrain<sup>102</sup>. La matrice solide, généralement de la silice, est alors insérée à l'intérieur de la puce et la purification fonctionne selon le même principe que décrit précédemment en colonne ou sur billes<sup>103</sup>.

L'ADN génomique peut également être immobilisé à l'intérieur d'une puce contenant des réseaux de micro pylônes dans lesquels l'ADN s'enchevêtre. Les débris cellulaires et petits fragments non enchevêtrés peuvent être évacués de la puce ce qui permet la purification. Certaines publications rapportent une efficacité d'extraction proche de 100 % sur 0,1 à 1 ng d'ADN extrait<sup>104</sup>. Ces faibles quantités collectées limitent cependant les étapes de caractérisation des échantillons. En effet, les instruments de contrôle qualité de l'ADN ont souvent des limites de sensibilité assez haute, de l'ordre du ng/ $\mu$ l comme nous le verrons dans la troisième partie de ce chapitre (voir 2.3).

A l'heure actuelle, aucune technologie microfluidique n'a vraiment apporté de réponse satisfaisante et globalement acceptée par la communauté à la préparation d'ADN de haut poids moléculaire. La miniaturisation des systèmes et les faibles volumes manipulés ne sont pas favorables aux besoins des étapes préparatives, i.e., extraire de grandes quantités de matière. Certaines équipes ont tout de même fabriqué des puces permettant de miniaturiser les étapes d'extraction et de préparation de librairie avant un séquençage NGS. Snider et al. ont récemment développé une puce intégrant ces étapes afin de préparer 660 pg d'ADN pour un séquençage Illumina<sup>105</sup>. Kim et al. ont également miniaturisé les étapes d'extraction et de préparation de librairie dans des microréacteurs. La miniaturisation permet de travailler avec de plus faibles quantités d'ADN et donc de séquencer des organismes difficilement cultivables comme certaines bactéries. Ils ont ainsi montré le séquençage Illumina de *Mycobacterium Tuberculosis* à partir de centaines de pg d'ADN<sup>106</sup>. Ces systèmes microfluidiques sont cependant uniquement compatibles avec des technologies NGS car les quantités d'ADN extraites sont trop faibles pour utiliser les techniques de 3<sup>ème</sup> génération décrites au premier chapitre.

Ainsi, de nombreuses publications rapportant l'extraction d'ADN en puces microfluidiques sont assez éloignées des quantités requises par les techniques de génomiques de 3<sup>ème</sup> génération (de ~20 ng à 5  $\mu$ g) ou possèdent de faibles débits de l'ordre du  $\mu$ L/h<sup>107-111</sup>. Afin d'obtenir une place parmi les techniques préparatives, la microfluidique doit développer des techniques de purification « continue » plutôt qu'en « batch » comme la plupart des techniques macrofluidiques actuelles<sup>112</sup>.

### 2.1.3 Extraction en « plugs » d'agarose

Lorsqu'une grande qualité d'ADN est nécessaire, il devient essentiel de limiter le cisaillement subi par les molécules d'ADN. Le travail en phase liquide est alors à proscrire et l'extraction se fait dans des matrices de gels, le plus souvent en agarose. Cette méthode consiste à encapsuler dans des « plugs » d'agarose les cellules ou noyaux afin de les lyser<sup>113</sup>. L'ADN génomique, de plusieurs Mb, se retrouve alors piégé dans cette matrice d'agarose, les pores étant trop petits pour que les grandes molécules d'ADN s'en échappent. En revanche les petites molécules, comme les sels et les contaminants cellulaires, peuvent diffuser librement à travers l'agarose<sup>114</sup>. On peut ainsi, en plaçant le plug dans différents tampons, effectuer la lyse des cellules et des noyaux, puis la purification de l'ADN. Le grand intérêt de cette méthode est d'obtenir des fragments jusqu'à plusieurs Mb. L'inconvénient majeur est le temps de purification, d'environ 3 jours pour la plupart des protocoles par un manipulateur expérimenté<sup>115</sup>. En effet, la diffusion des débris cellulaires à travers la matrice d'agarose peut être longue. De plus la collecte de l'échantillon d'ADN purifié peut s'avérer complexe car il faut l'extraire de la matrice agarose. Cela peut être fait via la dissolution de l'agarose par chauffage et grâce à l'utilisation d'une enzyme, l'agarase, qui digère l'agarose. Cependant, cette étape peut conduire à une fragmentation de l'ADN et son efficacité limitée entraîne souvent des débris d'agarose en solution pouvant perturber certaines applications. L'autre moyen d'extraire l'ADN de l'agarose est une électroélution, consistant à appliquer un champ électrique à travers le gel pour en faire sortir les fragments d'ADN purifiés par électrophorèse.

Ce protocole d'extraction est notamment recommandé par Bionanogenomics pour l'obtention de longues molécules d'ADN de qualité suffisante pour réaliser de la cartographie optique<sup>116,117</sup>. La méthode est également utilisée pour l'obtention de reads longs en séquençage Nanopore et PacBio<sup>77</sup>.

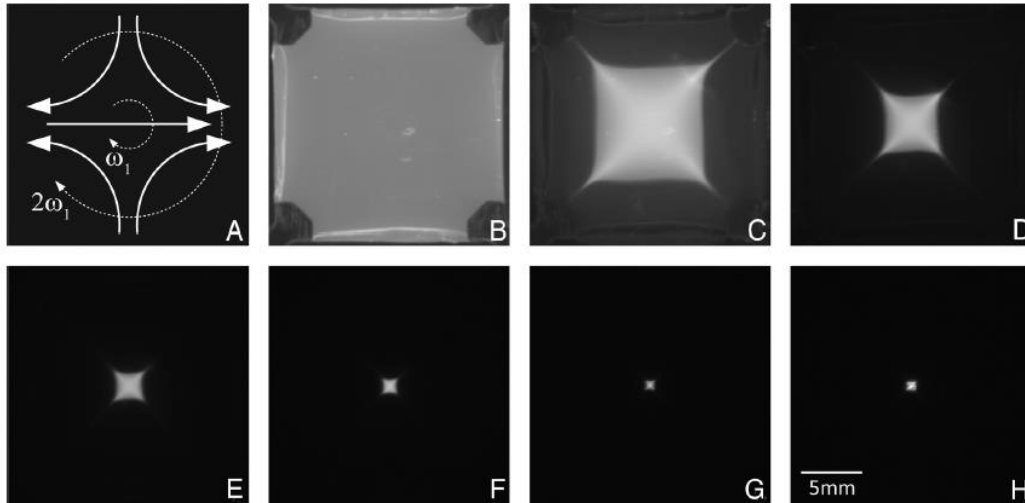
Les extractions en plug d'agarose étant particulièrement laborieuses, certaines entreprises ont développé des méthodes automatisées permettant de limiter l'intervention de l'opérateur. On peut citer la technologie développée par SageSciences, le SageHLS, qui permet d'effectuer des extractions d'ADN de haut poids moléculaire dans des cassettes de gel d'agarose pré-coulées. Jusqu'à 4 cassettes contenant 2 échantillons peuvent être utilisées par cycle d'environ 6h. Les cellules ou noyaux sont déposés dans un puits de la cassette avec les tampons de lyse. Une fois la lyse effectuée, les débris cellulaires sont évacués par électrophorèse à travers le gel pendant que l'ADN extrait, de plusieurs Mb, reste piégé. Afin d'extraire l'ADN, il est nécessaire de le fragmenter à l'aide d'enzymes de restrictions jusqu'à une taille lui permettant de migrer à travers le gel. Cette étape de purification par électrophorèse permet un gain de temps important comparé aux extractions en plug d'agarose où l'élimination des contaminants est réalisée par diffusion. La fonction champ pulsé de l'instrument

permet de fractionner grossièrement l'ADN en fonction de la taille et d'éluier l'échantillon dans 6 puits latéraux. Cette automatisation a cependant un coût non négligeable, outre l'instrument, les consommables sont associés à un prix de 50 \$ par extraction. De plus les rendements d'extraction n'atteignent pas ceux de la SPRI, SageScience revendiquant la purification de 3 µg d'ADN pour le chargement de 10 µg dans le système. L'instrument apporte d'autres fonctionnalités que nous détaillerons plus loin dans ce manuscrit comme la possibilité d'effectuer de la sélection en taille (voir 2.2.2.2).

## 2.1.4 Purification électrophorétique

### 2.1.4.1 SCODA

Des méthodes électrophorétiques ont été développées pour la purification d'ADN génomique. Nous pouvons notamment citer la technologie SCODA pour « synchronous coefficient of drag alteration<sup>118</sup> », développé par Marziali et al. en 2005 et commercialisé par Boreal Genomics. La méthode repose sur la relation non exactement linéaire entre la vitesse de migration de l'ADN et le champ électrique. Elle consiste à appliquer un champ électrique rotatif uniforme de fréquence  $\omega$  et un champ électrique rotatif quadripolaire de fréquence  $2\omega$ . La migration moyenne de l'ADN est dirigée vers le centre de ces champs électriques rotatifs et donc vers le centre du gel induisant la concentration de l'ADN. L'échantillon est déposé dans une cassette de gel d'agarose et l'ADN purifié est collecté au centre du gel (Figure 12). L'ADN peut ainsi être séparé de tous les contaminants mais également des petites molécules d'ADN, non affectées par ce phénomène. Cette technique permet de purifier jusqu'à 5 mL d'échantillon avec de très bons rendements et ainsi de purifier des échantillons très dilués<sup>119</sup>. L'ADN de bas poids moléculaire n'étant pas suffisamment affecté pour être concentré, la technique permet de collecter jusqu'à 20 µg d'ADN entre 50 kb et 1 Mb, le gel limitant le cisaillement sur les grandes molécules. La technique est implémentée sur un automate qui a cependant des temps de purification très long, de l'ordre de 30 h par échantillon<sup>120</sup>. Le prix de purification/échantillon n'est pas précisé. Une seule publication mentionne l'utilisation du système pour la purification et la concentration d'ADN de haut poids moléculaire pour du séquençage Nanopore<sup>121</sup>.



**Figure 12 : Processus de concentration de l'ADN au centre d'un gel d'agarose par la méthode SCODA. Un fragment de 2,7 kb est marqué en fluorescence et injecté dans le gel à une concentration de 0,2 ng/μL. Sous l'application des deux champs électriques l'ADN se concentre au centre du gel à une concentration de 100-200 ng/μL. Images prises toutes les 10 min.**

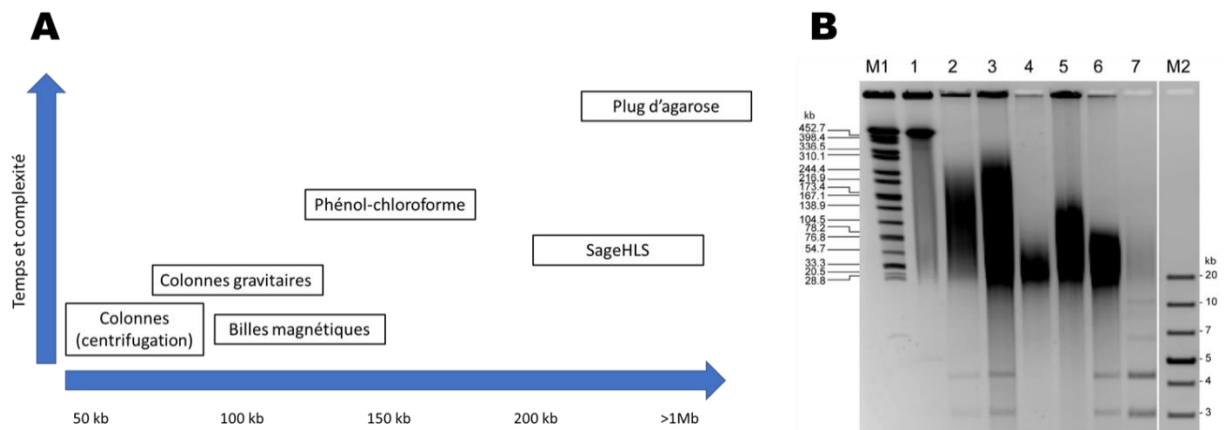
#### 2.1.4.2 Isotachophorèse

L'isotachophorèse est une technique électrophorétique initialement développée pour séparer et concentrer des espèces chargées dans un capillaire en utilisant des zones à mobilité ioniques variables pour focaliser l'échantillon. La zone contenant l'échantillon est encadrée par un ion pilote, avec une mobilité supérieure à celle des autres ions en solution, et un ion terminal, ayant une mobilité inférieure. Lors de l'application d'un potentiel électrique, un champ électrique fort est généré dans la zone terminale alors qu'un champ électrique faible est généré dans la zone pilote. Ainsi l'espèce à séparer migre plus rapidement dans la zone terminale que dans la zone pilote, conduisant à un état stable où l'échantillon est concentré à l'interface des deux zones.

L'équipe de Santiago et al. a montré la possibilité d'extraire un échantillon d'ADN par isotachophorèse<sup>122</sup>. Les ions pilotes et terminaux peuvent être choisis tels que seul l'ADN est concentré pendant que les débris cellulaires et les tampons de lyse migrent soit dans la direction opposée soit restent dans la zone terminale. La société Purigen commercialise un instrument sur ce principe d'extraction d'ADN et revendique une approche plus simple que la concurrence. Ce dispositif est censé permettre la quantification, être automatisable et limiter l'intervention de l'expérimentateur. Cependant assez peu de données sont disponibles concernant la pureté de l'échantillon obtenu, les rendements d'extraction ou sur les quantités d'ADN qu'il est possible de collecter. L'extraction se faisant dans un volume très réduit, il est probable que les quantités d'ADN manipulées soient faibles. Aucune publication ne mentionne d'ailleurs l'utilisation de ce système pour du séquençage 3<sup>ème</sup> génération.

## 2.1.5 Comparaison des méthodes d'extraction

Nous avons vu dans cette partie que de nombreuses méthodes permettent d'extraire de l'ADN. Cependant toutes les méthodes ne sont pas adaptées à la production d'ADN de haut poids moléculaire compatible avec du séquençage de 3<sup>ème</sup> génération. Les techniques les plus utilisées actuellement sont le kit MagAttract de Qiagen, recommandé notamment par 10X Genomics, l'extraction au phénol-chloroforme malgré l'utilisation de solvants toxiques, et les plugs d'agarose, laborieux, mais permettant d'obtenir la plus grande qualité d'ADN (Figure 13, A). On peut cependant noter que la plupart de ces méthodes conduisent à une fragmentation plus ou moins importante de l'ADN génomique avec la présence de fragments résiduels jusqu'à 20-30 kb (Figure 13, B). Il est alors nécessaire d'utiliser des méthodes de sélection en taille afin d'éliminer ces fragments de dégradation.



**Figure 13 : A : Comparaison des différentes méthodes d'extraction d'ADN en fonction de la taille de l'ADN obtenu et de la complexité de la méthode. B : Gel d'électrophorèse en champs pulsés après extraction d'ADN génomique de *K-Pneumoniae* 234-12 avec différentes méthodes. (M1-M2) Références de taille. (1) Extraction en plus d'agarose. (2) Extraction en colonne gravitaire avec le kit Genomic-tip de Qiagen. (3) Extraction avec billes magnétiques avec le kit MagAttract de Qiagen. (4) Extraction par précipitation au sel avec le kit MasterPure d'Epicentre. (5) Extraction par précipitation au sel avec le Kit Wizard Genomic de Promega. (6) Extraction en colonne avec centrifugation avec le kit DNeasy Blood and Tissue de Qiagen. (7) Extraction en colonne gravitaire le Plasmid Mini kit de Qiagen. Tiré de Becker et al.<sup>123</sup>.**

## 2.2 Sélection en taille d'ADN de haut poids moléculaire

L'étape d'extraction ne conduit pas toujours à l'obtention d'ADN de taille suffisamment grande et homogène pour réaliser un séquençage de qualité. Il est alors nécessaire d'effectuer une étape de sélection en taille de l'échantillon purifié<sup>124</sup>.

### 2.2.1 Billes magnétiques

Comme nous l'avons vu précédemment, l'ADN peut s'adsorber à la silice dans certaines conditions de pH et salinité. Or, plus une molécule d'ADN est grande et plus son affinité avec la silice est importante. Cette propriété a été utilisée par les entreprises commercialisant les kits pour effectuer de la sélection en taille après purification. Cette sélection s'effectue en jouant de manière précise sur la composition des tampons et sur le rapport nombre de billes/quantité d'ADN afin d'effectuer une élution ménagée avec un tampon de salinité intermédiaire permettant d'éliminer les plus petits fragments de l'échantillon<sup>125</sup>. Cette méthode de sélection en taille est notamment utilisée dans les protocoles de NGS pour l'élimination des amorces PCR, des chimères et autres petits fragments indésirables. On peut notamment citer le kit AMPure XP vendu par Beckman Coulter dans cette optique. La sélection en taille reste cependant assez peu précise et cible principalement des petits fragments d'ADN (< 1kb)<sup>126</sup>. Même si cela n'est pas clairement précisé, le protocole du kit MagAttract de Qiagen semble comprendre une étape d'élution intermédiaire pour l'élimination des petits fragments formés lors de l'extraction.

### 2.2.2 Gels d'agarose

Une sélection en taille peut également être effectuée grâce à une électrophorèse en gel d'agarose. Cette technique très classique de biologie moléculaire permet de séparer les acides nucléiques en fonction de leur taille dans une matrice d'agarose (voir 2.3.3.1). Il est ensuite possible de découper manuellement le gel afin d'isoler les fragments de la taille désirée via une électroélution ou une digestion de l'agarose<sup>90</sup>. Cette pratique est extrêmement laborieuse et manuelle, en particulier pour isoler de l'ADN de haut poids moléculaire. En effet, un gel d'électrophorèse classique ne permettra pas de séparer des fragments supérieurs à 40 kb environ. Au-delà, il est nécessaire d'utiliser des gels en champ pulsé dont les temps de migration sont très longs, jusqu'à 24h (voir 2.3.3.1). Le découpage des bandes d'intérêts peut également être complexe si l'ADN ne doit pas subir de marquage en fluorescence. Il est alors nécessaire d'analyser en parallèle une référence de taille, marquée, puis de découper le gel « à l'aveugle » à l'aide de cette référence. Cette méthode de sélection en taille manque de fidélité comme nous le verrons au cours du Chapitre 4. De plus, des contaminations sont possibles entre les différentes lignes du gel. Elle possède tout de même l'avantage

d'être facile à mettre en œuvre et très démocratisée ainsi que de permettre une sélection selon une fenêtre de taille accordable<sup>127</sup>.

La sélection en taille en gel d'agarose a été automatisée par plusieurs entreprises. Nous citerons deux technologies, le système Nimbus commercialisé par Coastal Genomics et le BluePippin développé par SageScience.

#### 2.2.2.1 *Nimbus*

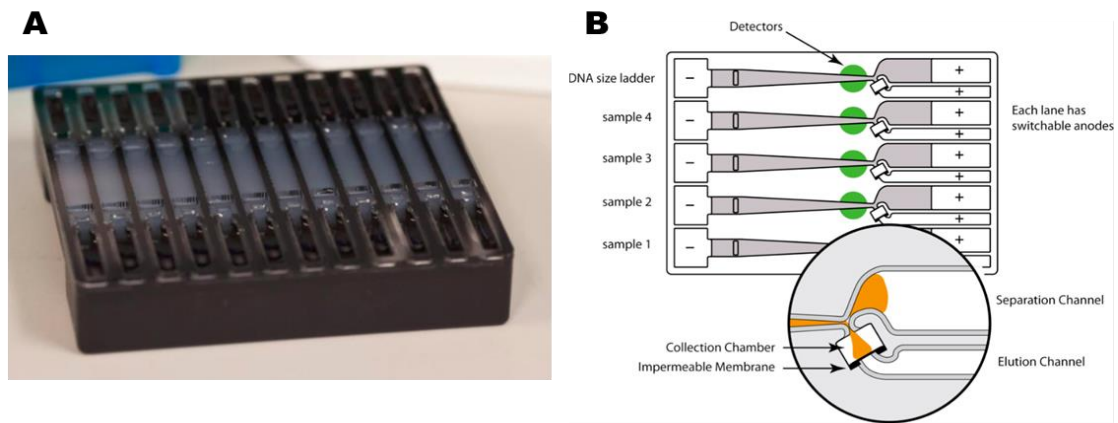
Cet instrument est une plateforme robotisée permettant de séparer l'ADN en gel d'électrophorèse et de le sélectionner en taille. Il utilise des cassettes d'agarose pré-coulées permettant d'analyser douze échantillons en parallèle. Chaque canal contient un puits pour déposer l'échantillon et un second pour le collecter (Figure 14, A). La migration de l'ADN est observée par fluorescence en temps réel ce qui permet de connaître le moment où la fraction d'intérêt entre dans le puits de collecte. Juste avant qu'elle n'entre, la migration est interrompue afin de le vider pour éliminer les petits fragments indésirables. La migration reprend ensuite et la fraction d'intérêt est collectée. Les volumes de collecte sont assez importants, d'environ 400  $\mu$ L pour l'ADN de haut poids moléculaire. L'instrument permet de faire de la sélection en taille de 50 bp à 30 kb avec une résolution allant de 10 bp à 5 kb. Il ne possède pas de fonction « champ pulsé » ce qui limite son utilisation pour la manipulation d'ADN de haut poids moléculaire. Il autorise néanmoins l'élimination de fragments inférieurs à 20 kb et l'isolation des plus grands fragments de l'échantillon. Jusqu'à 500 ng d'ADN génomique sont traités par canal pour un rendement d'environ 75%. Le marquage fluorescent de l'ADN par des intercalants peut être problématique pour certaines applications. Le principal avantage de cet instrument est son automatisation, du chargement à la collecte de l'échantillon, et son haut débit pouvant traiter 96 échantillons en parallèle<sup>128</sup>. Le prix de la plateforme n'est pas indiqué mais il s'agit d'un instrument assez encombrant et probablement destiné à des plateformes haut débit. Aucune publication ne mentionne son utilisation de ce système pour du séquençage 3<sup>ème</sup> génération.

#### 2.2.2.2 *BluePippin*

L'instrument est également fondé sur une électrophorèse en gel d'agarose avec collecte par électroélution. Des cassettes avec gel d'agarose pré-coulé contiennent cinq couloirs d'électrophorèse dont un pour analyser une référence de taille. Les canaux ont une forme conique amenant à une bifurcation en Y : une branche du Y part vers une poubelle, l'autre vers une chambre d'élution où l'ADN est retenu par une membrane d'ultrafiltration (Figure 14, B). L'instrument possède une fonction champ pulsé qui permet d'analyser des fragments d'ADN de 100 bp à 50 kb en fonction de la concentration en agarose dans la cassette (de 3% à 0,75%). Un détecteur optique est placé avant le Y afin d'identifier



les espèces à l'aide d'un marqueur interne et orienter la migration vers une voie ou l'autre selon la fraction à collecter. L'utilisateur peut également utiliser une référence de poids moléculaire externe, marquée par fluorescence, dans un des cinq canaux. Elle est analysée en parallèle des échantillons afin de connaître la vitesse de migration des différents fragments. Pour chaque canal, l'utilisateur indique la taille de début et la taille de fin de la collecte, en paire de bases. Il entre aussi la « méthode » d'électrophorèse, préprogrammée en fonction de la taille, et parfois de la résolution, voulue. La Justesse pour la sélection en taille est de 20% pour les grands fragments et 10 % pour l'élimination de petits fragments



**Figure 14 : A : Cassette d'agarose du système Nimbus contenant 12 couloirs d'électrophorèse. B : Schéma d'une cassette de l'instrument BluePippin contenant 5 couloirs d'électrophorèse.**

La sélection en taille de grands fragments d'ADN dure environ 10 h, hors étapes de préparation de l'échantillon et de calibration de l'appareil. Il est ensuite nécessaire de collecter et de purifier l'ADN, étapes qui ajoutent du temps d'expérimentation. En effet, un des inconvénients du système est son faible rendement sur l'ADN de haut poids moléculaire, entre 20 et 50%, car les fragments ont tendance à rester accroché à la membrane dans le puit de collecte. Afin d'améliorer ces rendements, SageScience préconise d'attendre au moins 45 min à la fin de la sélection en taille afin de laisser diffuser l'ADN. Après collecte des 40 µL d'échantillon, il est ensuite conseillé d'effectuer deux lavages avec une solution de Tween 20 à 0,1% afin d'éluer plus efficacement l'ADN. Ces collectes successives diluent l'échantillon et la présence de détergent peut perturber certaines applications postérieures. Une étape de purification/concentration sur billes magnétiques peut alors être effectuée pour augmenter la concentration de l'échantillon et l'éluer dans un tampon adapté. Ainsi la durée totale d'une sélection en taille efficace est relativement longue, d'au moins une journée. Les rendements étant assez faibles, il est important de charger au moins 500 ng d'ADN dans le système ce qui rend l'appareil peu adapté aux échantillons disponibles en faibles quantités. La plateforme de séquençage de Toulouse (GeT-PlaGe) a également remarqué que le BluePippin avait tendance à abîmer l'ADN en

créant des coupures simples brin. Ces coupures peuvent perturber les applications où l'ADN subit un passage simple brin comme le Chromium de 10X Genomics ou le nanopore. SageScience a récemment commercialisé une cassette spécialement adaptée à l'élimination des petits fragments d'ADN avant séquençage 3<sup>ème</sup> génération qui permettrait de réduire le temps de fonctionnement à 5 h, d'augmenter le chargement maximal à 10 µg et d'améliorer les rendements. Les cassettes d'agarose sont vendues 52 ou 60 \$ pour quatre échantillons.

Le BluePippin est l'instrument de référence actuellement pour effectuer une sélection en taille avant séquençage. Il est recommandé dans les protocoles de préparation de librairie de PacBio<sup>129</sup>, 10X Genomics<sup>130</sup> et Oxford Nanopore<sup>131</sup>. Son utilisation a amélioré les performances de séquençage dans de nombreux projets utilisant ces technologies<sup>132-136</sup>.

## 2.3 Contrôle qualité d'ADN de haut poids moléculaire

Une fois l'ADN extrait et sélectionné en taille, il est nécessaire d'en effectuer un contrôle qualité afin de vérifier qu'il corresponde au cahier des charges de l'instrument à utiliser. Il est ainsi important de mesurer sa pureté, sa concentration ainsi que sa distribution en taille. Les techniques spectrophotométriques sont les principales méthodes pour évaluer la pureté d'un échantillon ainsi que sa concentration. Les techniques fluorimétriques vont permettre d'évaluer la concentration totale d'un échantillon alors que les techniques électrophorétiques permettent de déterminer la distribution en taille.

### 2.3.1 Pureté de l'échantillon – Mesure d'absorbance

La pureté de l'échantillon est un paramètre essentiel car certains contaminants peuvent perturber les étapes de séquençage. Une extraction au phénol-chloroforme risque ainsi de laisser des traces de phénol alors que des sels type thyocyanate de guanidium peuvent être présents en quantités résiduelles après une extraction SPRI<sup>137</sup>. L'analyse spectrophotométrique permet d'évaluer la pureté d'un échantillon en mesurant la concentration des acides nucléiques mais également celle des phénols, de certains sels ou encore des ARNs. La méthode se base sur la mesure de l'absorbance des oligonucléotides par les UVs à la longueur d'onde de 260 nm ( $A_{260}$ )<sup>138</sup>. Les contaminations par les protéines et les phénols sont évalués par mesure de l'absorbance à 280 nm ( $A_{280}$ ) alors que la présence de sels de guanidines, de solvants ou de certains contaminants organiques est mesurée à 230 nm ( $A_{230}$ ). La pureté de l'ADN est acceptable lorsque le ratio  $A_{260}/A_{280}$  est compris entre 1,8 et 2 et que celui  $A_{260}/A_{230}$  est supérieur à 1,8. Un ratio  $A_{260}/A_{280}$  inférieur à 1,8 indique une contamination protéique ou par des phénols alors qu'un ratio supérieur à 2 indique la présence d'ARN dans l'échantillon. La concentration de l'échantillon en ADN peut ensuite être évaluée via la loi de Beer-Lambert<sup>139</sup>. Cette méthode ne permet cependant pas d'évaluer la présence de coupures simple brin ou de sites abasiques sur le fragment d'ADN.

La spectrophotométrie a l'avantage de donner une mesure de concentration directe, sans besoin d'étalon. Toute contamination perturbe la mesure, ce qui restreint la mesure de concentration à des échantillons purifiés. L'instrument de référence dans les laboratoires est le NanoDrop (Thermo Scientific). Il permet notamment de préserver l'échantillon car une goutte de quelques microlitres (0,5 – 2  $\mu\text{L}$ ) suffit pour la mesure. La plage de fonctionnement de l'instrument est de 2 à 15000 ng/ $\mu\text{L}$  pour une mesure en goutte d'ADN double brin. Certains instruments disposent d'une mesure en cuvettes de 50  $\mu\text{L}$  ayant une sensibilité de 0,4 ng/ $\mu\text{L}$ . La sensibilité de cet instrument peut ainsi être insuffisante pour certaines applications, notamment pour la mesure de concentration avant un séquençage 10X Genomics où la concentration des échantillons est  $\sim 1$  ng/ $\mu\text{L}$ .

### 2.3.2 Concentration de l'échantillon – mesures fluorimétriques

La méthode la plus répandue pour évaluer la concentration d'un échantillon d'ADN est la mesure fluorimétrique. Elle consiste à utiliser des rapporteurs fluorescents intercalés dans la séquence ou fixés à l'extrémité de la molécule d'ADN. De nombreux agents fluorescents ont été développés pour s'intercaler le long de la molécule d'ADN, au niveau des sillons majeurs ou mineurs<sup>140</sup>, ou en bout de chaîne, le plus souvent par incorporation enzymatique de nucléotides marqués<sup>141</sup>. Les agents intercalants sont généralement spécifiques de l'ADN et leur fluorescence augmente significativement lors de l'interaction ce qui permet de détecter un signal tout en limitant le bruit de fond. La mesure de l'intensité de fluorescence est corrélée avec la concentration en ADN<sup>142</sup>. Des intercalants tels que le SYBR Green (Applied Biosystems), le SYTOX Orange (Thermo Fisher) ou le YOYO-1 sont communément utilisés en solution. La concentration de l'intercalant est un paramètre important à maîtriser afin d'avoir un ratio signal/bruit maximal. Cependant, le marquage étant dynamique, il est favorisé lorsque les constantes d'associations – qui dépendent de la concentration des espèces – sont supérieures aux constantes de dissociations<sup>143</sup>. La pureté de l'échantillon peut également affecter la justesse de la mesure même si la plupart des kits actuels sont assez tolérants à la présence de sels, ARN, protéines... Certains contaminants peuvent tout de même être des compétiteurs d'interactions avec les fluorophores et ainsi fausser les mesures<sup>127</sup>. Un autre inconvénient de cette mesure est qu'elle se base sur une intensité de fluorescence qu'il faut comparer à celle d'un échantillon de concentration connue. Cela implique d'avoir des standards parfaitement calibrés, ce qui n'est pas toujours évident. En effet, les solutions d'ADN commerciales ont parfois des différences entre les concentrations théoriques et observées jusqu'à 20% dans l'expérience de Picometrics.

Le Qubit (Life Technologies) est le fluorimètre le plus communément utilisé dans les laboratoires et différents kits permettent de quantifier l'ADN, l'ARN ou les protéines. Un kit en version haute-sensibilité permet de détecter des concentrations jusqu'à 10 pg/ $\mu$ L<sup>144</sup>, il est donc bien adapté à la quantification des échantillons avant séquençage. La quantification d'ADN de très haut poids moléculaire après une extraction peut néanmoins manquer de fidélité. En effet, si l'on travaille avec de très grands fragments d'ADN la concentration au sein de l'échantillon peut être inhomogène. Il est d'ailleurs recommandé de fragmenter l'ADN de très haut poids moléculaire afin d'obtenir une concentration plus homogène lors de la mesure.

## 2.3.3 Distribution en taille – Techniques de séparation d'ADN

### 2.3.3.1 Les gels d'agarose

Nous avons vu lors du premier chapitre l'importance de la longueur des molécules pour la réussite d'un séquençage de 3<sup>ème</sup> génération. Il est donc essentiel d'évaluer la taille des fragments d'un échantillon avant d'initier la préparation de la librairie. La plupart des méthodes de séparation de l'ADN en fonction de la taille sont électrophorétiques. En solution, la mobilité électrophorétique de l'ADN ne dépend quasiment pas de la taille des fragments<sup>145</sup>. Afin d'induire des différences de mobilité entre les fragments, il est nécessaire d'utiliser des matrices en gel de polymères, introduites depuis le milieu du 20<sup>ème</sup> siècle. Le gel crée un réseau de pores à travers lequel les petites molécules se faufilent plus facilement que les grandes, créant ainsi une dépendance de la vitesse de migration en fonction de la masse moléculaire<sup>146</sup>. La migration de l'échantillon est opérée simultanément avec un marqueur de taille calibré servant de référence pour évaluer la taille des fragments de l'échantillon. L'observation des fragments est généralement faite via l'ajout d'un intercalant ADN fluorescent, comme le bromure d'éthidium. L'intensité de fluorescence de l'échantillon peut être comparée à celle de la référence pour une estimation de la concentration. L'électrophorèse sur gel d'agarose étant une méthode simple, à faible coût et efficace, elle est massivement utilisée dans les laboratoires pour la séparation d'ADN de 50 à 25000 bp environ<sup>147</sup>. En effet, au-delà de 25 kb, toutes les molécules migrent à la même vitesse quelle que soit leur taille rendant leur séparation impossible. Le gel d'électrophorèse classique n'est ainsi pas adapté pour évaluer la distribution en taille d'ADN de haut poids moléculaire.

Les gels d'électrophorèse en champ pulsé (PFGE), où la direction du champ électrique change périodiquement, ont été développés en 1984 par Schwartz et Cantor pour l'analyse des grands fragments d'ADN<sup>148</sup>. Cette technique joue sur le temps de réorientation de l'ADN selon le champ électrique qui est dépendant de la taille des molécules. Il est alors possible de séparer des fragments jusqu'à 10 Mb<sup>149</sup>. L'inconvénient principal de la méthode est le temps d'analyse considérablement allongé, d'une heure pour un gel classique à 20 h pour un PFGE. De plus, la sensibilité de l'analyse est limitée, de l'ordre de 10 ng/ $\mu$ L pour un smear ou 20 ng pour une bande unique, ce qui la rend inadaptée pour certains instruments de séquençage de 3<sup>ème</sup> génération et consomme de fortes quantités d'échantillon. La séparation d'ADN de haut poids moléculaire est un sujet qui a été beaucoup traité par la communauté scientifique dans les années 90 et au début des années 2000 afin d'améliorer le PFGE en termes de temps d'analyse et de sensibilité mais aucune innovation de rupture n'est apparue depuis<sup>112</sup>. Comme pour les méthodes d'extractions, l'apparition de la génomique de 3<sup>ème</sup> génération a considérablement relancé l'intérêt des chercheurs dans les méthodes d'analyses d'ADN de haut poids moléculaire.

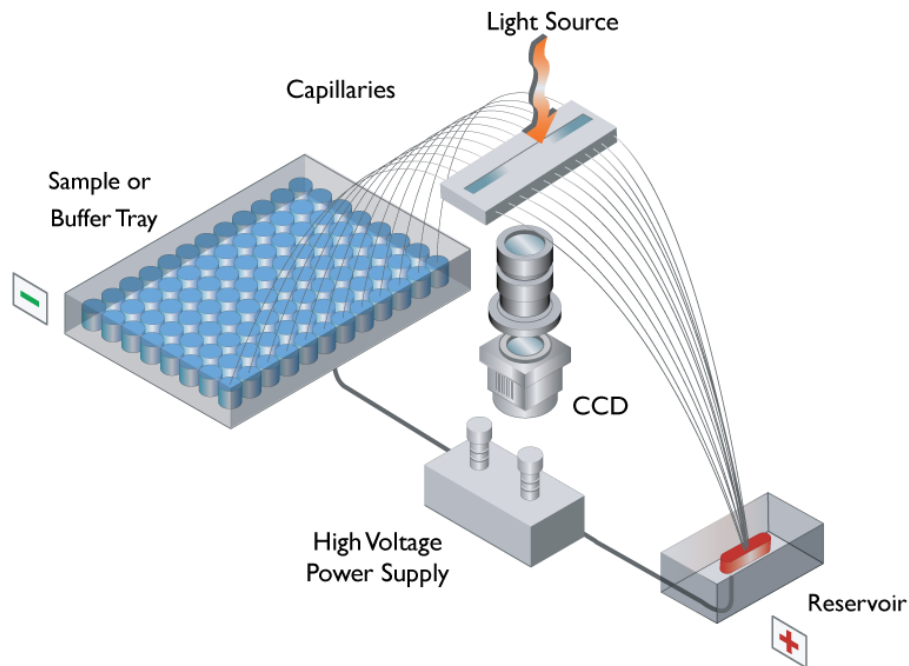
Les instruments BluePippin et SageHLS précédemment décrits pour la sélection en taille et l'extraction d'ADN de haut poids moléculaire permettent également d'évaluer la distribution en taille d'un échantillon. Le BluePippin permet surtout d'évaluer la part de fragments supérieurs à 50 kb environ et n'est donc que moyennement adapté à une analyse précise. La sensibilité et les temps d'analyse de ces instruments ne sont pas meilleurs que ceux du PFGE classique et leur utilisation pour du contrôle qualité reste marginale.

### 2.3.3.2 *L'électrophorèse capillaire*

L'inconvénient majeur des gels d'agarose pour le contrôle qualité d'échantillons est leur faible débit d'analyse. Avec l'introduction du séquençage de Sanger dans les années 70, l'évaluation précise de la taille d'un échantillon d'ADN est devenue une étape essentielle. Cette situation a conduit au développement de l'électrophorèse capillaire, plus performante. Fonctionnant selon le même principe physique, la séparation en format capillaire a l'avantage de limiter l'effet Joule grâce à une meilleure dissipation de la chaleur. En effet, les capillaires les plus utilisés ont un diamètre interne entre 25 et 150  $\mu\text{m}$  et possèdent donc un rapport surface/volume élevé. Cela permet l'application de forts champs électriques qui améliorent la vitesse et la résolution des séparations. L'électrophorèse capillaire possède les mêmes limitations que le gel d'agarose en termes de taille d'ADN qu'il est possible de séparer. L'utilisation de champs pulsé (FIGE pour field inversion gel electrophoresis) a également permis d'augmenter ce seuil mais jusqu'à 200 kb seulement<sup>150</sup>. Les travaux de Morris et al. dans les années 1990 rapportent des séparations jusqu'à 1,5 Mb en des temps records de 5 min mais n'ont jamais pu être reproduits<sup>151</sup>. De plus, des inversions ou duplications de bandes ont été reportées lors de séparations d'ADN de haut poids moléculaire en FIGE<sup>152</sup>.

D'un point de vue commercial la société Advanced Analytical Technologies Inc (AATI) propose l'instrument de FIGE le plus performant du marché, le FemtoPulse. Il permet la séparation d'ADN jusqu'à 165 kb en 3 h environ avec une sensibilité annoncée de 300 fg/ $\mu\text{L}$  sur des grands fragments uniques (>10 kb) et 3 pg/ $\mu\text{l}$  pour un smear d'ADNg. La fidélité de mesure est de 20 % CV sur la taille et de 25% CV sur la concentration. La justesse de la mesure en taille est meilleure, de 5% alors que la justesse de la mesure de concentration n'est pas précisée. Comme pour l'électrophorèse en gel, la taille et la concentration sont évaluées en comparant le temps de migration et l'intensité du signal de l'échantillon avec celui d'une référence calibrée. Un des principaux avantages de cet instrument est son haut débit, permettant l'analyse de douze échantillons en parallèle et le chargement de 288 échantillons (Figure 15). Cet instrument, commercialisé depuis fin 2016, répond donc bien à la demande de la génomique de 3<sup>ème</sup> génération en termes de contrôle qualité. Le système reste

cependant assez onéreux, 80000 € pour l'instrument et 2750 € par faisceau de capillaires, auxquels s'ajoutent les consommables en kit.



**Figure 15 : Schéma du système FEMTO Pulse d'AATI fonctionnant avec un faisceau de 12 capillaires et des plaques 96 puits. L'ADN marqué en fluorescence est séparé par électrophorèse et le signal est recueilli par une caméra CCD après excitation. Copyright Advanced Analytical.**

### 2.3.3.3 Séparation d'ADN en systèmes microfluidiques

Dans les années 2000, le développement des technologies microfluidiques<sup>153</sup> a permis l'apparition de nouveaux systèmes électrophorétiques performants en termes de sensibilité, consommant peu d'échantillon et capables de délivrer un résultat en quelques minutes. La dissipation de chaleur est encore meilleure qu'en capillaire et les distances de séparation plus courtes, ce qui permet de réduire le temps d'analyse.

De nombreux systèmes ont ainsi été développés commercialement, comme le Bioanalyzer et le TapeStation (Agilent), le Experion (BioRad) ou le LabChip GX Touch (Perkin Elmer). Cependant ces instruments ne possèdent pas de fonction champ pulsé et conservent les mêmes limitations en termes de séparation d'ADN de haut poids moléculaire que les gels d'agarose classiques.

D'un point de vue académique, des puces microfluidiques intégrant de nouvelles fonctions de séparation sont développées. En effet, les techniques de fabrication des puces offrent plus de versatilité que les capillaires et permettent le design de réseaux fluidiques sophistiqués, l'introduction de motifs dans les canaux et l'apparition de nouveaux modes de séparation. Parmi les nombreuses

méthodes développées, nous détaillerons brièvement les techniques de micro pylônes, de prisme à ADN et de piège entropique pour la séparation de longues molécules. Un panorama détaillé de ces techniques est disponible dans la revue de Dorfman à ce sujet<sup>154</sup>.

En 2002 Doyle et al. ont montré la possibilité de séparer de longs fragments d'ADN jusqu'à 150 kb en une dizaine de minute à travers un réseau de pylônes dans lequel migre l'ADN via l'application d'un champ électrique<sup>155</sup>. Lorsque la molécule d'ADN entre en contact avec un pylône, elle se déplie en formant un U autour de celui-ci, une branche de la molécule de part et d'autre du pylône. Une branche finit par reprendre la migration, débobinant l'autre branche du pylône (Figure 16, A). La vitesse à laquelle un fragment d'ADN s'extirpe du pylône est dépendante de la taille et permet ainsi la séparation. Des résolutions de 1 ont pu être obtenues entre une molécule de 48,5 kb et une de 166 kb avec des champs continus, ce qui reste assez limité. Comme pour l'électrophorèse en gel d'agarose, l'utilisation d'un champ pulsé permet d'améliorer les performances de la technique. Ainsi, une séparation a été faite entre un fragment de 48,5 kb et un de 166 kb en 10 s<sup>156</sup>. Cependant, aucune séparation de mélange complexe utilisant des réseaux de micro pylônes n'a été rapportée à ce jour<sup>157</sup>.

Une autre technique de microfluidique utilisant du champ pulsé est le prisme à ADN<sup>158</sup>. Le principe de séparation est similaire à celui développé dans les réseaux de pylônes mais avec un champ pulsé asymétrique, dans deux directions et de deux intensités différentes (Figure 16, C). L'ADN migre initialement vers le plus fort champ électrique jusqu'au basculement vers le champ électrique de faible intensité. Du fait de la faible intensité du champ, les plus grandes molécules n'ont pas le temps de se réorienter, ainsi leur migration moyenne est celle du fort champ électrique. Les petites molécules ont le temps de s'orienter dans le champ de faible intensité. Ainsi le degré de migration des petites molécules dépend de leur taille et permet la séparation. Un mélange de quatre tailles de 61, 114, 158 et 209 kb a pu être séparé par cette méthode en 15 s avec des résolutions proches de 1 entre les fragments<sup>158</sup>. La méthode n'est en revanche pas très sensible et nécessite un couplage avec des méthodes de concentration afin d'obtenir des performances suffisantes pour du contrôle qualité d'ADN de haut poids moléculaire. Cette méthode permet également de fractionner l'ADN en continu en collectant les différentes tailles dans des canaux après la séparation<sup>159</sup>.

Le piège entropique appliqué à la séparation de longues molécules d'ADN a été développé en 2000 par Han et Craighead. La méthode consiste à faire migrer l'ADN par électrophorèse à travers une alternance de zones larges et de zones fines par rapport à la taille des molécules à séparer (Figure 16, B). La profondeur des régions fines est plus faible que le rayon de giration de l'ADN à séparer, l'obligeant à se déformer, alors que dans les régions larges l'ADN retrouve sa forme de pelote aléatoire. Cette déformation dans les zones peu profondes entraîne une barrière entropique qui piège



momentanément les molécules à l'entrée de la région fine. Cette barrière entropique limite la mobilité de l'ADN qui devient dépendante de la taille, les grandes molécules migrant plus rapidement que les petites. En effet, plus une molécule est grosse et plus la probabilité est élevée qu'elle s'approche et entre dans la région peu profonde de fort champ électrique<sup>35</sup>. Han et Craighead ont montré la possibilité de séparer deux bandes de 164 et 38 kb en 15 min environ ou un mélange plus complexe de 10 à 50 kb en 25 min.

Globalement, les performances de ces instruments sont un peu limitées pour effectuer un contrôle qualité des échantillons d'ADN de haut poids moléculaire avant séquençage. Elles n'ont d'ailleurs jamais été industrialisées.

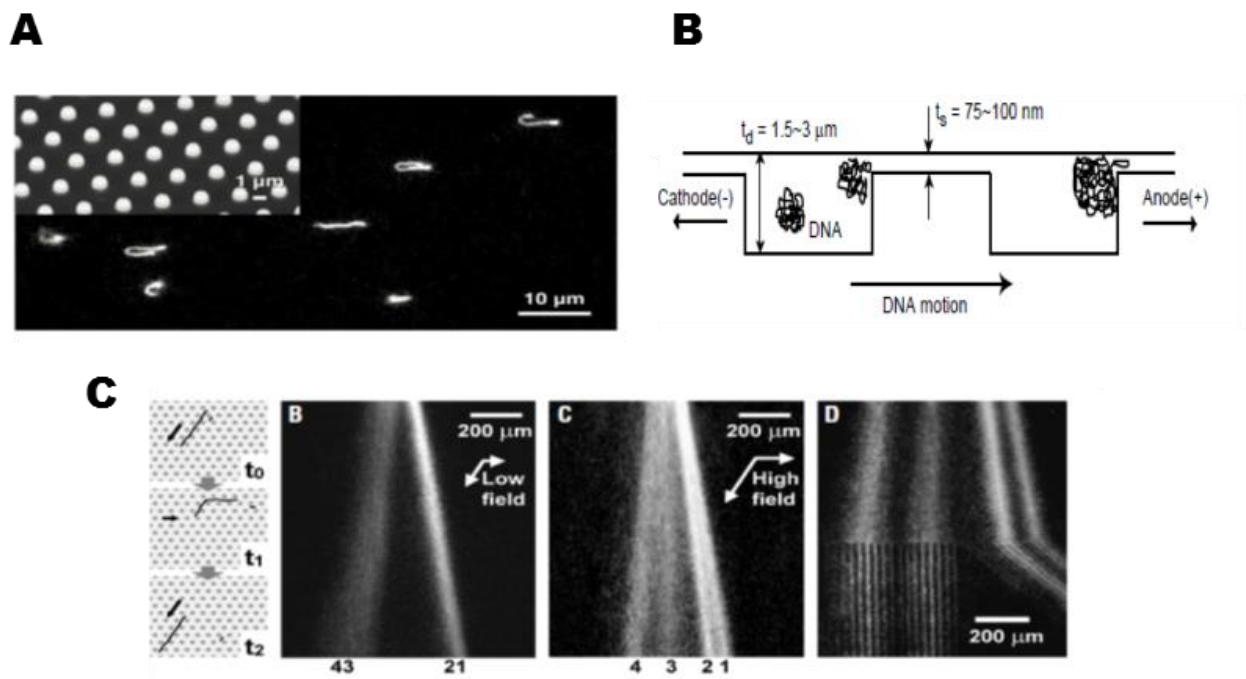


Figure 16 : Méthodes de séparations microfluidique d'ADN de haut poids moléculaire. A : Séparation en réseau de micro pylône, Doyle et al.<sup>155</sup>. B : Piège entropique, tiré de Han et al.<sup>35</sup>. C : Prisme à ADN, tiré de Dorfman et al.<sup>157</sup>

#### 2.3.3.4 Méthodes chromatographiques

Une des méthodes les plus classiques pour séparer des biomolécules en fonction de leur taille est la chromatographie d'exclusion stérique. Cette technique se base sur l'interaction plus ou moins forte des molécules à séparer avec une phase stationnaire poreuse. Les petites molécules pénètrent dans les pores de la matrice alors que les plus grandes molécules vont être exclues et éluées rapidement. Ces techniques, très utilisées pour la purification et l'analyse de protéines, sont en revanche peu utilisées pour la purification d'ADN à cause de la faible gamme de taille qu'il est possible de séparer. En effet, les colonnes de chromatographie d'exclusion stérique spécialement conçues pour

les grandes molécules ne permettent pas de séparer des fragments de plus de 5000 kDa, soit 8 kb environ. La matrice la plus couramment utilisée dans ce cadre est la Superose<sup>®</sup>, qui est une forme d'agarose réticulée et perlée<sup>160</sup>. Le débit d'analyse est assez faible avec les colonnes chromatographiques qui peuvent également se boucher.

## 2.4 Conclusion du Chapitre 2

Dans ce deuxième chapitre, nous avons vu que l'obtention de données de séquençage de qualité dépend grandement des étapes de préparation de l'échantillon. Cette préparation comprend une étape d'extraction de l'ADN, visant à récupérer des quantités suffisantes d'ADN purifié de haut poids moléculaire, de taille typique supérieure à 50 kb. Une sélection en taille de l'échantillon d'ADN est souvent nécessaire lorsque l'extraction conduit à de la fragmentation de l'ADN génomique. Enfin, une étape de contrôle qualité vérifie que l'échantillon correspond aux spécifications du séquenceur.

Le Tableau 3 récapitule les principales techniques d'extraction d'ADN de haut poids moléculaire ainsi que les caractéristiques de l'échantillon obtenu. On observe ainsi que beaucoup de techniques sont disponibles pour effectuer des extractions, aboutissant à différentes tailles d'ADN selon les besoins. La plus grande qualité d'ADN est obtenue avec l'utilisation de matrices d'agarose au prix de temps d'extraction allongés. Une automatisation de l'extraction en agarose est proposée par SageScience dans le but de réduire le temps de manipulation mais le rendement d'extraction est malheureusement faible. Les récents kits de billes magnétiques, comme le Magattract de Qiagen, sont un bon compromis à l'obtention d'ADN de qualité pour des temps et coûts d'extraction plus raisonnables. Ils ne sont cependant pas compatibles avec tous les organismes et la qualité du résultat est variable d'un utilisateur à un autre.

**Tableau 3 : Présentation des différentes méthodes d'extraction d'ADN de haut poids moléculaire ainsi que des caractéristiques des échantillons obtenus.**

Méthode	Billes magnétiques	Phénol-Chloroforme	Colonne gravitaire échangeuse d'anion	Plug d'agarose	Gel d'agarose automatisé
Fournisseur	Qiagen - MagAttract HMW DNA	N.A.	Qiagen - Genomic tip	N.A.	SageScience - SageHLS
Quantité	~10 µg	Selon besoin	10-500 µg	Selon besoin	5 µg
Rendement	80%	70%	80%	Variables	30%
Taille ADN	50-150 kb	100-200 kb	50-120 kb	> Mb	Mb
Temps	70 min	4 h	5 h	3 j	1 j
Coût	Moyen	Faible	Moyen	Faible	Élevé
Avantages	Simplicité, rendements, grands fragments, automatisable	Grands fragments, peu coûteux	Simplicité, rendements	Très grands fragments, qualité, peu coûteux	Très grands fragments, automatisation
Inconvénients	Variabilité, compatibilité	Solvant toxiques, expérience	Temps, difficilement automatisable	Laborieux, long, collecte ADN	Rendements, coût, temps

Malgré les précautions prises lors des étapes d'extraction, la préparation des échantillons pour le séquençage de 3<sup>ème</sup> génération fait souvent intervenir une étape de sélection en taille pour éliminer les fragments de dégradation. Contrairement aux méthodes d'extraction, peu d'instruments sont disponibles pour une sélection en taille rapide et efficace de l'ADN de haut poids moléculaire. Le Tableau 4 récapitule les différentes méthodes disponibles. Les billes magnétiques sont très utilisées pour la sélection en taille dans les protocoles NGS mais ne fonctionnent que sur des petits fragments d'ADN. Pour l'ADN de haut poids moléculaire, la seule méthode disponible est le gel d'agarose avec la possibilité d'une automatisation. Ainsi, l'instrument de référence pour effectuer cette sélection est le BluePippin malgré des rendements faibles sur les longs fragments d'ADN et des temps de préparation assez longs. Cette synthèse démontre donc qu'il existe un besoin technologique pour l'isolation d'ADN de haut poids moléculaire.

**Tableau 4 : Présentation des différentes méthodes de sélection en taille de l'ADN ainsi que leurs performances.**

Méthode	Billes magnétiques	BluePippin	Gel d'agarose	Nimbus
Plage sélection	100-1000 bp	Jusqu'à 50 kb	Jusqu'à 5 Mb	Jusqu'à 20 kb
Justesse	N.A.	10%	10%	10%
Quantité	0,5 µg/µL billes (MagAttract)	0,5 – 10 µg/canal	0,5 – 10 µg/ligne	0,5 µg/canal
Volumes initial	> 10 µL	30 µL	~30 µL	50 µL
Volume collecte	> 5 µL	120 µL	> 30 µL	400 µL
Distribution en taille (CV)	30%	20%	20%	10%
Rendement	80%	20-50%	~50%	75%
Temps	1 h	1 j	1-2 j	6 h
Cout	Moyen	Élevé	Faible	Très élevé
Avantages	Simple, automatisable	Grands fragments, automatisation	Accessible, très grands fragments	Automatisation, haut débit
Inconvénients	Petits fragments, fidélité	Rendements, coût, abime l'ADN	Laborieux, fidélité, contamination	Limité en taille, coût

Une fois l'échantillon extrait et sélectionné en taille, il est nécessaire de le caractériser et de contrôler sa qualité. La mesure de pureté est essentiellement réalisée grâce à la spectrophotométrie d'absorbance malgré sa faible sensibilité. Pour une mesure de concentration précise, la fluorimétrie avec la technique Qubit est communément adoptée bien qu'elle fasse appel à un étalon. Enfin

l'évaluation de la distribution en taille est effectuée en gel d'agarose en champs pulsés lorsqu'il est nécessaire d'analyser de l'ADN supérieur à 200 kb. Pour des fragments de taille inférieure (10-200 kb), l'instrument de référence est le FEMTO-Pulse permettant de séparer de l'ADN jusqu'à 160 kb avec une excellente sensibilité et une cadence d'analyse adaptée aux plateformes haut-débit (Tableau 5). Les technologies commerciales utilisant la microfluidique restent limitées à des séparations <50kb.

**Tableau 5 : Présentation des différentes méthodes de contrôle qualité en se focalisant sur l'analyse de l'ADN de haut poids moléculaire.**

Méthode	Spectrophotométrie	Fluorimétrie	PFGE	Electrophorèse capillaire (FemtoPulse)	Microfluidique (académique)
Mesure de pureté	Oui. A260/A280 et A260/A230	Non	Non	Non	Non
Mesure de concentration	Oui à 5%	Oui à 5%	Oui. Semi-quantitative	Oui à 25% CV	Non
Limite de détection	0,4 ng/μL	10 pg/μL	10 ng/μL en smear	300 fg/μL et 3 pg/μL en bande unique et smear	~1 ng/μL sans préconcentration
Taille d'ADN analysable	N.A.	N.A.	0,1 - 10 Mb	10 - 200000 bp	61 - 209 kb
Temps d'analyse	5 min	30 min	~24 h	3 h	15 s pour le prisme à ADN
Coût analyse	Faible	Faible	Moyen	Elevé	N.A.

### **3. $\mu$ LAS pour la manipulation d'ADN de haut poids moléculaire**

---

La technologie  $\mu$ LAS permet d'effectuer simultanément les opérations de concentration et de séparation de l'ADN en fonction de sa taille. Cette méthode d'analyse de l'ADN a été découverte durant la thèse d'Hubert Ranchon<sup>161</sup> en 2013 et a fait l'objet de deux brevets<sup>7,8</sup>. La technologie a ensuite été transférée industriellement vers Picometrics Technologies afin de développer un instrument commercialisable. Les premiers résultats significatifs ont été obtenus pour la séparation d'ADN de bas poids moléculaire, de 100 à 1500 bp. La concentration simultanée permet d'atteindre des limites de détection inférieures à l'état de l'art, de l'ordre de 10 fg/ $\mu$ L<sup>10</sup>. Cela a permis le développement d'une application centrée sur l'analyse de l'ADN circulant qui a également fait l'objet de la thèse de Rémi Malbec<sup>9</sup>. Nous décrirons au cours de ce chapitre le fonctionnement de cette nouvelle méthode de séparation ainsi que les éléments qui nous ont permis de manipuler des fragments d'ADN de taille supérieure. Nous aborderons les performances analytiques de cette technologie afin de les comparer aux techniques actuellement disponibles sur le marché et leur potentiel applicatif pour la génomique de 3<sup>ème</sup> génération.

## 3.1 Principe de fonctionnement de $\mu$ LAS

### 3.1.1 La séparation sans matrice solide

Contrairement aux techniques de séparation d'ADN décrites dans le chapitre 2 (2.3.3), la séparation  $\mu$ LAS n'utilise pas de matrice solide. L'utilisation de matrices a longtemps été indispensable pour induire des différences de mobilité entre les fragments d'ADN malgré certains inconvénients. En effet, l'utilisation de gels d'agarose est coûteuse en temps et en consommables. Les matrices utilisées en électrophorèse capillaire peuvent manquer de reproductibilité et boucher les capillaires du fait de leur viscosité importante, et les matrices micro fabriquées décrite en 2.3.3.3 sont complexes à fabriquer. Ainsi, certaines équipes ont travaillé au développement de nouvelles matrices plus fonctionnelles pour la séparation d'ADN. On peut citer les travaux de Barron et al. visant à utiliser des solutions de polymères dilués, moins visqueuses, qui vont faciliter leur injection dans les capillaires et accélérer les séparations<sup>162</sup>. Elles n'ont cependant pas permis d'effectuer des séparations au-delà de 23 kb.

En 2009, Wang et al. ont introduit la chromatographie hydrodynamique en solution libre pour la séparation de fragments d'ADN de 75 bp à 106 kb (Figure 17, B)<sup>163</sup>. La méthode repose sur l'encombrement stérique des grandes molécules qui sont repoussées des parois dans un micro capillaire. Lors de l'application d'un écoulement laminaire, les grandes molécules sont soumises à des vitesses d'écoulement plus importantes loin de la paroi et sont éluées en premier (Figure 17, A). Les séparations sont typiquement effectuées dans des capillaires de diamètre  $\sim 1 \mu\text{m}$ . Elles sont assez longues,  $\sim 2 \text{ h}$ , et nécessitent l'application de fortes pressions,  $\sim 20 \text{ Bar}$ . Le faible diamètre des capillaires les rend très sensibles au colmatage.

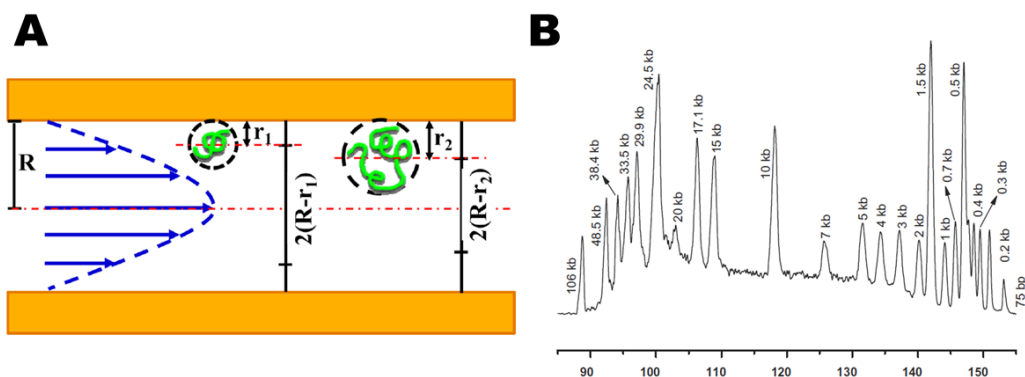


Figure 17 : A : Principe de séparation par chromatographie hydrodynamique dans un micro capillaire à l'aide d'un écoulement de Poiseuille. Les grandes molécules, plus encombrantes, migrent plus loin des parois et sont ainsi soumises à des vitesses plus importantes que les petites molécules. B : Séparation dans un capillaire de  $5 \mu\text{m}$  de diamètre sur une longueur effective de 14,95 m à 24 Bar d'un mélange de fragments d'ADN de 106 kb à 75 bp. Tiré de Wang et al.<sup>163</sup>.

Au début des années 2000, Yeung et Zheng ont montré la possibilité de séparer des fragments d'ADN dans des capillaires de diamètre supérieur, moins sensibles au colmatage, grâce au couplage d'une électrophorèse et d'un flux hydrodynamique<sup>164</sup>. Ils ont remarqué que la combinaison de ces deux forces entraînait la migration de l'ADN vers le centre ou les parois de la canalisation si elles sont appliquées conjointement ou en opposition respectivement. La vitesse de migration transverse est dépendante de la taille des molécules et induit donc des différences de mobilité. Cependant, une fois en régime permanent, la position de migration des molécules est indépendante de leur taille. La séparation est alors obtenue en alternant le champ hydrodynamique afin de maintenir les différences de mobilité. Ils ont ainsi montré la séparation de deux fragments de 48 et 5 kb mais les résolutions sont faibles et le protocole trop laborieux pour être utilisé en routine.

Bien que les migrations transverses décrites par Yeung et Zheng soient similaires à celles observées lors des séparations  $\mu$ LAS (voir 3.1.3), elles sont effectuées dans un fluide newtonien, sans propriétés viscoélastiques essentielles au phénomène  $\mu$ LAS. Ils utilisent cependant un coating de polyacrylamide linéaire pour supprimer l'électro-osmose dans leurs capillaires. On peut ainsi se demander si ce coating est capable d'introduire une viscoélasticité résiduelle dans leur solution.

### 3.1.2 Migration transverse viscoélastique

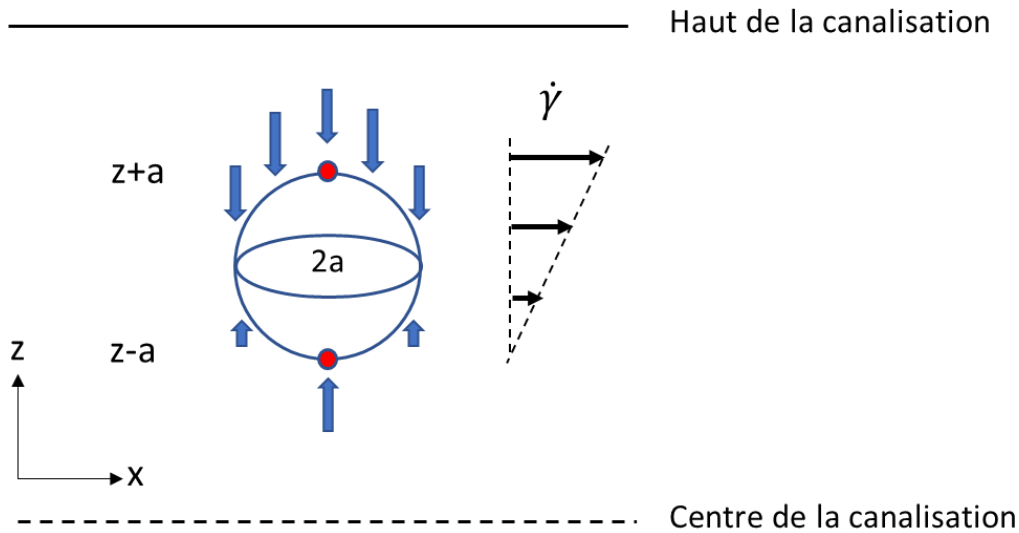
Contrairement aux migrations décrites ci-dessus par Wang et Yeung, des propriétés viscoélastiques du tampon sont nécessaires pour obtenir des séparations  $\mu$ LAS d'ADN de bas poids moléculaire<sup>6,10</sup>. Ces fluides sont caractérisés par un comportement intermédiaire entre un fluide visqueux et un solide élastique. Ils sont généralement obtenus par dissolution de polymères au-delà d'une concentration seuil,  $c^*$ , appelée concentration de recouvrement<sup>165</sup>. Elle est définie par :

$$c^* = \frac{3M_w}{4\pi R_g^3 N_A} \quad (3)$$

Avec  $M_w$  la masse molaire du polymère,  $R_g$  le rayon de giration du polymère et  $N_A$  le nombre d'Avogadro. Elle correspond à la concentration à partir de laquelle les pelotes formées par le polymère en solution commencent à interagir et former un réseau, et dépend donc de la nature du polymère ainsi que de la longueur de ses chaînes. Pour des solutions de PVP 360 kDa et 40 kDa on trouve un  $c^*$  de 7 g/L et de 58 g/L respectivement<sup>166</sup>.

La migration transverse de particules dans des fluides viscoélastiques est un phénomène bien décrit. La Figure 18 ci-dessous nous permet de schématiser la migration de particules dans une canalisation lors d'un écoulement laminaire.





**Figure 18 : Représentation schématique de l'écoulement viscoélastique d'une bille de rayon  $a$  dans une canalisation. Une différence de pression est générée de part et d'autre de la bille par le cisaillement.**

Leshansky et al. ont ainsi montré en 2007 la focalisation de billes de  $8 \mu\text{m}$  au centre d'une canalisation de  $20 \mu\text{m}$  sous l'application d'un écoulement dans une solution viscoélastique de polyvinylpyrrolidone 360 kDa à 8%<sup>167</sup>. Cette migration transverse est dépendante de la vitesse de l'écoulement, de la taille de la particule, et de la rhéologie de la solution ; cet effet n'est pas observé dans une solution purement visqueuse de glycérol. Ils ont ainsi déduit une expression de cette force transverse élastique :  $F_e \sim a^3 \left( \frac{\partial N_1}{\partial z} \right)$  avec  $a$  la taille de la particule et  $N_1$  la différence de contrainte normal pour un fluide de Maxwell :

$$N_1 = K\eta\tau\dot{\gamma}^2 \quad (4)$$

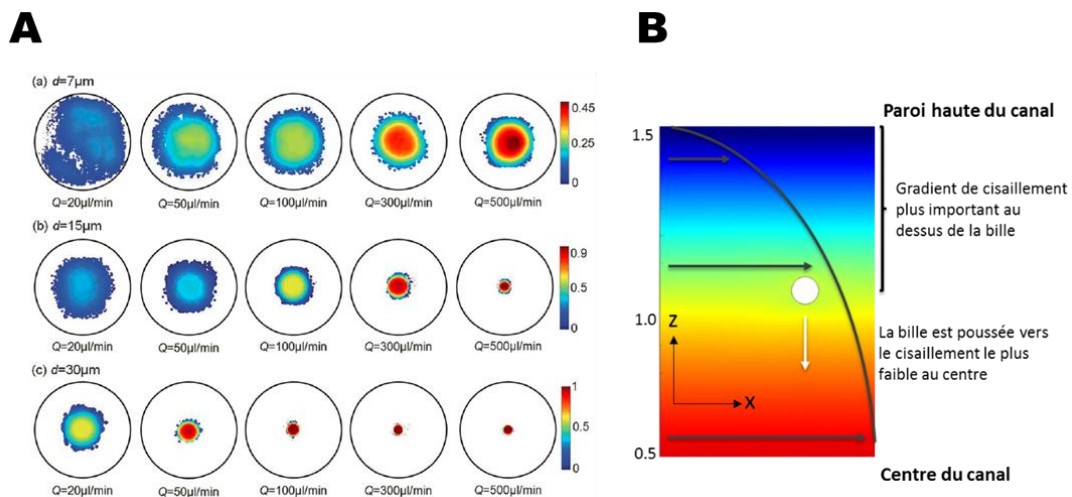
Avec  $K$  une constante de Maxwell,  $\eta$  la viscosité de fluide,  $\tau$  le temps de relaxation du fluide et  $\dot{\gamma}$  le cisaillement. On obtient ainsi l'expression de la force transverse suivante pour une position  $\varepsilon$  dans la canalisation :

$$F_T(\varepsilon) = -2\pi K \eta \tau a^3 \frac{\partial \dot{\gamma}^2}{\partial z}(\varepsilon) \quad (5)$$

Plus récemment, Seo et al. ont également montré l'augmentation de la force transverse avec la taille des particules et la vitesse d'écoulement pour la focalisation au centre d'une canalisation (Figure 19, A)<sup>168</sup>. Kim et al. ont ensuite réussi à focaliser des molécules d'ADN de phage  $\lambda$  (48,5 kb) et T4 (169 kb) au centre d'une canalisation grâce à ces propriétés d'écoulement viscoélastique, et ils ont montré qu'elle dépendait de la taille des molécules d'ADN<sup>169</sup>. Kang et al. ont directement utilisé l'ADN

afin de conférer au fluide des propriétés viscoélastiques. Ils ont ainsi focalisé des particules au centre d'une canalisation dans une solution d'ADN  $\lambda$  à 5 mg/L. Il est intéressant de noter que la concentration  $c^*$  de l'ADN  $\lambda$  est de 80 mg/L et que les focalisations décrites par Kang et al. sont faites en régime dilué. Les auteurs suspectent que l'ADN augmente fortement les propriétés élastiques de la solution grâce à sa grande taille et son temps de relaxation très long.

Cette focalisation au centre de la canalisation provient d'un gradient de cisaillement qui pousse la particule vers les zones de cisaillement les plus faibles, au centre du canal (Figure 19, B). Pour une description plus détaillée de ces mécanismes de migration transverse en fluides viscoélastique il est possible de se référer aux revues de d'Avino et al.<sup>170,171</sup>. Nous retiendrons qu'en présence de fluides viscoélastiques une force transverse à l'écoulement dirigée vers le centre de la canalisation apparaît, et dont l'intensité dépend de la rhéologie de la solution, de la taille des molécules, et de la vitesse d'écoulement.



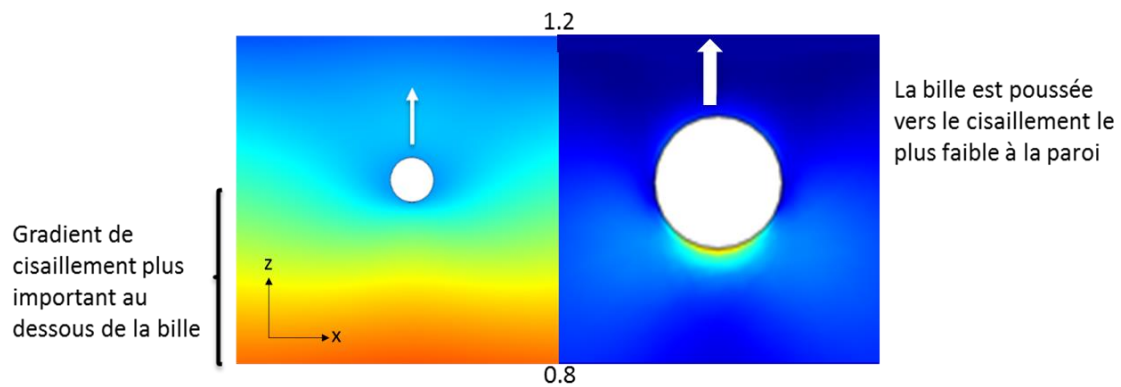
**Figure 19 : A : Variation de la concentration de particules de diamètre (a)  $d=7 \mu\text{m}$ , (b)  $d=15 \mu\text{m}$ , (c)  $d=30 \mu\text{m}$  dans une canalisation de  $d=300 \mu\text{m}$  dans une solution viscoélastique de 8% PVP 360 kDa à différentes vitesses d'écoulement selon l'axe du canal. Tiré de Seo et al.<sup>168</sup>. B : Simulation COMSOL réalisée par Vincent Picot, post-doctorant du LAAS, d'un écoulement de Poiseuille autour d'une bille avec un fluide viscoélastique.**

### 3.1.3 La séparation par $\mu$ LAS

Le principe de séparation  $\mu$ LAS repose sur l'application bidirectionnelle d'un champ électrique et d'une force hydrodynamique dans un fluide viscoélastique. Cette technique est mise en œuvre en format capillaire à Picometrics et en format puce microfluidique au LAAS. Dans les deux configurations, l'écoulement produit par la force hydrodynamique est de type Poiseuille, caractérisé par une vitesse maximale au centre de l'écoulement et qui décroît à mesure que l'on s'approche des parois du canal.

L'électrophorèse, en opposition de cet écoulement, freine les molécules d'ADN mais leur transport reste dominé par l'écoulement hydrodynamique.

Comme nous l'avons vu précédemment, le caractère viscoélastique de la solution introduit une troisième force, perpendiculaire à l'écoulement. Contrairement aux migrations décrites en 3.1.2, la migration transverse est orientée vers les parois du capillaire. En effet, l'électrophorèse en opposition modifie les zones de cisaillement autour de la bille. Ce fort cisaillement induit une force transverse, opposée au cas sans électrophorèse, qui entraîne la migration de la molécule vers les parois du canal (Figure 20).



**Figure 20 : Simulation COMSOL réalisée par Vincent Picot du champ de vitesse du fluide (panneau de gauche) et du champ de cisaillement (panneau de droite) autour d'une bille lors de l'application bidirectionnelle d'un flux hydrodynamique et d'une électrophorèse dans un fluide viscoélastique.**

L'origine de cette force transverse a été décrite et modélisée par l'équipe, et plus particulièrement la doctorante Bayan Chami <sup>172</sup>. Pour obtenir l'expression de cette force, ils ont cherché à exprimer la vitesse de migration d'une bille soumise à une contre-électrophorèse dans un écoulement de Poiseuille. Celle-ci s'écrit  $V_p(\varepsilon) - V_e$  avec  $V_p$  la vitesse du fluide à la position  $\varepsilon$  dans le canal et  $V_e$  la vitesse électrophorétique. On peut alors exprimer le taux de cisaillement au-dessus de la particule :

$$\dot{\gamma}_u \approx \frac{1}{\Delta z} \{V_p(\varepsilon + a + \Delta z) - [V_p(\varepsilon + a) - V_e]\} = \frac{dV_p}{dz}(\varepsilon + a) + \frac{\Delta z}{2} \frac{d^2V_p}{dz^2}(\varepsilon + a) + \frac{V_e}{\Delta z} \quad (6)$$

Avec  $\Delta z$  une distance caractéristique sur laquelle l'écoulement est déformé par la bille. Autrement dit, on fait l'hypothèse que la vitesse  $V_p$  est celle de l'écoulement Poiseuille sans bille au-delà de cette distance. On peut alors calculer la différence du carré du taux de cisaillement entre le dessus et le dessous de la particule :

$$\dot{\gamma}_u^2 - \dot{\gamma}_b^2 \approx \frac{2}{\Delta z} \frac{dV_p}{dz}(\varepsilon) \left[ 2V_e + \Delta z^2 \frac{d^2V_p}{dz^2}(\varepsilon) \right] \quad (7)$$

En supposant que la longueur caractéristique  $\Delta z$  soit du même ordre que  $a$ , on peut en déduire une expression de la force transverse :

$$F_T^{el}(\varepsilon) = -4\pi K \eta \tau a \frac{dV_p}{dz}(\varepsilon) \left[ 2V_e + a^2 \frac{d^2V_p}{dz^2}(\varepsilon) \right] \quad (8)$$

Avec  $K$  une constante de Maxwell,  $\eta$  la viscosité,  $\tau$  le temps de relaxation élastique du fluide,  $a$  le rayon de la particule,  $h$  le diamètre ou hauteur de la canalisation. En intégrant l'expression du champ de vitesse parabolique  $V_p(z) = 4V_{p_0} \frac{z}{h} \left(1 - \frac{z}{h}\right)$  dans (8) on obtient :

$$F_T^{el}(\varepsilon) = -32\pi K \eta \tau \frac{a}{h} V_{p_0} \left( V_e - 4V_{p_0} \frac{a^2}{h^2} \right) \left( 1 - \frac{2\varepsilon}{h} \right) \quad (9)$$

On remarque que sans contre-électrophorèse, *i.e.* avec  $V_e=0$ , on retombe sur l'expression (5) de la force transverse décrite par Leshansky et al.<sup>167</sup>. Si la taille de la particule  $a$  est faible par rapports aux dimensions de la canalisation  $h$ , et que les vitesses électrophorétique et hydrodynamique sont de même ordre, alors l'expression (9) se simplifie en :

$$F_T^{el}(\varepsilon) = -32\pi K \eta \tau \frac{a}{h} V_{max} V_e \left( 1 - \frac{2\varepsilon}{h} \right) \quad (10)$$

Il est intéressant de noter qu'un travail de modélisation de Butler et al. (ref.<sup>175</sup>) sur le transport de particules en forme d'altères dans un écoulement purement visqueux avec une contre électrophorèse a montré un comportement physique comparable avec une force transverse visqueuse  $F_T^v$  qualitativement écrite selon :

$$F_T^v(\varepsilon) \sim -NV_{p_0} V_e \left( 1 - \frac{2\varepsilon}{h} \right) \quad (11)$$

Cette force n'a pas été étudiée précisément dans notre système de séparation. Une étude pour distinguer l'amplitude relative de  $F_T^{el}$  et  $F_T^v$  reste donc à mener. Les valeurs types des différents facteurs de l'équation (10) retrouvées dans nos expériences sont regroupées dans le Tableau 6 :

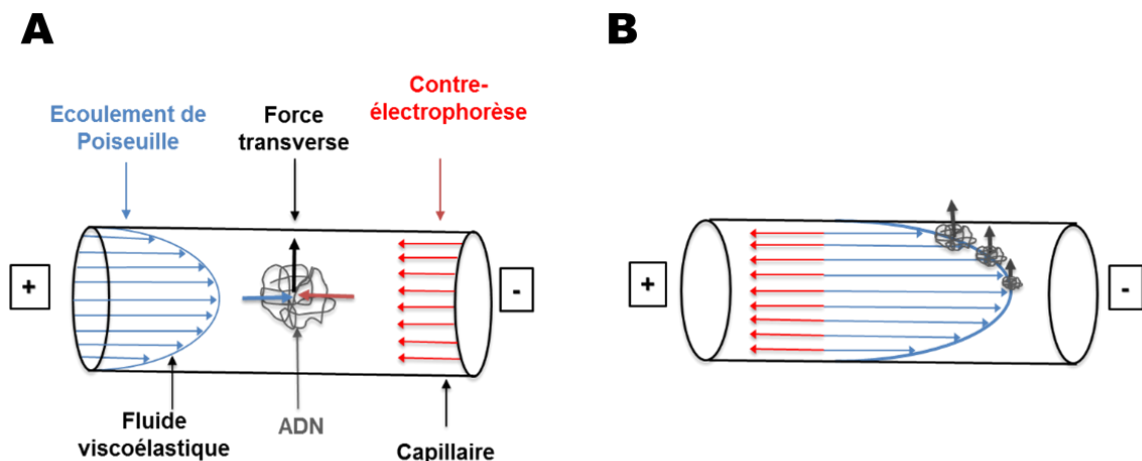
**Tableau 6 : Valeurs types des différentes constantes et variables de l'expression de la force transverse viscoélastique.**

K	$\eta$ (mPa.s)	$\tau$ (ms)	$a$ ( $\mu$ m)	$h$ ( $\mu$ m)	$V_{max}$ (mm/s)	$V_e$ ( $\mu$ m/s)
0.1	1 à 40	15	1	25 $\mu$ m	1 à 10	10 à 100

La viscosité des solutions a été mesurée par électrophorèse capillaire à partir du temps de migration d'un « plug » de fluorescéine dans un capillaire de dimensions connues à une pression donnée. On peut alors déterminer la vitesse d'écoulement à cette pression et donc la viscosité de la solution. Celle-ci s'est avérée constante sur la gamme de taux de cisaillement à laquelle nous travaillons ( $50-1000 \text{ s}^{-1}$ ), indiquant un comportement de fluide newtonien.

La viscoélasticité des solutions de PVP a également été évaluée par micro rhéologie et suivi de particules. En effet, grâce au modèle de Maxwell, il est possible de relier le déplacement quadratique moyen de traceurs fluorescent à la viscosité d'une solution aux temps longs et à son élasticité aux temps courts<sup>172</sup>.

L'équation (10) nous indique donc que l'application conjointe d'une force hydrodynamique et d'une électrophorèse en opposition entraîne l'apparition d'une force transverse orientée vers les parois du capillaire s'exerçant sur les molécules d'ADN (Figure 21, A). Elle dépend des caractéristiques du fluide (viscosité, élasticité), des caractéristiques de la canalisation (hauteur ou diamètre), des paramètres d'actionnement (vitesse d'écoulement et électrophorétique) mais surtout de la taille des molécules. Ainsi plus une molécule est grande ( $a$ ) et plus la force transverse est importante. Les molécules sont alors focalisées en différentes position ( $\epsilon$ ) de la canalisation en fonction de leur taille. Or, la vitesse de l'écoulement est dépendante de cette position dans le canal. Les grandes molécules sont ainsi dans des zones d'écoulement plus lentes ce qui permet la séparation en fonction de la taille (Figure 21, B).



**Figure 21 : Principe de séparation par  $\mu$ LAS en format capillaire. A : Les molécules d'ADN migrent dans le capillaire grâce à un écoulement de Poiseuille. L'application d'une contre-électrophorèse induit une force transverse à l'écoulement, dirigée vers les parois du capillaire. B : Cette force dépend de la taille des molécules qui sont plaquées en différentes position dans la canalisation. Or, la vitesse de l'écoulement dépend de la position dans la canalisation ce qui induit une séparation en fonction de la taille.**

### 3.1.4 Caractéristiques du tampon

Afin d'obtenir une séparation, les propriétés viscoélastiques du tampon sont essentielles. Pour cela nous ajoutons dans les tampons d'analyse de la polyvinylpyrrolidone (PVP), un polymère organique obtenu par polymérisation de la *N*-vinylpyrrolidone. De nombreuses références sont disponibles, de différents fournisseurs et différents poids moléculaires. Les travaux menés par Picometrics et Rémi Malbec lors de sa thèse ont montré l'importance de la concentration et de la taille des chaînes de PVP pour la séparation d'ADN de bas poids moléculaire<sup>10</sup>. Pour les expériences menées sur de petits fragments d'ADN, du PVP 360 ou 1300 kDa est utilisé à 5% dans du tampon Tris-Borate-EDTA (TBE). Cependant la dispersion en taille des chaînes de polymère n'est jamais donnée par les fabricants. Nous avons fait analyser ces polymères en chromatographie d'exclusion stérique<sup>173</sup> par le laboratoire des Interactions Moléculaires et Réactivités Chimiques et Photochimiques (IMRCP) de l'Université Paul Sabatier de Toulouse, et découvert que la distribution en taille de ces polymères commerciaux était mal maîtrisée (Tableau 7). Ainsi nous avons fait synthétiser par les IMRCP des polymères PVPs de plus faible taille, mais de bonne pureté, et surtout avec une distribution en taille mieux maîtrisée, plus proche de la monodispersité (Tableau 7). Ces PVPs ont ensuite été testés pour la séparation d'ADN de haut poids moléculaire (voir 3.2.1).

**Tableau 7 : Comparaison de sept différents PVP basée sur leur masse moléculaire moyenne en nombre ( $M_n$ ), leur masse moléculaire moyenne en poids ( $M_w$ ) et leur dispersité (rapport  $M_w/M_n$ ). Les propriétés de ces PVP ont été déterminées par chromatographie d'exclusion stérique<sup>173</sup>.**

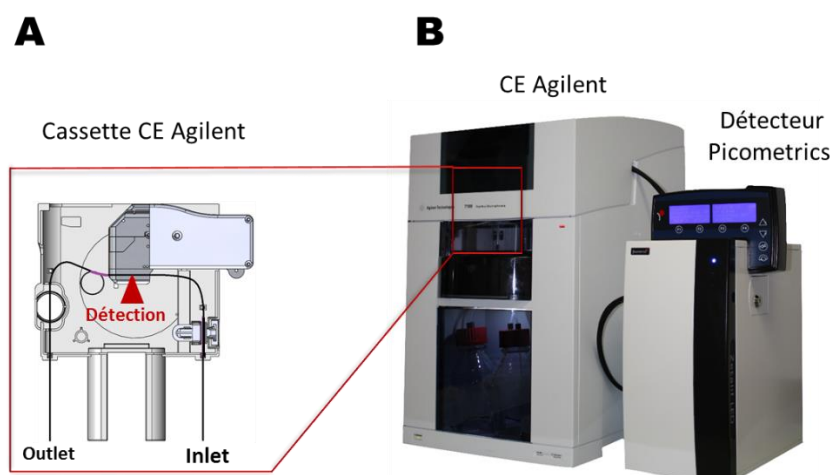
PVP	Sigma 10 kDa	IMRCP 12.5 kDa	IMRCP 30 kDa	Sigma 40 kDa	IMRCP 43 kDa	Sigma 360 kDa	Sigma 1300 kDa
$M_n$ (kDa)	9.3	12.5	30.3	26.5	43.6	427	323
$M_w$ (kDa)	16.4	13.5	41.1	59.3	66.2	740	672
Dispersité	1.8	1.1	1.4	2.2	1.5	1.8	2.1

### 3.1.5 Mise en œuvre de la séparation par $\mu$ LAS

#### 3.1.5.1 Format capillaire

En format capillaire, les séparations sont mises en œuvre sur un système d'électrophorèse capillaire Agilent 1600 CE équipé d'un détecteur Zetalif LED 480 nm de Picometrics (Figure 22, B). Les capillaires sont achetés chez Polymicro, de nombreux diamètres sont disponibles de 5 à 320  $\mu$ m de diamètre interne. Une fenêtre de détection est créée dans le capillaire qui est inséré dans la cellule de détection et dans la cassette de l'instrument (Figure 22, A). Avant les séparations, les capillaires sont conditionnés par un lavage à l'acide chlorhydrique (HCl), un coating avec du polyvinylalcool à 1% et par le tampon d'analyse, composé de Bis-Tris (30 mM), PIPES (10 mM), EDTA (1 mM) et supplémenté en

PVP. La détection de l'ADN est faite par fluorescence via un marquage au YO-PRO-1, un intercalant commercialisé par ThermoFisher, présent dans le tampon d'analyse à 200 nM. Pour les expériences de séparation deux références de taille ont été utilisées : le 1 kb extend DNA ladder de NEB contenant treize fragments de 0,5 kb à 48,5 kb et un mélange de quatre fragments de 50, 100, 150 et 210 kb. Les échantillons d'ADN sont analysés à une concentration de 1 ng/ $\mu$ L à l'aide d'une injection électrocinétique à 80 V/cm durant 10 s.

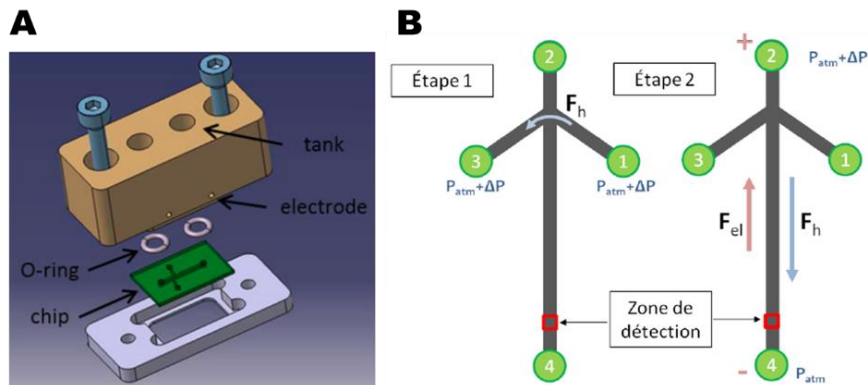


**Figure 22 : Instrumentation pour la séparation par  $\mu$ LAS au format capillaire. A : Insertion du dispositif capillaire dans la cassette Agilent. La flèche rouge désigne la zone de détection. B : La cassette est ensuite introduite dans un instrument d'électrophorèse capillaire Agilent 7100 CE relié à un détecteur Picometrics Zetalif LED.**

### 3.1.5.2 Format puce microfluidique

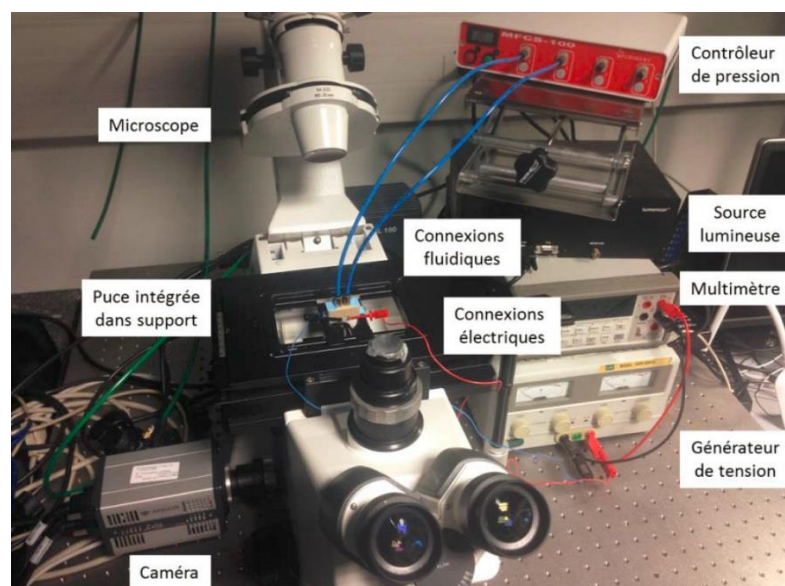
Les puces sont fabriquées dans la salle blanche du LAAS, le protocole de fabrication (voir A) ayant été développé par Hubert Ranchon durant sa thèse<sup>161</sup>. Brièvement, elles sont composées d'un substrat en silicium dans lequel sont gravés les canaux et d'un capot en verre qui sert à encapsuler la puce et à effectuer l'observation par microscopie. Elles sont intégrées dans un support permettant d'appliquer pression et tension et servant de réservoir pour les solutions (Figure 23, A). Ces puces sont composées d'un motif en forme de croix, avec deux canaux latéraux (1-3) permettant l'injection et un canal central (2-4) permettant d'effectuer la séparation. Les quatre extrémités correspondent à des réservoirs connectés à des contrôleurs de pression où les solutions tampons (réservoir 2,3 et 4) et l'ADN marqué (réservoir 1) sont injectés. La première étape consiste à régler les pressions de sorte que l'ensemble des molécules s'écoulent du réservoir 1 au 3, ceci dans l'objectif de former une bande ou « plug » d'ADN (Figure 23, B). Dans une seconde étape, la configuration des pressions change afin d'envoyer le « plug » vers le réservoir 4 tout en appliquant un champ électrique avec des électrodes insérés dans les réservoirs 2 (anode) et 4 (cathode). La détection par fluorescence est réalisée en fin

de canal de séparation grâce à un marquage de l'ADN avec un intercalant, le YOYO-1, à une concentration de 1 molécules pour 5 bp d'ADN.



**Figure 23 : A : Support permettant d'intégrer les puces composées de réservoirs pour les différentes solutions, de connectiques et d'électrodes pour appliquer pression et tension dans la puce. B : Schéma illustrant les étapes de la séparation. L'étape 1 consiste à former un « plug » d'ADN à la jonction des canaux. L'étape 2 correspond à l'étape de migration de l'ADN grâce à une force hydrodynamique ( $F_h$ ) et freiné par une force électrophorétique en opposition ( $F_{el}$ ).**

Afin d'appliquer pression et tension dans la puce et d'effectuer l'observation de la migration, le support est placé sous un microscope à épifluorescence Zeiss muni d'une caméra EMCCD (Electron Multiplying Charge Coupled Device, Andor Zyla) et d'une source de lumière Lumencore Light Engine. La prise d'images et de vidéos est réalisée à l'aide du logiciel open source Micro-Manager. La pression est générée par couplage d'un contrôleur de pression (Fluigent MFCS 1 Bar) directement dans le support à l'aide de connectiques pneumatiques. La tension est appliquée par un générateur de tension DC directement connecté aux électrodes (Figure 24).



**Figure 24 : Présentation du montage expérimental. Le support contenant la puce est placé sur un microscope inversé à épifluorescence et connecté à un contrôleur de pression, un générateur de tension et une caméra.**



## 3.2 Séparation d'ADN de haut poids moléculaire par $\mu$ LAS

Dans cette partie, nous détaillerons les conditions expérimentales qui ont permis de séparer et concentrer de grands fragments d'ADN avec  $\mu$ LAS. Nous verrons dans un premier temps que le choix du tampon et des paramètres d'actionnement est essentiel pour obtenir une bonne résolution de séparation dans une gamme de taille d'ADN. Nous aborderons ensuite les séparations sur puces microfluidiques, permettant de réduire les temps d'analyse mais comportant certaines limites, que nous discuterons en conclusion.

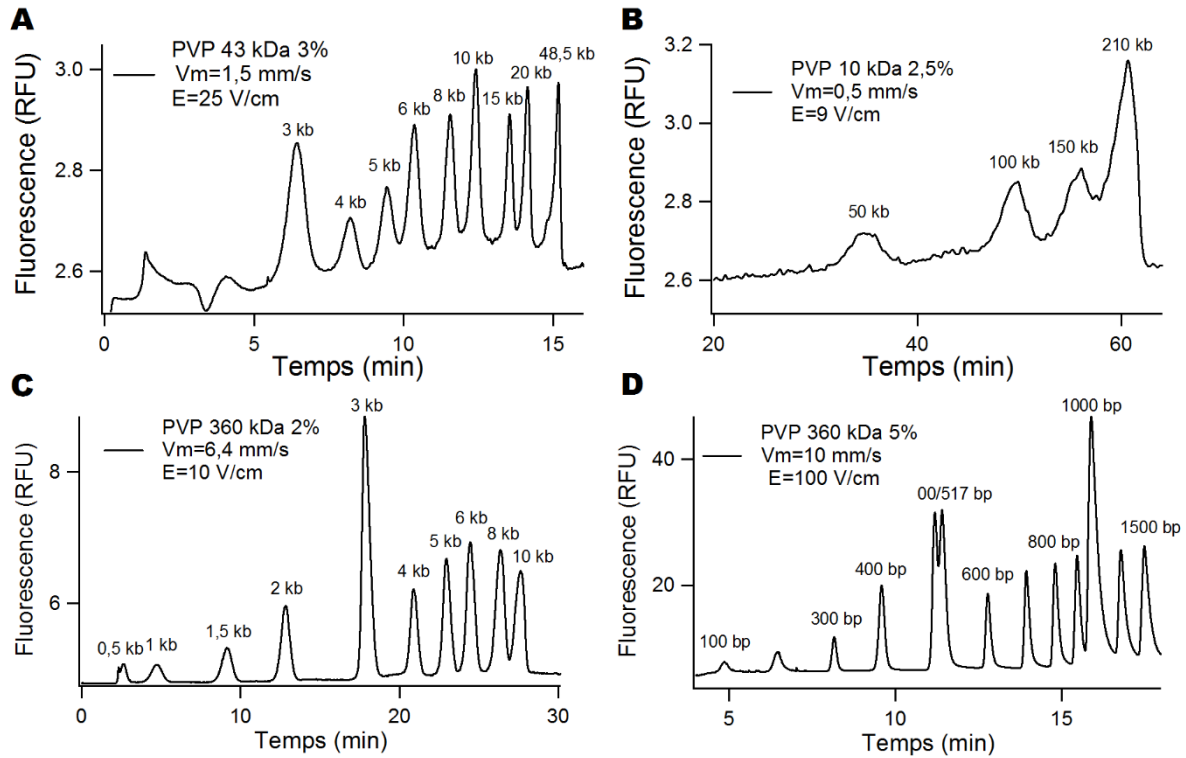
### 3.2.1 Optimisation des conditions opératoires en capillaire

Les travaux précédents cette thèse effectués par Picometrics ont montré la possibilité de séparer le kb extend ladder (0,5 – 48,5 kb) en 20 min avec des résolutions (voir Lexique) comprises entre 0,5 et 1,5 pour les plus grands fragments. Ces séparations ont été effectuées dans un tampon TBE supplémenté de PVP 360 kDa à 2% pour une vitesse moyenne d'écoulement de 6,4 mm/s et un champ électrique de 10 V/cm<sup>6</sup>. Il n'a cependant pas été possible de séparer des fragments plus grands ou d'augmenter la résolution de séparation avec ces paramètres.

Afin d'améliorer ces séparations et de travailler avec de plus grands fragments les différents paramètres influençant la séparation  $\mu$ LAS ont été optimisés (voir 3.1.3) : vitesse d'écoulement, champ électrique, poids moléculaire, concentration et origine (commerciale ou IMRCP) de la PVP dans le tampon. Les résultats de cette optimisation sont détaillés dans la publication constituant le Chapitre 4 de ce manuscrit.

Cette optimisation a débouché sur l'utilisation d'une PVP synthétisée par les IMRCP (Toulouse, France) d'un poids moléculaire de 43 kDa dissoute à 3% (w/v) dans un tampon Bis-Tris Pipes, d'une vitesse d'écoulement de 1,5 mm/s et d'un champ électrique de 25 V/cm dans des capillaires de 25  $\mu$ m de diamètre. Ces paramètres permettent la séparation du kb extend ladder en 15 min pour une résolution maximale de 3 entre le fragment de 20 kb et celui de 48,5 kb (Figure 25, A). Cette formulation du tampon a été utilisée dans la suite des expériences de ce manuscrit.

La séparation d'un échantillon composé de quatre fragments de 50, 100, 150 et 210 kb a également été effectuée en 1 h avec ce tampon en abaissant la vitesse d'écoulement à 0,5 mm/s et le champ électrique à 9 V/cm (Figure 25, B).



**Figure 25 : Séparation par  $\mu$ LAS de différents ladders ADN grâce à différentes formulations du tampon d'analyse. A : Séparation du kb extend ladder avec de la PVP 43 kDa à 3%. B : Séparation d'un ladder de quatre tailles avec de la PVP 10 kDa à 2,5%. C : Séparation du 1 kb ladder avec de la PVP 360 kDa à 2%. D : Séparation du 100 bp ladder avec de la PVP 360 kDa à 5%.**

Ce travail d'optimisation a mis en lumière l'importance des conditions d'actionnement (pression, tension) et des caractéristiques rhéologiques du tampon d'analyse pour travailler avec différentes gammes de taille d'ADN<sup>1</sup>. Ainsi, pour séparer des petits fragments il est nécessaire d'avoir une force transverse importante et donc des fluides fortement viscoélastiques (Figure 25, D) alors que la séparation de grands fragments requiert des conditions d'écoulement plus douces et un tampon avec de la PVP très diluée. La Figure 25 récapitule les performances de la technologie  $\mu$ LAS dans différentes gammes de taille d'ADN en faisant varier la concentration et la taille des chaînes de PVP dans le tampon pour moduler la viscoélasticité de la solution.

Il est également intéressant de noter que les solutions de PVP utilisées pour séparer les grands fragments d'ADN sont diluées en dessous de la concentration critique de recouvrement, i.e. 5,8% pour du PVP 40 kDa<sup>166</sup>. Au contraire, les séparations d'ADN de bas poids moléculaire sont faites en réseaux enchevêtrés avec de la PVP 360 kDa entre 2 et 5% soit 2,5 à 6 fois au-dessus de la concentration de

<sup>1</sup> Notez que dans l'optique d'augmenter les débits fluidiques et de manipuler de plus grandes quantités d'ADN, nous avons réalisé des expériences de séparation d'ADN dans des capillaires de 25, 40 et 50  $\mu$ m de diamètre sans que cela n'affecte la qualité des séparations.

recouvrement de 0,8%. Comme contrôle, des séparations d'ADN de 3 à 48 kb dans des tampons « newtoniens » ne contenant pas de PVP ont été effectuées. Les séparations ont également pu être réalisées dans ces conditions, moyennant l'utilisation de vitesses d'écoulement et électrophorétique légèrement supérieures à celles appliquées avec un tampon contenant de la PVP. Il semblerait donc que la PVP ne soit pas essentielle au phénomène de séparation  $\mu$ LAS pour l'ADN de haut poids moléculaire. Comme nous le verrons dans le chapitre 4 ou dans les différents travaux de l'équipe sur l'ADN de bas poids moléculaire<sup>6,9,10</sup>, l'ajout de PVP augmente tout de même considérablement la force viscoélastique de plaquage vers les parois. De plus, l'optimisation décrite dans le chapitre 4 de ce manuscrit indique qu'un tampon contenant de la PVP permet une séparation optimale de la gamme 3 à 48 kb.

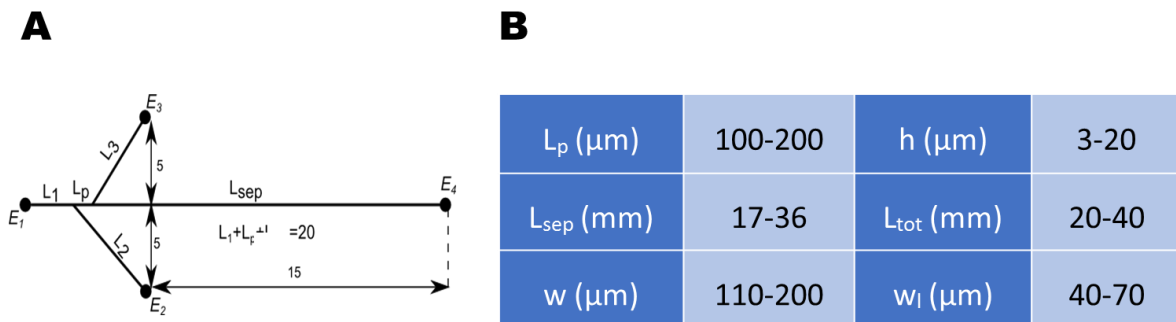
Mais comment expliquer l'origine des séparations obtenues dans des tampons sans propriétés viscoélastiques ? En l'absence de PVP, il est possible qu'une viscoélasticité résiduelle subsiste, éventuellement induite par les molécules d'ADN comme ont pu le décrire Kang et al<sup>174</sup> (voir 3.1.2). Il peut également s'agir d'un phénomène migratoire semblable à celui décrit par Yeung et Zheng, bien que dans nos migrations, les positions axiales d'équilibre sont dépendantes de la taille contrairement à leurs observations (voir 3.1.1). En 2007, Butler et al. ont effectué des simulations numériques sur la migration de polymères, modélisés comme des altères viscoélastiques, dans des fluides newtoniens sous l'actionnement d'un écoulement laminaire et d'une électrophorèse. Leurs simulations ont mis en évidence une migration vers les parois de la canalisation lorsque l'écoulement et l'électrophorèse sont en opposition<sup>175</sup>. Ce résultat indique donc que les propriétés viscoélastiques d'un fluide induisent une migration de bille ou d'ADN vers la paroi en présence d'électrophorèse, mais aussi que les molécules longues et déformables sont susceptibles d'être plaquées aux bords du canal. Autrement dit, les effets viscoélastiques et de déformabilité co-existent probablement dans nos expériences, sans que l'on sache encore évaluer leur importance respective.

## 3.2.2 Vers des séparations rapides en puces microfluidiques

### 3.2.2.1 *Optimisation de la géométrie des puces*

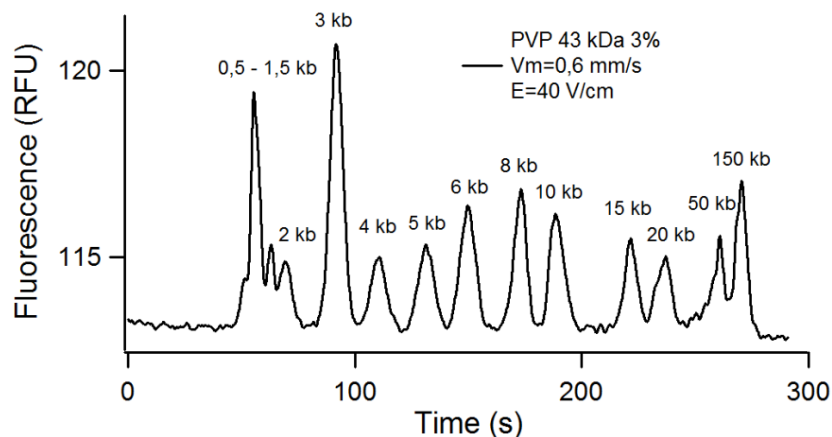
L'utilisation de puces microfluidiques permet d'accélérer les séparations et de diminuer la consommation de tampons et d'échantillons. Une des principales limitations des technologies de séparation d'ADN de haut poids moléculaire est la durée des analyses. Ainsi, le développement d'une technologie sur un format micro puce nous a semblé pertinent afin d'évaluer rapidement la distribution en taille et la concentration d'un échantillon d'ADN de haut poids moléculaire. Les résultats établis durant la thèse d'Hubert Ranchon montraient la séparation d'un marqueur de taille de 0,5 à 10 kb en environ 10 min. Le contrôle qualité des échantillons pour la génomique de 3<sup>ème</sup>

génération nécessite des technologies analytiques performantes jusqu'à 200 kb. Pour atteindre ces tailles, les optimisations de tampons identifiées sur capillaire ont été transposées au format puce. Les dimensions de la puce ont également été optimisées (Figure 26) en travaillant notamment sur la hauteur de la canalisation et les distances de séparation.



**Figure 26 : A : Schéma général des puces utilisées pour effectuer les séparations par  $\mu$ LAS. B : Dimensions testées pour les différents canaux de la puce. W et Wl étant les largeurs du canal central et des canaux latéraux respectivement.**

Cette optimisation a conduit à l'utilisation de puces de hauteur  $h=3 \mu\text{m}$  et une longueur de séparation de 36 mm. Des séparations du kb extend ladder (0,5 – 48,5 kb) supplémenté d'un fragment de 150 kb ont ainsi été obtenues avec des résolutions de 1,5 entre les fragments de 48 et 150 kb. De plus, ces migrations sont faites en moins de 5 min, bien plus rapidement que les séparations sur capillaires (Figure 27). Les vitesses d'écoulement de 0,6 mm/s sont assez proches de celles utilisées en capillaires mais les champs électriques utilisés sont plus importants, de 40 V/cm. De manière générale, plus la hauteur de la puce (ou le diamètre du capillaire) est faible et plus le champ électrique nécessaire pour avoir une séparation est important. En effet, lorsque la hauteur de la canalisation diminue, le cisaillement augmente. L'ADN plaqué contre les parois, du fait de sa taille et de sa flexibilité, est également soumis à des vitesses plus rapides du centre de la canalisation. Une contre électrophorèse plus importante est alors nécessaire pour freiner l'ADN.



**Figure 27 : Séparation du kb extend ladder + 150 kb en puce avec un tampon PVP 43 kDa à 3%, une vitesse moyenne d'écoulement de 0,6 mm/s et un champ électrique de 40 V/cm.**

Lors de cette optimisation des puces ayant une hauteur de canalisation  $h=20\ \mu\text{m}$  ont également été testées. Dans ces conditions, l'état transitoire avant que les molécules d'ADN migrent à leur position d'équilibre dans l'écoulement est assez long. En effet, lors de la transition en conditions de séparation, une partie du plug d'ADN part très rapidement dans la canalisation alors qu'une autre partie est plaquée à la paroi et migre lentement. Cette observation a été nommée « l'effet inertiel », correspondant à la migration rapide d'une partie de l'ADN au centre de la canalisation qui n'est pas affectée immédiatement par l'effet  $\mu$ LAS. Les molécules initialement proches de la paroi vont être immédiatement ralenties alors que celles du centre, où le cisaillement est nul, ne sont pas affectées. Cela entraîne une forte dispersion transitoire des vitesses des molécules d'ADN et empêche la séparation. Les puces de  $20\ \mu\text{m}$  de profondeur ne sont donc pas adaptées pour ces séparations. Cet effet n'est pas ou peu observé à Picometrics sous format capillaire alors que les diamètres des canalisations sont pourtant de  $50\ \mu\text{m}$ . Une hypothèse est que l'effet devient négligeable devant les longueurs de séparation en capillaires.

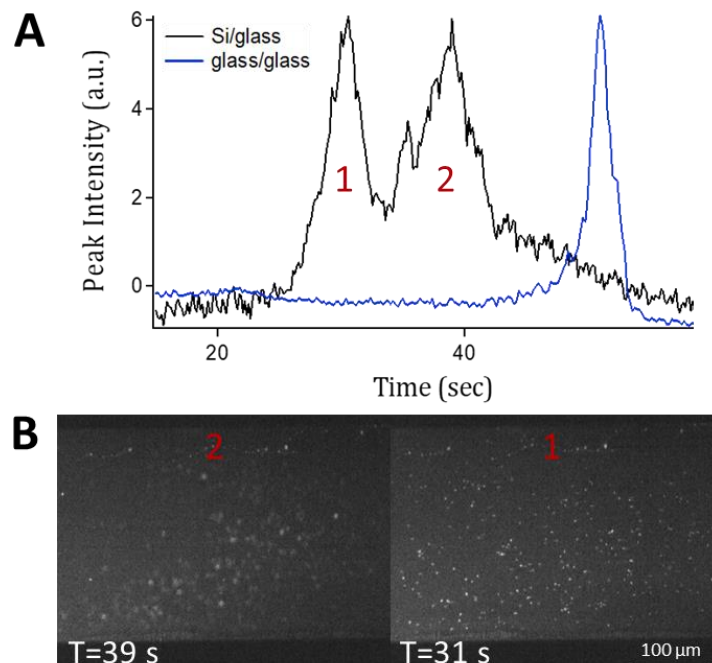
### 3.2.2.2 *Une limite de séparation vers 100 kb*

Les séparations du kb extend ladder supplémenté d'une bande unique d'ADN étant prometteuse, le travail a été poursuivi avec un échantillon plus complexe comprenant un mélange de quatre fragments de 50, 100, 150 et 210 kb. Cependant, la séparation de ce marqueur n'a pas pu être effectuée de façon satisfaisante quels que soient les paramètres d'actionnement testés, les fragments de 100, 150 et 210 kb ayant des mobilités très proches.

Nous avons vu lors du Chapitre 2 que l'utilisation d'un champ électrique variable permettait d'augmenter la gamme de taille d'ADN séparable en électrophorèse en jouant sur le temps de réorientation des molécules. Ce principe a été transposé à nos séparations sur puces en utilisant un générateur de tension alternée. La modulation de différents paramètres tels que l'offset (voltage moyen appliqué), l'amplitude (variation du voltage au-dessus et en dessous de l'offset), et la période (le temps entre chaque variation de voltage) a été testée afin de séparer de plus grands fragments. Cependant, les différents tests effectués n'ont pas permis d'améliorer la séparation des plus grands fragments. Les conditions testées n'ont pas non plus permis d'identifier clairement l'influence de chacun de ces paramètres sur la qualité de la séparation. L'introduction de champs électriques variables augmente encore le nombre de paramètres pouvant influencer la migration  $\mu$ LAS. Ainsi, de nouvelles expériences faisant appel à des plans d'expériences seraient utiles afin d'identifier plus clairement les bénéfices de ces paramètres pour la séparation d'ADN de haut poids moléculaire.

### 3.2.2.3 Anomalies de migration en puces microfluidiques

Ces travaux sur puce ont également permis de mettre en évidence un phénomène de dédoublement de bande d'ADN dans les puces Silicium/verre (substrat/capot). En effet, des migrations effectuées avec une seule taille d'ADN dans l'échantillon ont conduit à la présence de deux populations en fin de canal (Figure 28, A, tracé noir). L'observation de ces populations au microscope a révélé une différence de positionnement verticale entre la première (Figure 28, B, population de droite nette) et la seconde population d'ADN (Figure 28, B, population de gauche floue). Cela montre que les différences de mobilité dépendent de la hauteur de migration de l'ADN dans le canal. Les molécules d'ADN se répartissent en deux groupes, l'un au « plafond » et l'autre au « plancher » du canal de séparation.



**Figure 28 : Dédoublement de bande d'ADN en puces microfluidiques. A : Signal de fluorescence lors de la migration d'un échantillon d'ADN de 75 kb en puces Silicium/Verre (noir) ou en puces Verre/Verre (bleu). Deux populations distinctes (1 et 2) sont observées en fin de canalisation pour une seule taille d'ADN. B : Observation des deux populations 1 et 2 en fin de canalisation. Le panneau de droite correspond à une fraction de l'ADN de 75 kb plaquée au plafond de la puce et migrant plus rapidement. Le panneau de gauche correspond à la seconde fraction de l'ADN de 75 kb plaquée au plancher du canal et migrant plus lentement. L'observation est faite au même endroit dans la puce et sans modification de la mise au point.**

Ce dédoublement provient probablement d'une différence d'état de surface entre le capot en verre et le substrat en silicium de la puce. Cela entraînerait une différence d'électro-osmose entre le haut et le bas du canal. En effet, ce dédoublement a été observé principalement à fort voltage et il est facile à mettre en évidence après un lavage à la soude qui régénère les silanols de surface et maximalise

le phénomène d'électro-osmose. Afin de résoudre ce problème, il est nécessaire d'obtenir deux surfaces identiques ou de supprimer l'électro-osmose. Dans un premier temps, différents coatings, dynamiques ou covalents, ont été testés afin de limiter ou d'homogénéiser l'électro-osmose sur la puce. Cependant aucun n'a permis de résoudre complètement le dédoublement d'une bande unique.

L'équipe au LAAS, et principalement Bayan Chami et Inga Tjunelyte, ont donc développé un nouveau procédé de fabrication de puces utilisant un substrat et un capot en verre identiques afin d'obtenir les mêmes états de surface. Nous avons utilisé une résine photosensible, la PermiNex™ 2000 de MicroChem, qui permet à la fois de faire les canaux via une simple photolithographie, mais également de coller le capot au substrat. Une publication relative à cette nouvelle méthode de fabrication est en cours d'écriture. L'utilisation de ces nouvelles puces a permis de résoudre le dédoublement des bandes avec l'obtention d'une seule population d'ADN en fin de migration lorsqu'une seule taille est introduite dans la puce (Figure 28, A, tracé bleu).

### 3.2.3 Comparaison des formats puce et capillaire

En termes de performances de séparation, le format capillaire permet d'analyser des fragments d'ADN sensiblement plus grands que le format puce. Le format capillaire facilite également le balayage des conditions opératoires grâce à l'automatisation de l'instrument bien que les séparations soient plus longues. Les expériences au format puce sont plus rapides à conduire mais nécessitent une préparation en amont beaucoup plus importante ainsi qu'une intervention constante de la part de l'expérimentateur (Tableau 8). Elles ont néanmoins permis de nombreuses observations sur la migration de l'ADN et ont grandement participé aux progrès effectués en capillaire. En effet, il est possible de suivre au microscope les molécules d'ADN tout au long de la séparation. La formation d'agrégats de molécules lorsque la concentration était trop élevée a ainsi été observée. Ces agrégats entraînent des anomalies de migration qu'il était plus complexe d'interpréter en format capillaire. De plus, le format puce offre plus de modularité en termes de dimensions des canaux. Nous constaterons dans le paragraphe suivant que les dimensions des assemblages de capillaires sont fortement contraintes par les caractéristiques de l'instrument d'électrophorèse capillaire et par leur fragilité.

Comme nous le verrons dans les chapitres 4 et 5 de ce manuscrit la plupart des résultats obtenus durant cette thèse l'ont été en format capillaire. Il est vrai que le problème de dédoublement de bandes illustré précédemment a longtemps freiné les progrès au format puce. D'autre part, nous nous sommes beaucoup intéressés à des problématiques de préparation d'échantillon et le format puce est moins adapté à la collecte des échantillons après analyse. L'instrument d'électrophorèse capillaire Agilent possède lui un carrousel pouvant contenir 40 flacons ce qui facilite les collectes.

Même s'ils sont assez faibles, les débits du format capillaire restent plus importants que ceux du format puce et permettent de travailler de plus grandes quantités de matière (Tableau 8).

**Tableau 8 : Comparaison des performances du format puce et du format capillaire pour la séparation d'ADN de haut poids moléculaire avec la technologie  $\mu$ LAS.**

Format	Puce	Capillaire
Gamme séparation	50 – 150 000 bp	100 – 200 000 bp
Résolution 50-150kb	1.5	3.7
Temps d'analyse	5 min	60 min
Temps de préparation	60 min	15 min
Volume échantillon	1 $\mu$ L	5 $\mu$ L
Débit	30 nL/min	180 nL/min
Automatisation	Non	Oui

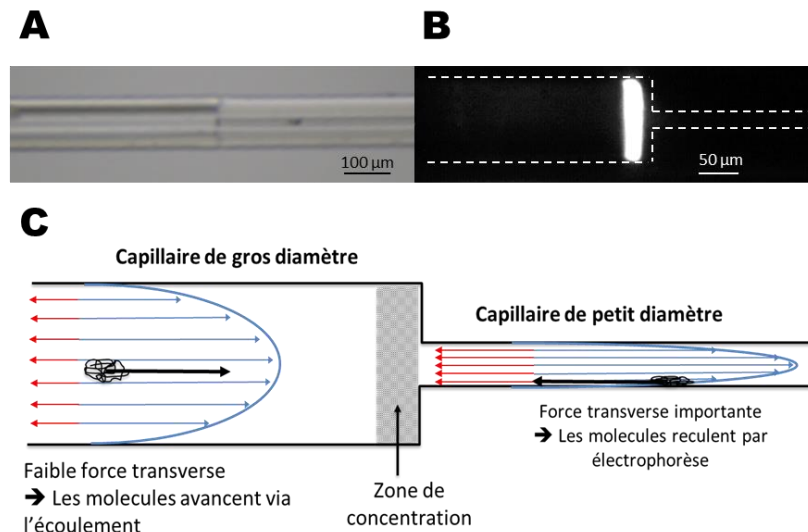


### 3.3 La concentration par $\mu$ LAS

La concentration d'ADN par  $\mu$ LAS est possible aux formats puce<sup>6</sup> et capillaire<sup>10</sup>. Cependant, la concentration de grands fragments d'ADN s'est avérée beaucoup délicate à mettre au point de manière automatisée avec le format puce. Ainsi, durant cette thèse, ce procédé a été uniquement réalisé sur des capillaires. Nous verrons dans la section suivante son principe de fonctionnement ainsi que ses performances afin de les comparer à l'état de l'art.

#### 3.3.1 Principe de fonctionnement

La concentration par  $\mu$ LAS est réalisée en modulant spatialement l'intensité des vitesses électrophorétiques et hydrodynamiques via l'introduction d'une constriction à la jonction de deux capillaires de différent diamètre (Figure 29, A). La conservation du débit implique une grande vitesse d'écoulement dans le capillaire de petite section par rapport à celle dans le capillaire de grand diamètre. Il en va de même pour la force électrophorétique. Ainsi, selon l'équation 10 (3.1.3), la force transverse est très importante dans le capillaire de petit diamètre et permet de plaquer toutes les molécules d'ADN aux parois du capillaire. Or, proche des parois, la vitesse d'écoulement est quasiment nulle. L'électrophorèse est alors supérieure à l'écoulement et les molécules d'ADN remontent le long des parois du capillaire. Dans le capillaire de gros diamètre, la force transverse est très faible et la migration reste dominée par l'écoulement hydrodynamique (Figure 29, C). Cela crée une zone de concentration d'une dizaine de microns de longueur à la jonction des deux capillaires (Figure 29, B)<sup>10</sup>.



**Figure 29 : Concentration  $\mu$ LAS au format capillaire. A : Assemblage de deux capillaires de 100  $\mu$ m et 20  $\mu$ m de diamètre interne afin de créer une constriction. B : Concentration  $\mu$ LAS d'un fragment de 5 kb marqué au YOYO-1 à la jonction de deux capillaires. C : Dans le capillaire de grand diamètre la force transverse est faible et la force hydrodynamique domine ; l'ADN avance jusqu'à la constriction. Dans le capillaire de petit diamètre, la force transverse est importante et plaque toutes les molécules aux parois du capillaire. Proche des parois l'électrophorèse domine et l'ADN remonte jusqu'à la constriction permettant l'enrichissement.**

Dans un premier temps, l'échantillon est injecté en forte quantité, ie, le volume du capillaire de gros diamètre, puis concentré via l'application d'un fort voltage (Figure 29, B et C). Ensuite le voltage est diminué afin de relarguer progressivement l'ADN de la constriction et de le séparer dans le capillaire de petit diamètre comme précédemment décrit (voir 3.2.1).

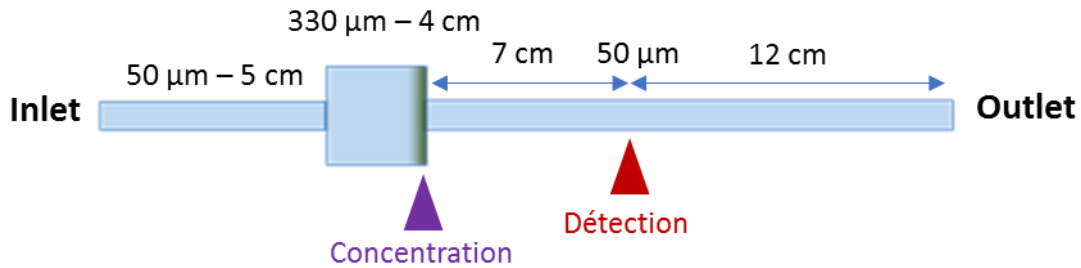
Les étapes de concentration et de séparation sont donc séquentielles. Le format puce microfluidique permet d'effectuer ces étapes de manière simultanée comme l'ont montré Hubert Ranchon et Rémi Malbec dans leur thèse pour les ADN de bas poids moléculaire<sup>6,9,161</sup>. Ce format n'est en revanche pas adapté à la concentration d'ADN de haut poids moléculaire. En effet, lorsque les grandes molécules d'ADN sont concentrées dans les puces, elles collent fortement aux parois ce qui empêche toute séparation et rend la puce inutilisable. Ces effets d'adsorption aux parois sont beaucoup plus marqués pour les grandes molécules que pour les petites<sup>176</sup>. Ainsi les expériences impliquant la concentration d'ADN de haut poids moléculaire ont été exclusivement menées au format capillaire durant cette thèse.

### 3.3.2 Dimensions du dispositif

L'instrument d'électrophorèse capillaire Agilent impose un certain nombre de contraintes sur les dimensions des capillaires. En effet, le capillaire est placé dans une cassette qui est elle-même insérée dans l'instrument (voir 3.1.5.1, Figure 22). Ainsi le dispositif doit, au minimum, faire une longueur de 30 cm afin d'être intégré dans la cassette. Il faut également que le capillaire soit inséré dans la cellule de détection, à une position fixe dans la cassette.

Ensuite, l'instrument peut appliquer des pressions uniquement comprises entre 0 et 50 mBar et entre 2 et 10 Bar. Nous avons vu précédemment que l'utilisation de capillaires de 50  $\mu$ m de diamètre interne avec des écoulements proches de 1,5 mm/s permettaient d'obtenir des séparations de qualité jusqu'à 200 kb. Afin d'obtenir ces écoulements, il faut ajuster la résistance hydraulique (voir B) du dispositif avec les diamètres et longueurs des capillaires pour opérer dans la gamme de pression applicable par l'instrument. Il faut ensuite définir la longueur du capillaire de gros diamètre servant de chambre d'injection et créant la constriction avec le capillaire de 50  $\mu$ m. Plus cette chambre est grande et plus le volume injecté (et donc la sensibilité) est grand. Cependant, l'écoulement étant particulièrement lent dans le capillaire de grand diamètre, de trop grandes distances induisent des temps d'analyses très longs. Nous avons ainsi opté pour un capillaire de 330  $\mu$ m de diamètre et de 4 cm de long. Ce diamètre de 330  $\mu$ m permet d'emmancher le capillaire de séparation de 50  $\mu$ m. Le capillaire de 330  $\mu$ m s'abîme rapidement lorsqu'il est placé directement dans l'électrode. Ainsi, un autre capillaire de 50  $\mu$ m est emmanché à l'entrée du dispositif, qui sert également de filtre pour limiter le bouchage du dispositif (Figure 30). Les capillaires emmanchés sont collés à l'aide d'une colle

époxy<sup>10</sup>. De plus, les jonctions entre les capillaires sont des zones particulièrement fragiles et sont sensibles à la moindre contrainte mécanique. Il est donc important que les dimensions de l'assemblage n'induisent pas de courbure au niveau des jonctions lors de l'insertion dans la cassette de l'instrument d'électrophorèse capillaire. Ces différents éléments nous ont amené à l'utilisation d'un dispositif aux dimensions suivantes :



**Figure 30 : Dimensions du dispositif permettant d'effectuer les concentrations et les séparations par  $\mu$ LAS. Le dispositif est un emmanchement de deux capillaires de  $50\ \mu\text{m}$  dans un capillaire de  $330\ \mu\text{m}$ . Le capillaire de  $330\ \mu\text{m}$  correspond à la chambre d'injection et de concentration et le capillaire de  $50\ \mu\text{m}$  permet d'effectuer la séparation. La détection est effectuée à  $7\ \text{cm}$  de la zone de concentration.**

### 3.3.3 Analyse haute sensibilité de l'ADN de haut poids moléculaire

Une fois le design définitif établi nous avons mis au point une méthode d'analyse de l'ADN dans la gamme  $2 - 100\ \text{kb}$  comprenant une étape de concentration suivie d'une séparation. L'échantillon est injecté pour un volume maximum de  $1,7\ \mu\text{L}$ , i.e., un demi-volume de la chambre d'injection de  $330\ \mu\text{m}$  de  $4\ \text{cm}$ . Il est ensuite concentré pendant  $20\ \text{min}$  à la constriction avec un champ électrique de  $166\ \text{V/cm}$  et une vitesse moyenne d'écoulement de  $1,5\ \text{mm/s}$  ( $50\ \text{mBar}$ ) dans le capillaire de  $50\ \mu\text{m}$  (Figure 31, courbe rouge). Enfin, le voltage est abaissé progressivement jusqu'à  $25\ \text{V/cm}$ , i.e., le champ électrique établi en 3.2.1 pour une séparation optimale, afin de relâcher progressivement les fragments du concentrateur et de les séparer dans le capillaire de  $50\ \mu\text{m}$ . Le détecteur, placé à  $7\ \text{cm}$  de la constriction, permet de collecter l'intensité de fluorescence en fonction du temps.

Avec ce dispositif nous pouvons séparer des fragments d'ADN de  $2$  à  $100\ \text{kb}$  en  $1\ \text{h}$  environ avec des résolutions de  $2$  entre un fragment de  $50\ \text{kb}$  et un de  $100\ \text{kb}$  (Figure 31, Courbes noire et bleue). De plus l'injection d'un grand volume permet d'avoir une excellente sensibilité, de l'ordre de  $10\ \text{fg}/\mu\text{L}$  sur un fragment de  $10\ \text{kb}$  ou encore  $300\ \text{aM}$  sur le fragment de  $50\ \text{kb}$ . En électrophorèse capillaire classique, le volume d'injection est de l'ordre de quelques nL. Ici, l'injection de  $\sim 1\ \mu\text{L}$  offre une sensibilité environ  $1000$  fois supérieure.

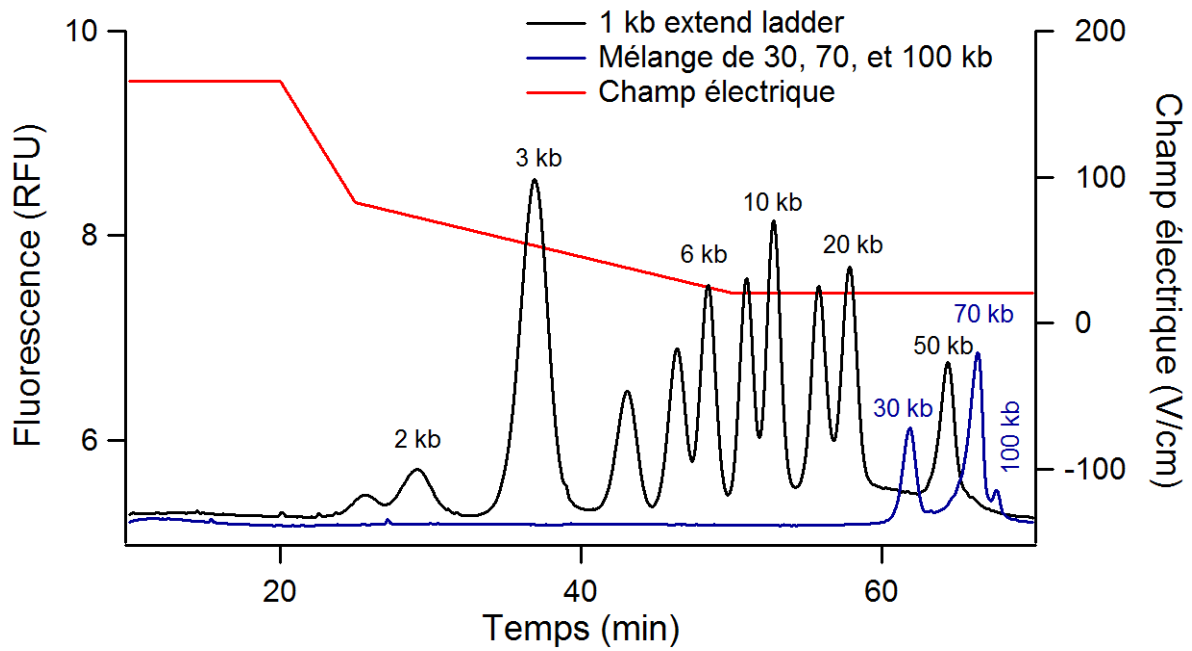


Figure 31 : Analyse par  $\mu$ LAS du kb extend ladder et d'un mélange contenant trois fragments de 30, 70 et 100 kb. Le courbe rouge représente l'évolution du champ électrique en fonction du temps. Une première étape de 20 min à 166 V/cm permet de concentrer l'échantillon. Puis le champ électrique diminue afin de séparer l'ADN dans le capillaire de petit diamètre. La courbe noire représente l'intensité de fluorescence en RFU collectée lors de la séparation du kb extend ladder à 100  $\mu$ g/ $\mu$ L. La courbe bleue correspond à la séparation du mélange de trois fragments à 25  $\mu$ g/ $\mu$ L.

Cette méthode nous permet d'évaluer la taille et la concentration d'un échantillon en comparant les temps de migration et l'intensité de fluorescence à une référence de taille et de masse calibrée. L'analyse est faite de manière séquentielle, la référence étant toujours analysée avant une série d'échantillons inconnus.

Nous avons également évalué la fidélité et la justesse des mesures de taille et de concentration en effectuant plusieurs séries d'analyses du kb extend ladder (onze analyses). L'évaluation en taille d'un échantillon est précise à 2% CV et sa justesse est de 3% (Tableau 9). La quantification d'un échantillon est précise à 5% CV et sa justesse est d'environ 15% (Tableau 10).

**Tableau 9 : Fidélité et justesse de l'évaluation en taille sur la gamme 3 – 48,5 kb d'un échantillon par la méthode d'analyse développée sur le principe  $\mu$ LAS en capillaire.**

Taille attendue (bp)	Taille calculée (bp)	SD	Fidélité (% CV)	Justesse (%)
<b>3000</b>	2947	21.3	1	2
<b>4000</b>	3941	26.4	1	1
<b>5000</b>	4935	37.4	1	1
<b>6000</b>	5900	58	1	2
<b>8000</b>	7955	100	1	1
<b>10000</b>	9930	209	2	1
<b>15000</b>	15530	667.7	4	3
<b>20000</b>	19735	857.8	4	1
<b>48500</b>	51300	1249.2	2	5

**Tableau 10 : Fidélité et justesse de la quantification d'un échantillon sur la gamme 3 – 48,5 kb par la méthode d'analyse développée sur le principe  $\mu$ LAS en capillaire.**

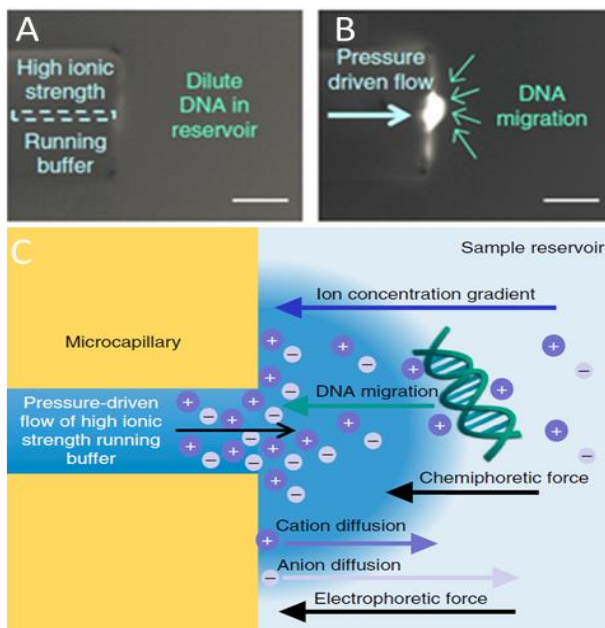
Fragment (kb)	Masse attendue (pg)	Masse calculée (pg)	SD	Fidélité (% CV)	Justesse (%)
<b>3</b>	100	104.1	1.1	1	4
<b>4</b>	26	27.5	0.3	1	6
<b>5</b>	33	31.9	0.8	3	3
<b>6</b>	40	38.9	1.1	3	3
<b>8</b>	33	32.7	1.2	4	1
<b>10</b>	33	35.0	1.8	5	6
<b>15</b>	45	38.0	1.3	4	16
<b>20</b>	30	33.6	2.5	6	12
<b>50</b>	30	34.0	2.8	9	13

### 3.3.4 Comparaison des performances avec l'état de l'art

Afin d'analyser des échantillons dilués, les méthodes classiques d'électrophorèse d'ADN ont également besoin d'étapes de concentration. Celles-ci peuvent être réalisées en amont de l'analyse ou effectuées directement par l'instrument avant la séparation. La méthode la plus classique en électrophorèse capillaire est le « stacking », basé sur la différence de vitesse de migration de l'échantillon dans son tampon ou dans celui d'analyse<sup>177</sup>. Pour cela, on place l'échantillon dans un tampon de plus faible conductivité que le tampon d'analyse. Ainsi, le champ électrique sera plus important dans la zone de faible conductivité et les fragments d'ADN vont rapidement migrer vers la zone de forte conductivité. Une fois atteinte, la vitesse des fragments diminue ce qui crée une zone de

concentration à l'interface des deux tampons. Cette méthode permet d'obtenir des facteurs de concentration de l'ordre de 100 avant l'étape de séparation<sup>177</sup>. Elle reste cependant limitée aux échantillons purifiés et élués dans des tampons peu salés. L'isotachophorèse, décrite en 2.1.4.2, permet également la concentration de l'échantillon à l'interface de deux tampons de différentes conductivités avec des facteurs d'enrichissement jusqu'à 1000 pour l'ADN<sup>178,179</sup>.

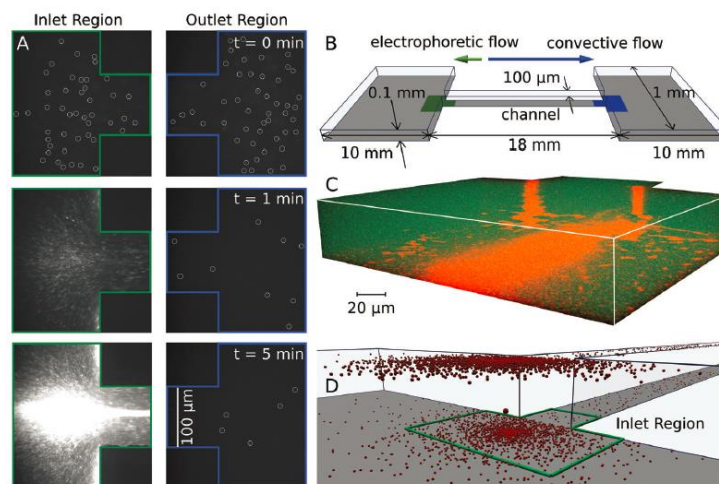
Plus récemment, Friedrich et al. ont développé une méthode appelée rhéotaxie moléculaire, et obtenu des facteurs de concentration de 10000 en quelques minutes<sup>180</sup>. Lorsque de l'ADN est placé dans un réservoir de faible conductivité et qu'un tampon de forte conductivité est injecté via un capillaire dans ce réservoir, l'ADN migre contre le flux et se concentre à l'entrée du capillaire (Figure 32, A-B). Le mécanisme de migration proposé repose sur l'utilisation d'anions ayant une diffusivité plus importante que les cations. Il se crée ainsi un gradient de concentration d'ions en sortie du capillaire induisant un champ électrique qui fait migrer l'ADN vers l'entrée du capillaire (Figure 32, C). Cette méthode d'enrichissement peut être couplée à une séparation hydrodynamique en solution libre (décrite en 3.1.1) en inversant le flux hydrodynamique après la concentration. Ils ont ainsi montré la séparation d'ADN de 2 à 27 kb à une concentration de 150 aM. La méthode reste limitée par la faible résolution de séparation qu'il est possible d'obtenir et par l'obligation de diluer l'ADN dans un tampon de faible conductivité. Les facteurs de concentration obtenus par cette méthode n'en restent pas moins impressionnants. L'entreprise Circulomics envisage le développement d'un instrument basé sur le couplage de ces deux technologies.



**Figure 32 : Principe de fonctionnement de la rhéotaxie moléculaire. A : L'ADN est dilué dans un réservoir peu salé connecté à un capillaire de faible diamètre (5  $\mu$ m) contenant un tampon plus salé. B : Un flux hydrodynamique injecte du tampon salé dans le réservoir ce qui entraîne la migration de l'ADN contre ce flux qui se concentre à l'entrée du capillaire. C : L'utilisation d'anions diffusant plus rapidement que les cations entraîne un gradient de concentration induisant une force électrophorétique dirigée vers l'entrée du capillaire.**

En 2016, un mois après la première publication de l'équipe sur la technologie  $\mu$ LAS<sup>6</sup>, l'équipe de Butler et al. en Floride a montré la concentration de grands fragments d'ADN génomique dans une constriction via un actionnement électro-hydrodynamique<sup>181</sup>. Le mécanisme de migration de l'ADN vers les parois de la puce ressemble fortement au phénomène  $\mu$ LAS bien que la publication décrive un modèle similaire aux travaux de Yeung et Zeng<sup>182</sup> (voir 3.1.1). La publication de Butler et al. mentionne d'ailleurs l'utilisation de la PVP (0,5% à 40 kDa) dans leur tampon afin d'effectuer un coating pour supprimer l'électroosmose qui perturberait leur concentration. Or, nous avons montré durant cette thèse la possibilité de séparer des grands fragments d'ADN dans des tampons contenant de faibles quantités de PVP voire même sans PVP. La publication mentionne le fait que sans coating PVP, la concentration d'ADN est moins performante ce confirme également nos observations. Avec ce dispositif (Figure 33) ils obtiennent des facteurs de concentration de l'ordre de 1000 par minute, très similaires à ceux obtenus par Rémi Malbec et Hubert Ranchon durant leur thèse.

Plus récemment, l'équipe de Butler a publié une seconde étude sur cette méthode de concentration d'ADN. Ils mentionnent les travaux de l'équipe sur  $\mu$ LAS comme étant un mécanisme différent, du fait des vitesses d'écoulement et électrophorétique plus élevées, et des propriétés viscoélastiques des tampons plus importantes<sup>183</sup>. Il est cependant très probable que leurs observations soient issues des mêmes phénomènes migratoires que l'effet  $\mu$ LAS. Dans cette publication, ils décrivent également un phénomène de saturation de la zone de concentration comme nous avons pu l'observer à Picometrics lorsque de trop grandes quantités d'ADN sont introduites dans le système (Chapitre 5).



**Figure 33 : Concentration d'ADN à l'entrée d'un canal microfluidique. A : En début d'expérience l'ADN est présent à l'entrée (vert) et la sortie du canal (bleu) puis se concentre à la constriction lors de l'application d'un écoulement de 1,2 mm/s et d'une électrophorèse de 50  $\mu$ m/s. B : Représentation schématique du dispositif utilisé. C : Microscopie confocale montrant les zones de concentration de l'ADN. D : Représentation schématique indiquant la concentration de l'ADN aux parois de la puce. Tiré de Butler et al.<sup>181</sup>.**

D'un point de vue commercial, l'instrument de référence pour l'analyse de l'ADN de haut poids moléculaire est le FemtoPulse d'AATI, pouvant séparer des fragments jusqu'à 165 kb avec une sensibilité de 300 fg/ $\mu$ L (voir 2.3.3.2). Bien que la technologie  $\mu$ LAS offre une meilleure sensibilité de 10 fg/ $\mu$ L, le pouvoir résolutif reste inférieur à celui du FemtoPulse sur les plus grands fragments (>100 kb). De plus, l'instrument d'AATI possède une cadence d'analyse bien plus importante, ce qui le rend plus adapté au contrôle qualité d'échantillons avant séquençage. La technologie  $\mu$ LAS permet cependant d'explorer des applications du côté préparatif grâce à la fonction de collecte qui n'est pas disponible sur le FemtoPulse qui est un instrument purement analytique. Comme nous le verrons dans les chapitres 4 et 5, la technologie  $\mu$ LAS permet des applications préparatives comme la collecte spécifique de grands fragments ou la sélection en taille d'échantillons pour le séquençage.



### 3.4 Conclusion du Chapitre 3

Dans ce troisième chapitre, nous avons décrit le principe de fonctionnement de la technologie  $\mu$ LAS. Elle repose sur l'existence d'une force transverse à l'axe d'un canal lors de l'application bidirectionnelle d'un flux hydrodynamique et d'un champ électrophorétique. Nous avons vu que la migration transverse est bien décrite dans la littérature dans le cas d'un simple écoulement ; elle est alors associée à la focalisation de particules ou d'ADN au centre de la canalisation. L'ajout d'un champ électrique en opposition change l'orientation de cette force. Nous avons présenté les résultats d'un modèle analytique décrivant cette force transverse avec la dépendance en termes d'actionnement, de propriétés rhéologiques de la solution et de la taille des molécules d'ADN, permettant ainsi leur séparation dans un écoulement de Poiseuille.

Un travail d'optimisation des propriétés viscoélastiques du tampon d'analyse a permis de réaliser des séparations d'ADN jusqu'à 200 kb dans des capillaires ou des puces microfluidiques. Nous avons ensuite implémenté une fonction de concentration en créant une constriction à la jonction de deux capillaires de différents diamètres. Ce dispositif de capillaire permet de concentrer jusqu'à 1,7  $\mu$ L d'échantillon et de le séparer en atteignant des limites de détection de 10 fg/ $\mu$ L, parmi les plus performantes de l'état de l'art.

Mise en œuvre sur un instrument d'électrophorèse capillaire automatisé, la technologie  $\mu$ LAS permet d'évaluer la taille et la concentration d'un échantillon avec une excellente fidélité ( $\sim$ 5%) et une bonne justesse ( $\sim$ 10%). Nous allons voir dans la suite de ce manuscrit que ces bonnes performances nous ont permis de répondre au cahier des charges de deux applications liées à la préparation d'échantillons d'ADN de haut poids moléculaire.

## 4. Capture spécifique de longs fragments d'ADN via un couplage entre la technique CRISPR/Cas9 et la technologie $\mu$ LAS

---

Nous avons vu dans notre premier chapitre que les technologies actuelles de séquençage sont capables de produire une quantité phénoménale de données, jusqu'à 15 Tb pour les plus performantes, et ainsi de séquencer des génomes entiers en quelques jours. Cependant, cette quantité de données dépasse parfois les capacités de stockage des plateformes et le traitement des données devient un des coûts les plus importants du séquençage<sup>184</sup>. De plus, il n'est pas toujours nécessaire de séquencer la totalité des génomes, certains projets requérant l'information d'une petite partie du génome seulement. Il existe donc un besoin d'effectuer du séquençage de manière ciblée afin d'extraire uniquement l'information requise et non pas des génomes entiers, beaucoup plus laborieux à étudier, même avec les récentes techniques de 3<sup>ème</sup> génération<sup>185</sup>. La complexité de certains génomes motive également le développement d'outils pour du séquençage ciblé. En effet certains génomes, de plantes notamment, peuvent faire plusieurs Gb de long avec jusqu'à 90% de séquences répétées pouvant faire plusieurs centaines de kb<sup>186</sup>. Le traitement bio-informatique de ces données de séquence est extrêmement long et difficile et induit des coûts importants qu'il est nécessaire de limiter<sup>187</sup>.

Les méthodes de séquençage spécifiques sont déjà particulièrement bien implantées en NGS. On distingue deux types de méthodes ; en capturant la séquence d'intérêt à l'aide de sondes ou en l'amplifiant par PCR. On peut citer notamment la méthode développée par Gnirke et al. reposant sur l'utilisation de sondes biotinilées, complémentaires de la séquence d'intérêt, qui vont être capturées via une interaction avec la streptavidine<sup>188</sup>. Pour un inventaire plus détaillé de ces différentes méthodes il est possible de se référer à la revue de Kozarewa<sup>189</sup>. Ces méthodes ne permettent cependant pas la capture de longs fragments d'ADN. De même, le séquençage spécifique via la PCR montre ses limites lors de l'études de régions riches en GC<sup>190</sup> ou contenant de longues répétitions, supérieures à 10 kb<sup>191</sup> et peut induire des biais de séquençage. Le développement de méthodes alternatives pour du séquençage ciblé de grands fragments est donc en plein essor.

La découverte en 2012 de la technologie CRISPR/Cas9<sup>192</sup>, permettant de cibler n'importe quelle région d'un génome avec une précision inédite, est une incroyable opportunité pour le développement des techniques de séquençage spécifique. Cependant, une des problématiques liées à cette méthode

de ciblage réside dans l'isolation du fragment une fois digéré. En effet il se retrouve dans des proportions équimolaires avec le reste du génome, et représente ainsi une très faible proportion de l'échantillon qu'il est difficile de récupérer. Il apparait alors un besoin pour des technologies ultrasensibles afin de pouvoir détecter et isoler le grand fragment d'intérêt au milieu de l'ADN génomique. La technologie  $\mu$ LAS nous semblant bien adaptée à cette problématique, nous avons entrepris une collaboration en ce sens avec le Centre National de Ressources Génomique Végétales (CNRGV). L'objectif du projet est d'isoler spécifiquement un grand fragment d'ADN en fonction de sa séquence puis de le séquencer avec des technologies de 3<sup>ème</sup> génération.

L'étude suivante, présentée sous la forme d'un article et en cours de soumission chez Nucleic Acid Research, présente les résultats de cette collaboration. J'attire l'attention du lecteur sur le fait que la version de l'article présentée dans ce manuscrit contient des données de séquençage d'un échantillon ayant été contaminé et dont nous ne comprenons pas l'origine. Ces données ne sont pas présentées dans la version finale de l'article mais elles nous semblaient pertinentes dans le cadre de cette thèse.

## **Technology for DNA isolation coupled to Cas9-assisted targeting for sequencing and assembly of a 30 kb region in plant genome**

Nicolas MILON<sup>1, 2\*</sup>, Céline CHANTRY-DARMON<sup>3\*</sup>, Carine SATGE<sup>3</sup>, Margaux-Alison FUSTIER<sup>3</sup>, Stéphane CAUET<sup>3</sup>, Sandra MOREAU<sup>5</sup>, Caroline CALLOT<sup>3</sup>, Arnaud BELLEC<sup>3</sup>, William MARANDE<sup>3</sup>, Tslil GABRIELI<sup>4</sup>, Laure SAÏAS<sup>2</sup>, Audrey BOUTONNET<sup>2</sup>, Frédéric GINOT<sup>2</sup>, Hélène BERGES<sup>3</sup> and Aurélien BANCAUD<sup>1</sup>

<sup>1</sup>*CNRS, LAAS, 7 Avenue du Colonel Roche, F-31400, Toulouse, France.*

<sup>2</sup>*Picometrics, 478 Rue de la Découverte, 31670 Labège, France*

<sup>3</sup>*French Plant Genomic Resource Center, INRA - CNRGV, 24 Chemin de Borde Rouge – Auzeville, CS 52627, 31326 Castanet Tolosan Cedex, France [helene.berges@toulouse.inra.fr](mailto:helene.berges@toulouse.inra.fr)*

<sup>4</sup>*Laboratory of Plant-Microbe Interactions, INRA - LIPM, 24 Chemin de Borde Rouge – Auzeville, CS 52627, 31326 Castanet Tolosan Cedex, France*

<sup>5</sup>*School of Chemistry, Center of Nanoscience and Nanotechnology, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel*

\*These authors equally contributed to this work.

Correspondence: [abancaud@laas.fr](mailto:abancaud@laas.fr), [helene.berges@inra.fr](mailto:helene.berges@inra.fr)

Keywords: DNA separation, DNA concentration, DNA isolation, CRISPR/Cas9

## Abstract

Cas9 targeting in complex genomes has been viewed as unique tool to obtain high-quality and continuous sequence data. So far, however, size selection of target loci has been essentially carried out with pulse field gel electrophoresis (PFGE), but the quality of the resulting DNA was insufficient to perform *de novo* sequencing. Here, we develop the  $\mu$ LAS technology to capture and purify high molecular weight DNA in the size range of 3 to 200 kb. First, the technology is optimized to perform high sensitivity DNA profiling and detection in order to monitor the presence and control the quality of residual DNA fragments in complex genomic mixtures. We notably report a lower limit of detection of 20 fg/ $\mu$ L for 50 kb fragments obtained with an analytical time of 50 minutes. Furthermore,  $\mu$ LAS is operated to isolate a 31.5 kb locus cleaved by Cas9 from the plant *Medicago truncatula*. After a validation with a model system based on a Bacterial Artificial Chromosome (BAC) clone of 97 kb, the technology is operated on whole genome following different purification methods involving or not PFGE. Based on PacBio sequencing, we report optimal enrichment factors of 11.9 and 892 in BAC clone or whole genome, respectively, and obtain one contig of 31,009 and 29,441 bp with 99% sequence identity to the reference sequence showing that the quality of DNA purified with  $\mu$ LAS is compatible with sequencing and assembly.

## Introduction

The emergence of third-generation sequencing technologies with long-read capabilities has raised great interest to speed up the production of quality sequencing data (Huddleston et al., 2014, 2014; Pecrix et al., 2018). However, *de novo* assembly of eukaryotic genomes remains a feat due to their complexity in terms of size, polyploidy or high percentage of repetitive elements (Schatz et al., 2012), as well as the requirement for long and high-quality DNA molecules (Mayjonade et al., 2016). In situations where whole genome sequencing is not necessarily required, sequence capture methods appear to offer a relevant solution to obtain quality sequence data in large genotype panels with advantages with respect to time to results and computational cost. This statement is particularly true to for crop improvement, in which a key challenge consists in linking a phenotype to a genomic region (Bevan et al., 2017).

For sequences of less than 10 kb, specific enrichment is mostly achieved using PCR amplification or on-array capture (Mamanova et al., 2010). Yet, the typical size of a gene in the human genome is 10 to 15 kb (Strachan and Read, 1996), requiring the use of multiple amplification reactions that can only be performed in documented genomic regions. For longer tracks with a high degree of polymorphism (Kidd et al., 2008), the multifunction CRISPR/Cas9 technology (Jinek et al., 2012; Lander, 2016; Tian et al., 2017) has been viewed as a unique opportunity to overcome some limitations of current methods, which require whole-genome clone libraries and haplotype-specific iterative sequencing (Bellec et al., 2016; Jain et al., 2018). This approach, called CATCH (Cas9-Assisted Targeting of CHromosome segments) or CISMIR (CRISPR-mediated Isolation of Specific Megabase-size Regions of the genome), has been implemented to isolate a gene target of 100 kb in bacteria (Jiang et al., 2015; Jiang and Zhu, 2016), as well as a 200 kb target containing BRCA1 gene in the human genome (Gabrieli et al., 2018) or a 2.3 Mb region in mouse (Bennett-Baker and Mueller, 2017). After Cas9-directed restriction, the target region is isolated according to its size using pulsed field gel electrophoresis (PFGE), band excision and agarose digestion. Direct sequencing of the purified DNA fragment, which was characterized by an enrichment of ~200-fold, enabled mapping and exploration of structural diversity but not the assembly of the fragment sequence.

We recently introduced the DNA processing technology  $\mu$ LAS (Ranchon et al., 2016), which allows us to perform the operations of separation and concentration by tuning electro-hydrodynamic actuation in viscoelastic (VE) liquids (Chami et al., 2018). This technology reached good detection performances down to 10 fg/ $\mu$ L for DNA molecules in the size range 100 to 1,500 bp (Andriamanampisoa et al., 2018). Because  $\mu$ LAS is operated in continuous flow, we reasoned that it was well-suited for DNA purification and more specifically the collection of DNA of target fragments using a size selection criterion defined by CRISPR/Cas9 restriction. To prove this assertion, we first establish the settings to separate, concentrate, and detect DNA molecules of molecular weight (MW) spanning 3 to 200 kb. We report a limit of detection of 20 fg/ $\mu$ L for fragments of 50 kb, equivalently 360 molecules/ $\mu$ L, for an analytical time of 1 hour. We also demonstrate the isolation, sequencing and assembly of a locus of 31.5 kb extracted from a BAC and

from genomic DNA (gDNA) of the plant *Medicago truncatula* (Pecrix et al., 2018; Young et al., 2011). Altogether, this study establishes the relevance of  $\mu$ LAS technology for high molecular weight (high MW) DNA isolation and suggests its potential for targeted sequencing in complex genomes.

## Material and Methods

**Chemical and Instrumentation.** Molecular biology grades chemicals were purchased from Sigma-Aldrich (France). The buffer of the experiments was TBE 1X (Tris-HCl 89 mM, boric acid 89 mM, EDTA 1 mM, pH = 8.3) supplemented with the different PVP polymers, purchased from Sigma-Aldrich, BASF, or custom synthesized by the IMRCP laboratory. Quick-Load® 1 kb Extend DNA Ladder (# N3239S) was purchased from New England Biolabs and used at 1 ng/ $\mu$ L in separation experiments by  $\mu$ LAS. The Lambda Ladder PFG Marker (New England Biolabs) was used for PFGE experiments. The intercalating dye (Picometrics Technologies, France, # 16-BB-DNA1K/01) was used at a 2X final concentration in the buffer solution. All the solutions were filtered at 0.22  $\mu$ m.

Measurements and sample purification were performed with an Agilent 1600CE system (Germany) equipped with a Zetalif LED 480 nm detector (Picometrics Technologies, France). Fused silica capillaries were purchased from Polymicro Technologies (Phoenix, US). Capillary assemblies were performed as previously described (Andriamanampisoa et al., 2018). Electrokinetic injection was used at 100 V/cm for 10 s for separation experiments. Hydrodynamic injection at 50 mBars during 180 s was used for concentration and collection experiments.

**Quantification of DNA size, concentration and limit of detection.** Quantification and sizing of the samples with  $\mu$ LAS technology were performed using the kb extend DNA ladder as reference and linear interpolation schemes, as described in (Andriamanampisoa et al., 2018). Noise was estimated with Open LAB CDS (Agilent Technologies) by averaging over a lag of 1 min at the beginning and the end of the experiment. The limit of detection (LOD) was calculated by extrapolating the DNA concentration corresponding to a signal to noise ratio (SNR) of three from the four dilutions.

**BAC and genomic DNA agarose plug production.** The Mth2-17015 (CR931730) *Medicago truncatula* BAC clone containing the 31 kb region was isolated from the Mtr-B-mth2 *Medicago truncatula* BAC library available at the CNRGV ([https://cnrgv.toulouse.inra.fr/library/genomic\\_resource/Mtr-B-%20mth2](https://cnrgv.toulouse.inra.fr/library/genomic_resource/Mtr-B-%20mth2)). The BAC clone was extracted with the NucleoBond® Xtrac Midi Plus kit (Macherey-Nagel) following the manufacturer's instructions using 100 mL of LB media with a chloramphenicol selection at 12.5  $\mu$ g/ml. Agarose plugs with BAC DNA were produced by mixing 40  $\mu$ L of an agarose solution InCert® (Lonza) at 1.5% with 40  $\mu$ L of BAC DNA at 50 ng/ $\mu$ L during 10 min at 45°C. The plugs were then casted in CHEF Mapper® XA System 50-Well Plug Mold (BioRad) placed on ice and incubated at 4°C for 30 minutes. The plugs were stored in a buffer containing 10 mM Tris-HCl (pH=7.4) and EDTA 1 mM (TE) at 4°C.

Agarose plugs with genomic DNA were produced from 40 g of fresh A17 *Medicago truncatula* leaves at the three-leaf seedling stage pretreated in the dark for 2 days, using the first step of the protocol published in Zhang and collaborators (Zhang et al., 2012). Four centrifugations at 2,400 g were performed to wash



nuclei pellets, and the resulting material was dissolved in 40 agarose plugs of 80  $\mu$ L at an agarose concentration of 0.75%. gDNA was extracted in-plug with a lysis buffer, washed and treated with PMSF, and the resulting plugs were stored in TE at 4°C.

**Cas9 digestion.** The sgRNA were designed using the Crispor V4.3. software (Haeussler *et al.*, 2016; <http://crispor.tefor.net/>), their sequence being GTTTGTTAATTCTCGACCGG(AGG) and CTGCTCGACAATTTACCAG(AGG) in Medtr5g005120.1 and Medtr5g005050.2 (the 3 nucleotides in parenthesis correspond to the PAM sequence). The RNA guides provided by Integrated DNA Technologies were rehydrated and complexed using IDT Nuclease-Free Duplex Buffer following the supplier recommendations at 300 ng/ $\mu$ L. Before starting the Cas9 digestion reaction, the plugs were washed in fresh TE during one hour at room temperature with agitation at 450 rpm, as recommended in (Jiang and Zhu, 2016). Each plug was then cut into 32 pieces and preincubated in 2 mL of 1X Cas9 nuclease reaction buffer (New England Biolabs) at room temperature with agitation at 450 rpm during 30 min. In parallel, 12  $\mu$ g of each sgRNA was preincubated with 40 pmol of EnGen Cas9 NLS (20  $\mu$ M, New England Biolabs) in 1X Cas9 reaction buffer at 25°C during 30 min. The two sgRNA complexes were finally incubated with gDNA in the plugs during 2 hours at 37°C. To stop Cas9 digestion, 240  $\mu$ g of Proteinase K (20 mg/ml) was added and incubated during 3 hours at 43°C. The plugs were eventually rinsed in 1X TAE by overnight incubation at 4°C.

#### **PFGE, electroelution and FEMTO Pulse® analysis.**

PFGE was used either for removing small DNA from the gDNA plugs (1), quality control (2) and separation of the fragment of interest (3). (1) Small and medium MW DNA were removed from the plugs by running a short PFGE cycle of 1 hour in a 0.25X TBE agarose gel at 12°C with an electric field of 9 V/cm, 120° tilt angle, 0.5 s initial switch time, 1.5 s final switch time, linear ramping (CHEF Mapper® XA Pulsed Field Electrophoresis System (BioRad)). The plugs were eventually removed from the PFGE stage and stored in TE at 4°C. (2) Quality control of high MW DNA plugs was carried out by running a long PFGE cycle of 24 hours. One half of a plug was placed on a 0.25X agarose gel and the electric field was set to 6 V/cm, 120° tilt angle, pulse time 60 s to 120 s, linear rampling at 12°C (Supplementary Fig. S1). In order to evaluate DNA concentration in the plugs, one half of a plug was melted at 70°C for 2 minutes and digested by 1 unit of Agarase (Thermo Fisher Scientific) during 45 min at 43°C. DNA concentration was determined with the Qubit™ dsDNA BR Assay Kit and the associated Qubit 3.0 Fluorometer (Invitrogen) after 10 minutes of sonication. (3) To separate the fragment of interest from the rest of gDNA, a 1% agarose 0.25X TBE PFGE gel was casted with a low melting window following Peterson and collaborators protocol (Peterson *et al.*, 2000). The NEB Lambda Ladder PFG Marker and the BAC DNA digested by Cas9 were used as size makers. The Cas9 digested plug pieces were run for 16 hours with the following conditions: 6 V/cm, 120° tilt angle, pulse time 5 s to 15 s, linear rampling at 12°C. A band corresponding at the 30kb fragment size was cut from the gel. The DNA was eluted in about 100

$\mu$ L of TAE 1X from the plug pieces using the BioRad 422 Electro-Elutor for 2 hours following the manufacturing recommendations. The DNA fragments were analyzed on the FEMTO *Pulse*® Automated Pulsed-Field CE Instrument (Advanced Analytical Technologies) with the gDNA 165kb methods.

**PacBio sequencing.** The 31 kb fragments isolated from BAC or plant gDNA were amplified using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) starting from 2  $\mu$ L of material. The amplified fragments were sequenced in a tagged library generated using the Pacific Biosciences library preparation protocol for 10 kb libraries. It was sequenced in one SMRT cell using the P4 polymerase in combination with the C2 chemistry. The work was conducted following the standard operating procedures of the manufacturer and the sequencing was performed on a PacBio RS II at the INRA sequencing platform GeT-PlaGe in Toulouse, France (<http://get.genotoul.fr/la-plateforme/get-plage/>) or the NGI Platform in Uppsala, Sweden (<http://snpsseq.medsci.uu.se/about-us/ngi/>).

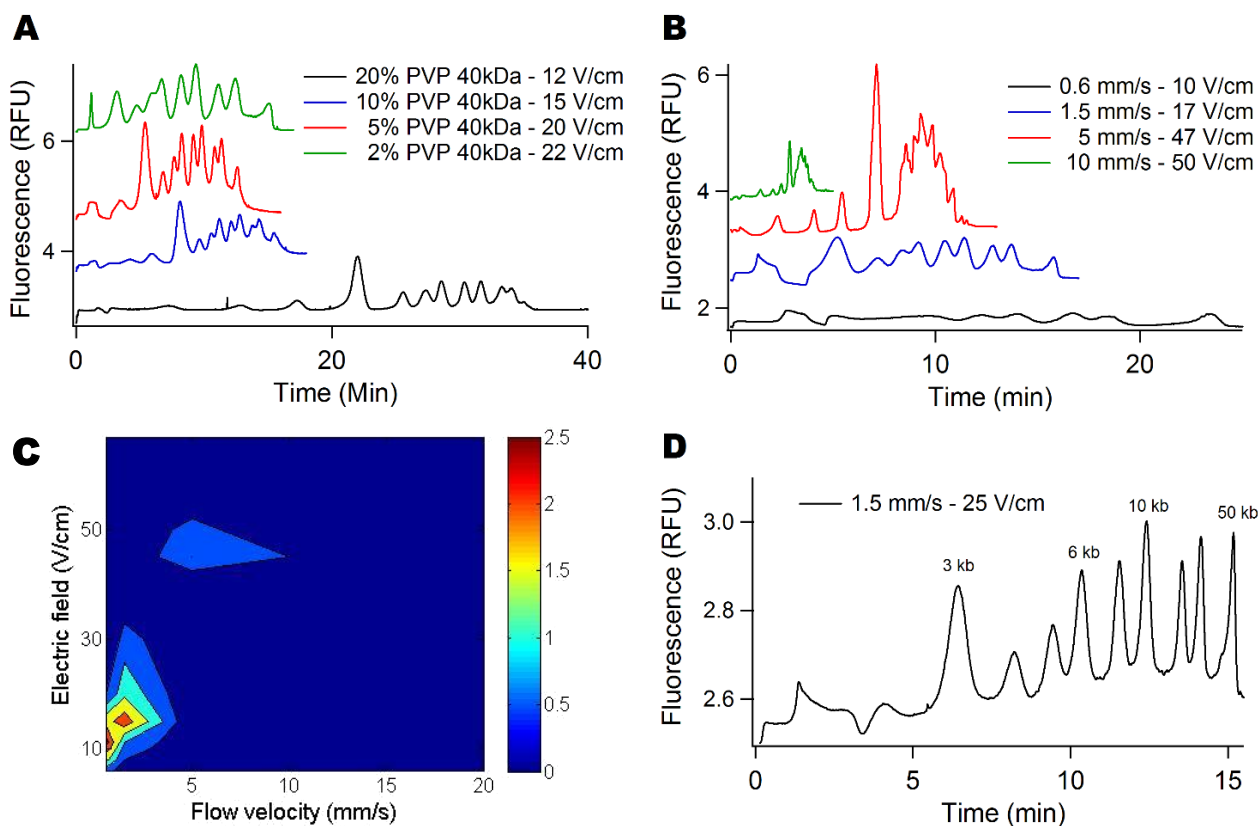
**Bioinformatics analysis.** The data were demultiplexed and filtered, low quality reads (read quality <0.75 and read length <500 bp) were removed. As WGA generates chimeric reads, filtered reads were then mapped against their corresponding reference -BAC clone (GenBank CR931730.1) for the BAC sample, and the sequence of the v4.0 *Medicago truncatula* genome for gDNA samples- with minimap2 once with option -cx map-pb --secondary=n to generate PAF format output and once with the option -ax map-pb --secondary=n to generate SAM format output (Li, 2018). This step allows the splitting of chimeric reads, and non-chimeric corresponding fastq sequences were generated using homemade python scripts. Those sequences were then assembled and corrected with the pipeline of the CANU assembler (Koren *et al.*, 2017). Using samtools-1.8 (Li *et al.*, 2009), the SAM files were sorted and indexed with function *sorted* and *index*, then depth coverage of the target region and the rest of the matrix were estimated with the function *bedcov*. Enrichment was calculating as the ratio of the depth of the targeted region over the depth of the other part of the genome. Coverage along the genome was calculated using the function *depth* and plots coverage were generated using R-3.4.3 (R Core Team 2017, <https://www.r-project.org/>).

## Results

### Viscoelastic fluid formulation for high MW DNA separation by $\mu$ LAS

DNA separation and concentration by  $\mu$ LAS is dictated by transverse migration in VE liquids (Chami *et al.*, 2018). For DNA molecules of less than 1 kb (Andriamanampisoa *et al.*, 2018), we used a high MW formulation of poly-vinylpyrrolidone (PVP) and a volumic concentration of 5%, which was associated to a viscosity of 30 mPa.s and an elasticity of 0.7 Pa (Chami *et al.*, 2018). The isolation of longer fragments of ~30 kb requires the definition of an appropriate VE fluid formulation as well as actuation conditions. Eight different PVP polymers with average molar mass ( $M_n$ ) spanning 10 to 360 kDa have been purchased or custom-synthesized ((Guinaudeau *et al.*, 2012), Table 1). For each PVP, we carried out

series of separation experiments with a ladder containing 13 bands in the range 0.5 to 50 kb (see Methods) using the resolution  $R_{50}$  between the bands of 20 and 50 kb as readout.



**Figure 1: High MW DNA separation by  $\mu$ LAS.** **A:** The four curves correspond to chromatograms obtained with the kb extend DNA ladder using four different concentrations of PVP 40 kDa, as indicated in inset. The flow velocity  $v_m$  is set to 1.5 mm/s. **B:** Using 5% PVP 40 kDa, the graph presents the separation of the kb extend DNA ladder for four different flow velocities  $v_m$  ranging from 0.6 to 10 mm/s as indicated in inset. **C:** The contour plot shows the resolution  $R_{50}$  as a function of the electric field and flow velocity with a buffer containing 5% PVP 40 kDa. **D:** Chromatogram of the kb extend DNA ladder with the custom-synthesized PVP 43 kDa dissolved at a volumic concentration of 3%.

We first selected a reference polymer with an average MW of 40 kDa in our library (Table 1), and optimized high MW DNA separation by tuning its concentration, the flow velocity, and the electric field. PVP 40 kDa was thus dissolved in a buffer for electrophoresis at four different concentrations of 2 to 20%, which corresponded to viscosities of 1.4 to 18.6 mPa.s, respectively. The chromatograms shown in Fig. 1A were obtained for mean flow velocity  $v_m$  set to 1.5 mm/s, and electric fields adjusted to optimize  $R_{50}$  as reported in Supplementary Fig. S2A. We noticed that  $R_{50}$  increased from 0.7 to 2.3 as the concentration of PVP decreased from 20% to 5%, and then decreased to 1.7 for a concentration of 2%. Notably,  $v_m$  was set to 1.5 mm/s because we obtained an optimal  $R_{50}$  of 2.3 for a minimal separation time of 15 min for flow velocity spanning 0.6 to 20 mm/s and electric field in the range 10 to 50 V/cm (Fig. 1B-C). This first set of experiments therefore showed us that high MW DNA separation required a formulation of PVP with a

viscosity 15-fold lower than that for low MW DNA separation. In addition, the flow velocity and the electric field were decreased by a factor of 10 in comparison to the conditions to separate molecules of 0.1-1 kb (Andriamanampisoa et al., 2018).

We then compared the different PVP polymers by dissolving them at a concentration that corresponded to a viscosity of 2.6 mPa.s, *i.e.* as for 5% PVP 40 kDa, as reported in Table 1. The flow velocity was set to 1.5 mm/s. Using high MW PVP of 360 kDa or k90 (Table 1), the resolution  $R_{50}$  was lower than 1.3, showing that long PVP chains are not adequate for high MW DNA separation. Optimal separation performances with an optimum  $R_{50}$  of 3.0 were achieved with custom-made PVP for 43 kDa (Fig. 1D). The separation was performed in ~15 minutes, and the theoretical plate number was  $\sim 4.10^5$  plate/m for the 50 kb peak. This result compares well to the best reported value of  $10^6$  plates/m obtained with nanofabricated pillar matrices (Kaji et al., 2004). It is much greater than  $10^3$  plate/m achieved with matrix-free separation based on cyclic electro-hydrodynamic actuation (Zheng and Yeung, 2003). Further, using this optimal PVP formulation, we could separate a sample containing four bands of 50, 100, 150, and 210 kb in less than 1 hour (Supplementary Fig. S2B). This time to result is faster than the 4 hours required by PFGE (Birren and Lai, 1994). Altogether, this study shows that  $\mu$ LAS is relevant to perform DNA separation in the size range 10 to 200 kb in less than an hour.

**Table 1 : Characteristics of PVP polymers.** Eight different PVP formulations are compared according to their number average molecular weight  $M_n$ , average molecular weight  $M_w$ , mass:volume concentration corresponding to a viscosity of 2.6 mPa.s, and separation performance based on resolution between the 20 and 50 kb fragments. Structural properties are inferred from supplier information or from chromatography, as described in (Guinaudeau et al., 2012). The Sigma 40 kDa marked with asterisks corresponds to the reference formulation for our study. (†)

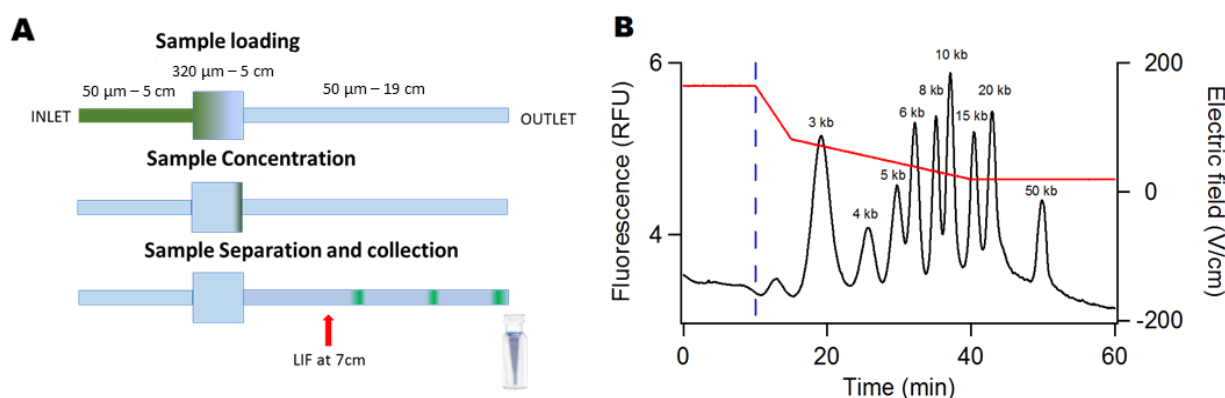
Obtained from supplier information.

PVP	Sigma 10 kDa	Custom 12.5 kDa	BASF k30 (†)	Custom 30 kDa	Sigma 40 kDa (**)	Custom 43 kDa	BASF k90 (†)	Sigma 360 kDa
$M_n$ (kDa)	9.3	12.5	14	30.3	26.5	43.6	325	427
$M_w$ (kDa)	16.4	13.5	50	41.1	59.3	66.2	1400	740
Concentration for a viscosity of 2.6 mPa.s	10%	8%	5%	4%	5%	3%	2.5%	0.7%
Resolution $R_{50}$	1.5	2.8	1.5	2.1	2.3	3.0	1.3	1.0

### High MW DNA concentration and determination of the limit of detection

The  $\mu$ LAS technology allows us to concentrate DNA molecules of 0.1 to 1 kb at the junction of two capillaries of different diameters ((Andriamanampisoa et al., 2018), Fig. 2A). For this study, we engineered a capillary system composed of a large and a narrow capillary of 320  $\mu$ m and 50  $\mu$ m in inner diameter, respectively (see details in Methods). The former contains the injected volume of 4.3  $\mu$ L, *i.e.* 1000 times more than that of conventional CE, and the latter is dedicated to size separation. In this capillary system,

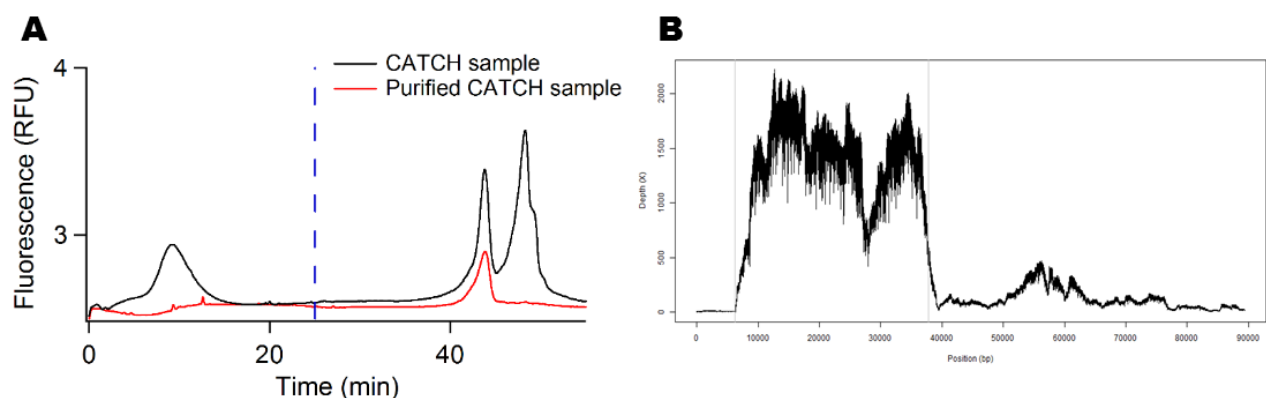
concentration operations are operated by delivering a high electric field, which impedes the migration of DNA molecules in the separation channel and forces DNA enrichment at the constriction. The electric field is then progressively reduced in order to allow the migration of the DNA species in the separation channel. A typical concentration and separation experiment is reported in Fig. 2B. The first concentration phase lasts 10 min with a flow velocity set to 1.5 mm/s and a high electric field of 165 V/cm (red curve in Fig. 2B). During the next 30 minutes, the electric field is gradually reduced to 25 V/cm, *i.e.* the separation conditions defined in the previous paragraph. Notably, the progressive decrease of the electric field allows us to maintain good separation performances after the concentration phase (not shown). The overall analytical process lasts  $\sim$ 50 min, and the resolution  $R_{50}$  remains equal to 3 with this two-step process (black curve in Fig. 2B). The limit of detection (LOD), as measured from the analysis of serial dilutions of the kb extend DNA ladder (Supplementary Table S1), is 13 and 20 fg/ $\mu$ L for the bands of 10 and 50 kb, respectively. We estimate the gain in sensitivity of this concentration and separation process in comparison to a simple separation to be 1000-fold (not shown). Furthermore, the LOD for a fragment of 50 kb is 600 aM, *i.e.* two orders of magnitude lower than that reached by conventional sample stacking (Osborn et al., 2000). This LOD is also 15-fold enhanced in comparison to recent commercial systems, such as the FEMTO Pulse<sup>®</sup>, but the sensitivity of recent microfluidic technologies based on molecular rheotaxis of 150 aM is 4 times greater (Friedrich et al., 2017). Finally, we performed multiple analysis of the ladder sample in order to estimate the sizing and the quantification precision and accuracy, which were 3% and 13%, respectively for a 50 kb fragment (Supplementary Table S2 and S3).



**Figure 2:  $\mu$ LAS for online concentration and separation.** **A:** Schematics of a capillary system composed of (i) an inlet capillary of 50  $\mu$ m in inner diameter and 5 cm in length, (ii) a loading chamber of 320  $\mu$ m in inner diameter and 5 cm in length, and (iii) an outlet capillary of 50  $\mu$ m in inner diameter and 19 cm in length. The detector is placed 7 cm downstream of the interconnection between the loading chamber and the outlet capillary. Collection is performed by isolating a target volume at the outlet during a calibrated time period. **B:** The black curve shows the intensity vs. time for the kb extend ladder at a concentration of 100 pg/ $\mu$ L and the red curve is the temporal evolution of the electric field. The sample is concentrated during the first ten minutes of the process (dashed blue line).

### Purification of a 30 kb fragment

We then decided to exploit the separation performance of  $\mu$ LAS between the bands of 20 and 50 kb to collect and purify a fragment of 30 kb. This size range is relevant to genetic analysis of most eukaryotic genes; for example, it is twice longer than the average size of human genes of 10 to 15 kb (Strachan and Read, 1996) despite a broad variability from  $\sim$ 0.2 to  $\sim$ 2,500 kb. As a proof of concept, we used a 100 kb BAC clone from *Medicago truncatula* (see Methods), in which we identified two sgRNA for Cas9 restriction to produce molecules of 31,512 bp and 65,337 bp, hereafter referred to as the 31 and 65 kb fragments. Our process of purification and cleavage was carried out by embedding 1.5  $\mu$ g of BAC DNA in agarose plugs in order to emulate standard protocols for gDNA purification (see Methods). Quality control of the digestion reaction was carried out by PFGE and  $\mu$ LAS, showing three bands around 100, 70, and 30 kb corresponding to the linearized BAC and the two cleaved fragments (Supplementary Fig. S3A-B). The yield of the cutting reaction was 93% according to profiling and titration with the FEMTO Pulse  $\text{\textcircled{R}}$  system (Supplementary Fig. S3C and Table S4). Finally, the digested BAC was electroeluted from the agarose laden, resulting in 106  $\mu$ L of DNA at a concentration of 25 ng/ $\mu$ L, as inferred from Qubit Fluorometry. The amount of DNA after electroelution of 2.6  $\mu$ g was nearly twice that of the input material, suggesting that sgRNA contribute to  $\sim$ 50% of the fluorescence signal. The analysis of the electroeluted sample with  $\mu$ LAS confirms this hypothesis (black curve in Fig. 3A), because the detection of a strong and broad fluorescence peak during the concentration phase is very likely associated to the passage of sgRNA, which are too small to be retained at the junction. Furthermore, the two peaks of 31 and 65 kb are well resolved together with a high MW residue that corresponds to the linearized BAC. The concentration of the 31 kb fragment is  $3 \pm 0.4$  ng/ $\mu$ L based on our calibration with the kb extend DNA ladder (Supplementary Fig. S3B), meaning that the yield of the collection process is  $70 \pm 9\%$ .



**Figure 3: Capture of a 31 kb fragment from a BAC.** **A:** The chromatograms show the Cas9-digested BAC sample (black curve) with two main peaks of 31 and 65 kb, and the purified sample with one single peak of 31 kb (red curve). The dashed blue line corresponds to the end of the concentration step. **B:** The graph presents the read depth as a function of the physical position on the BAC clone. The vertical gray lines represent the position of the sgRNA.

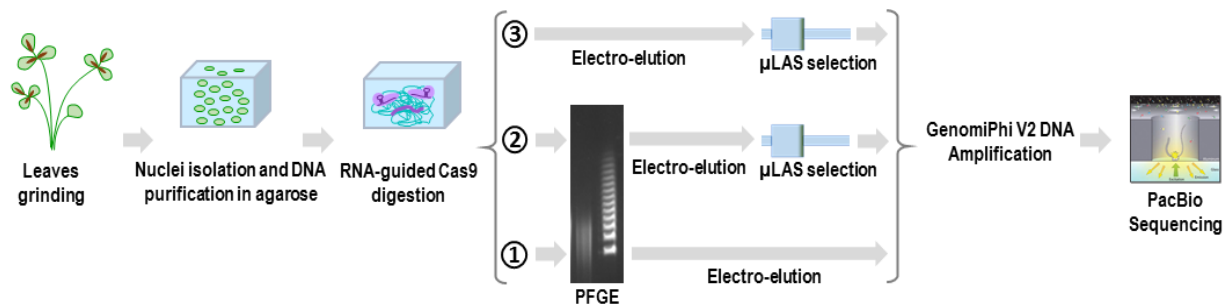
Sample purification was achieved by monitoring the passage of the sample at the outlet of the capillary and defining a temporal window for switching between two collection tubes of the carousel of the capillary electrophoresis system (Fig. 2A). We performed this operation by injecting  $\sim 125$  pg of Cas9 digested DNA in a volume of 2  $\mu$ L. Note that the sample has been diluted in order to avoid saturation and uncontrolled DNA leak during the concentration phase (manuscript in preparation). The calibration of the size selection window, which has been obtained by measuring the size distribution of a sample extracted from a broad smear pattern (Supplementary Fig. S4), indicates that  $\mu$ LAS purification selects molecules in the range 25 to 35 kb. In order to obtain a total volume of 25  $\mu$ L compatible with downstream molecular biology operations, in particular whole DNA amplification, sample collection procedure has been repeated 10 times consecutively. Fig. 3A reports the characterization of the resulting sample with  $\mu$ LAS, showing the discard of sgRNA and most of the 65 kb fragments. Specifically, a residual intensity signal in the 65 kb region is recorded, corresponding to  $\sim 9 \pm 1$  % of the total signal, meaning that the purification rate is  $\sim 92 \pm 2$  %. The final product concentration is  $4 \pm 0.5$  pg/ $\mu$ L, equivalently  $100 \pm 13$  pg are recovered, and the yield of the DNA isolation process is  $67 \pm 9$ %.

The purified 31 kb fragment was then amplified using whole DNA amplification in order to produce at least 8  $\mu$ g for PacBio RSII long read sequencing. 6,124 reads have then been obtained, among them 5,856 reads mapped against the published sequence of the Mth2-17015 BAC according to Minimap2 (Li, 2018), Table 2). The average read depth was 1525X for the 31 kb region of interest and of 129X for the rest of the BAC (Fig. 3B). We calculated the enrichment factor by comparing the depth of coverage in the region of interest against the rest of the BAC sequence and inferred an enrichment factor of 11.9. Using the CANU assembler (Koren et al., 2017), we obtained a consensus sequence with 1 contig of 31,009 bp (Supplementary Data Sequence “Contig\_CATCH\_30kb\_BAC”) representing 98% of query coverage with 99% of identity and 47 gaps with the reference sequence using the Blast 2 sequences online software (Altschul *et al.*, 1990). In figure 3B, the alignment of the sequences against the BAC reference sequence (CR931730.1) shows a small pic in the region corresponding to the rest of the BAC, which is in good agreement with the residual signal observed in Figure 3A. However, the coverage in the region of interest is  $\sim 12$  time greater and the average read depth of 1525X allows us to obtain an assembly of the sequence of interest in a unique contig.

### **Capture of a fragment of interest from plant genomic DNA**

We finally focused our study on the capture of the same 31 kb fragment in the 450 Mb genome of *Medicago truncatula*, (Pecrix Y. *et al.* 2018). The fragment of interest represents a fraction of  $7.10^{-5}$  of the genome. Three different extraction methods have been compared (Fig. 4). As a reference and for comparison to recent studies (Bennett-Baker and Mueller, 2017; Gabrieli et al., 2018), DNA sequencing has been performed on a sample obtained by band excision after PFGE and electro-elution (method 1, lower arrow in Fig. 4). The same electro-eluted sample has been processed by  $\mu$ LAS to evaluate the benefit

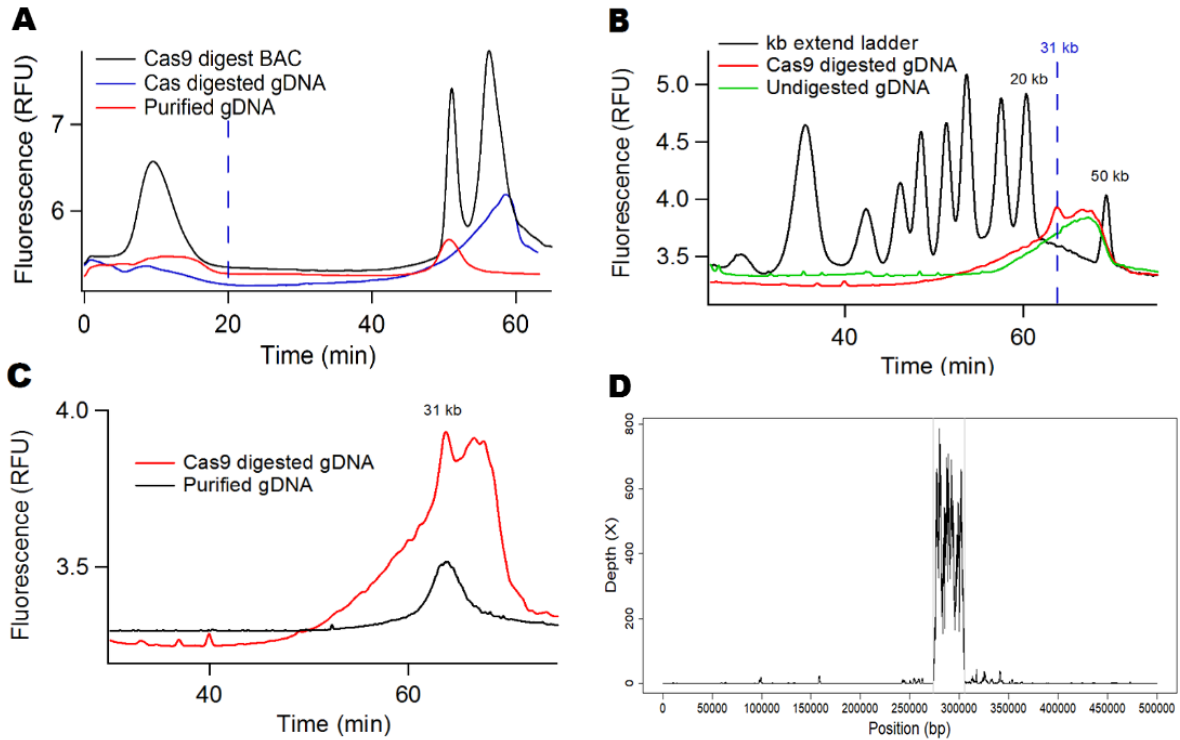
of this technology for sequencing (method 2, middle arrow in Fig. 4). Finally, method 3 has consisted in performing size selection by  $\mu$ LAS on the material electro-eluted from the plugs with no steps of size selection by PFGE (upper arrow in Fig.4).



**Figure 4: Workflow of the CATCH method for plant genomic DNA purification and sequencing:** The cartoons represent the consecutive steps for the isolation of a target region from plant genomes. We used three different strategies: 1-with PFGE alone; 2-with PFGE and  $\mu$ LAS isolation and 3-only using  $\mu$ LAS isolation.

For this study, twenty agarose plugs have been prepared, each containing  $\sim 10$   $\mu$ g of gDNA. Assuming the Cas9 digestion reaction to be complete, we expect the maximal mass of 31 kb fragment to be 0.7 ng per plug. For method 3, we first analyzed the size distribution of the electroeluted Cas9-digested gDNA by  $\mu$ LAS and observed a broad size pattern in the range 20 to more than 100 kb without any specific signature at 30 kb (blue curve in Fig. 5A). In addition to the fact that this profile was observed with or without Cas9 restriction, the measurement of the total mass of Cas9-digested gDNA after electroelution of 300 ng, *i.e.* 400 times more material than expected for the 31 kb fragment, indicated that electroelution was insufficiently selective to isolate the target locus. Conversely, when DNA was processed by PFGE and band excision (method 1 & 2), we only collected 10 ng of gDNA per plug after the electro-elution step (Supplementary Fig. S5A). Although this quantity still represents 15 times more than the 0.7 ng expected for the 31 kb locus, we detected the presence of the 30 kb fragment as a spike in a broad smear pattern in the range 10 to 60 kb (red curve in Fig. 5B). We further confirmed the presence of this 30 kb peak with FEMTO *Pulse*® analysis (Supplementary Fig S5B). Notably, we obtained the same smear pattern without the 30 kb spike for the negative control without Cas9 digestion, meaning that gDNA degradation occurs during gDNA extraction or in the course of PFGE (green curve in Fig. 5B). The 31 kb fragment represented  $\sim 2.5$  % of the total DNA mass in the smear pattern, meaning that we collected  $0.26 \pm 0.03$  ng of 31 kb fragment. The overall recovery yield was therefore  $37\% \pm 5\%$ .





**Figure 5: Capture of a 30 kb DNA fragment from genomic DNA.** **A:** The blue chromatogram corresponds to the Cas9-digested gDNA obtained according to method 3, without purification by PFGE. The red curve is the  $\mu$ LAS purification of the same sample using a selection window around 30 kb. The black curve is the Cas9-digested BAC which serves as reference. Note that the dashed blue curve to the end of the concentration step. **B:** The chromatograms represent the quality control of Cas9-digested gDNA obtained according to method 2, after PFGE, band excision around 30 kb and electro-elution. Black curve represents the kb extend ladder, serving as size and concentration reference. Red and green curves correspond to the Cas9-digested and undigested gDNA, respectively. **C:** The red chromatogram corresponds to the Cas9-digested gDNA obtained according to method 2. The black chromatogram represents the product purified with  $\mu$ LAS from the same sample. **D:** The graph presents the reads depth as a function of the physical position on the first 500 kb of the *Medicago truncatula* chromosome V sequence. The vertical blue lines represent the position of the sgRNA on the reference BAC sequence. The X-axis corresponds to the position in the sequence in bp and the Y-axis to the depth.

We then purified the 31 kb fragment using  $\mu$ LAS and a size selection window centered at 30 kb with a breadth of  $\{+/-\}$  5 kb (Supplementary Fig. S4), as described in method 2 and 3. We injected  $\sim 0.2$  ng of DNA in the capillary device, and fractionated the sample 10 times consecutively to obtain 25  $\mu$ L at 3 pg/ $\mu$ L or 30  $\mu$ L at 5 pg/ $\mu$ L for the method 2 and 3, respectively. We finally performed whole DNA amplification with the different reaction products, which were subsequently sequenced with the PacBio RS II technology.

Our control method 1 solely based on PFGE allowed us to collect 139,450 reads with 129,477 reads that mapped against the published genome of *Medicago truncatula* (Table 2). We obtained a coverage of 103X for the target fragment and of 1.2X for the matrix and inferred an enrichment factor of

84. This result is similar to recent reports, which were 75 (263 kb), 174 (610 kb) and 39 (2.3 Mb) in mouse for Bennett-Baker and Muller (Bennett-Baker and Mueller, 2017) and 237 (200 kb) in human for Gabrieli and collaborators (Gabrieli et al., 2018). A fastq file of sequences mapping in the target region was generated and assembled with CANU in two not overlapping contigs of 18,733 bp (Supplementary Data Sequence “Contig1\_CATCH\_30kb\_ADNg\_Method1”) and 7,760 bp (Supplementary Data Sequence “Contig2\_CATCH\_30kb\_ADNg\_Method1”). Further selection of the target genomic locus with  $\mu$ LAS after PFGE (method 2) led to 41,714 reads that were mapped against the published sequence of *Medicago truncatula* (Fig. 5C). We obtained a 448X coverage of the region of interest and a final enrichment factor of 892. A fastq file of sequences mapping in the target region was generated and assembled with CANU in 1 contig of 29,441 bp (Supplementary Data Sequence “Contig\_CATCH\_30kb\_ADNg\_Method2”) representing 93% of query coverage with 99% of identity and 11 gaps with the reference sequence using the Blast 2 sequences online software (Altschul *et al.*, 1990). The direct isolation of the target fragment after electroelution from the agarose plug lead to 99,028 reads but only 12% were mapped against the published genome of *Medicago truncatula* (Supplementary Fig. S6). We found that most reads came from human DNA although we cannot explain the origin of this contamination. By removing the unwanted reads, we obtained a coverage of the target region of 22X and an enrichment factor 227 that is similar to previous studies. Unfortunately, we obtained too few reads on the target region to have an assembly of the sequence.

**Table 2 :** The table presents sequencing data obtained with the three purification methods shown in Fig. 4 and for BAC DNA.

Method	# bases (bp)	# reads	N50	Mapped bases on target region	Cov target	Cov matrix	Enrichment
BAC 30 kb	59,356,468	6,124	9,692	80%	1,525	129	12
ADNg - Method 1	680,112,424	139,450	5,591	1%	103	1	84
ADNg - Method 2	325,378,430	45,210	8,265	5%	448	1	892
ADNg - Method 3	473,492,538	99,028	5,625	0.2%	22	0.1	227

Nonetheless,  $\mu$ LAS isolation following a PFGE increases the percent of bases that were mapped on the target region by a factor of 5 and the enrichment of the target sequence by a factor 10 compared to PFGE alone (Table 2). Notably, we obtained an enrichment factor of 892 higher than previous studies in eukaryote genome with the CATCH/CISMR technique (Bennett-Baker and Mueller, 2017; Gabrieli et al., 2018) and we report for the first time an assembly of a unique contig of the sequence of interest with 99% of identity with the reference sequence.

## Conclusion

We presented a simple and versatile technology to perform the main operations of nucleic acids processing, namely concentration, separation, and purification. We first optimized the formulation of the viscoelastic solution for high MW DNA separation for fragments of up to 200 kb. We then engineered the geometry of a capillary system to concentrate DNA samples and reached a limit of detection of  $\sim 10$  fg/ $\mu$ L for 50 kb fragments. Finally, taking advantage of the fact that  $\mu$ LAS technology is operated without separation matrix, we developed an assay to purify a fragment of interest of 30  $\{+/-\}$  5 kb. The technology is subsequently applied to collect a target fragment of 30 kb from *Medicago Truncatula* genomic DNA cleaved by Cas9. We report that the assembly of one contig of 29,441 bp, which represents 93% of the target sequence, can be obtained in an eukaryote genome by combining  $\mu$ LAS with PFGE.

Altogether, we argue that the  $\mu$ LAS technology presents unique performances for DNA detection and purification, which are for quality control of genomic DNA samples as well as for high MW sample collection. In addition to the recent efforts to improve the efficiency of Cas9 and reduce off-target nuclease activity (Zischewski et al., 2017), our methodological development appears to be timely to obtain high-quality sequencing data and assembly of large targeted regions of interest in plant and eukaryotic genomes.

## Acknowledgments

We are thankful to Mathias Destarac (Paul Sabatier University, Toulouse) for the synthesis of custom PVP. We thank Claudette Icher, Fabrice Devoilles and the TPMP platform for the growing of the plants. We are thankful to Kit-Sum Wong and William Amoyal from Advanced and Analytical who performed the FEMTO *Pulse*<sup>®</sup> analysis. We thank the platforms GeT-PlaGe and NGI for the Pacbio sequencing. This project is funded by two grants one from European Union and Occitanie Pyrénées-Méditerranée region and a second one from the Plant2Pro Carnot Institute.

## Bibliography

- Andriamanampisoa, C.-L., Bancaud, A., Boutonnet-Rodat, A., Didelot, A., Fabre, J., Fina, F., Garlan, F., Garrigou, S., Gaudy, C., and Ginot, F. (2018). BIABooster: Online DNA Concentration and Size Profiling with a Limit of Detection of 10 fg/ $\mu$ L and Application to High-Sensitivity Characterization of Circulating Cell-Free DNA. *Analytical Chemistry* 90, 3766–3774.
- Bellec, A., Courtial, A., Cauet, S., Rodde, N., and Vautrin, S. (2016). Long Read Sequencing Technology to Solve Complex Genomic Regions Assembly in Plants. *Next Generat Sequenc & Applic* 3, 10–4172.
- Bennett-Baker, P.E., and Mueller, J.L. (2017). CRISPR-mediated isolation of specific megabase segments of genomic DNA. *Nucleic Acids Research* 45, e165–e165.
- Bevan, M.W., Uauy, C., Wulff, B.B., Zhou, J., Krasileva, K., and Clark, M.D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346.
- Birren, B., and Lai, E. (1994). Rapid pulsed field separation of DNA molecules up to 250 kb. *Nucleic Acids Research* 22, 5366–5370.
- Chami, B., Socol, M., Manghi, M., and Bancaud, A. (2018). Modeling of DNA transport in viscoelastic electro-hydrodynamic flows for enhanced size separation. *Soft Matter* 14, 5069–5079.
- Friedrich, S.M., Burke, J.M., Liu, K.J., Ivory, C.F., and Wang, T.-H. (2017). Molecular rheotaxis directs DNA migration and concentration against a pressure-driven flow. *Nature Communications* 8, 1213.
- Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y., and Ebenstein, Y. (2018). Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Research* 46.
- Guinaudeau, A., Mazières, S., Wilson, D.J., and Destarac, M. (2012). Aqueous RAFT/MADIX polymerisation of N-vinyl pyrrolidone at ambient temperature. *Polymer Chemistry* 3, 81–84.
- Huddleston, J., Ranade, S., Malig, M., Antonacci, F., Chaisson, M., Hon, L., Sudmant, P.H., Graves, T.A., Alkan, C., and Dennis, M.Y. (2014). Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Research* gr-168450.
- Jain, M., Olsen, H.E., Turner, D.J., Stoddart, D., Bulazel, K.V., Paten, B., Haussler, D., Willard, H.F., Akeson, M., and Miga, K.H. (2018). Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology* 36, 321.
- Jiang, W., and Zhu, T.F. (2016). Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nature Protocols* 11, 960.
- Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., and Zhu, T.F. (2015). Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature Communications* 6, 8101.
- Jinek, M., Chylinski, K., Fonfara, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- Kaji, N., Tezuka, Y., Takamura, Y., Ueda, M., Nishimoto, T., Nakanishi, H., Horiike, Y., and Baba, Y. (2004). Separation of long DNA molecules by quartz nanopillar chips under a direct current electric field. *Analytical Chemistry* 76, 15–22.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., and Antonacci, F. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* gr-215087.

Lander, E.S. (2016). The heroes of CRISPR. *Cell* 164, 18–28.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 1, 7.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7, 111.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* 61, 203–205.

Osbourn, D.M., Weiss, D.J., and Lunte, C.E. (2000). On-line preconcentration methods for capillary electrophoresis. *ELECTROPHORESIS: An International Journal* 21, 2768–2779.

Pecrix, Y., Staton, S.E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinaud, M.-F., Latrasse, D., and Zouine, M. (2018). Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nature Plants* 1.

Peterson, D.G., Tomkins, J.P., Frisch, D.A., and Paterson, A.H. (2000). Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. *Journal of Agricultural Genomics* 5, 1–100.

Ranchon, H., Malbec, R., Picot, V., Boutonnet, A., Terrapanich, P., Joseph, P., Leïchlé, T., and Bancaud, A. (2016). DNA separation and enrichment using electro-hydrodynamic bidirectional flows in viscoelastic liquids. *Lab Chip* 16, 1243–1253.

Schatz, M.C., Witkowski, J., and McCombie, W.R. (2012). Current challenges in de novo plant genome sequencing and assembly. *Genome Biology* 13, 243.

Strachan, T., and Read, A.P. (1996). *Human molecular genetics* (BIOS Scientific Publishers).

Tian, P., Wang, J., Shen, X., Rey, J.F., Yuan, Q., and Yan, Y. (2017). Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synthetic and Systems Biotechnology* 2, 219–225.

Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., and Schoof, H. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520.

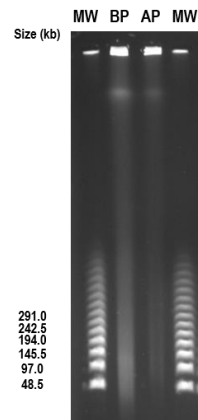
Zhang, M., Zhang, Y., Scheuring, C.F., Wu, C.-C., Dong, J.J., and Zhang, H.-B. (2012). Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nature Protocols* 7, 467.

Zheng, J., and Yeung, E.S. (2003). Mechanism for the separation of large molecules based on radial migration in capillary electrophoresis. *Anal. Chem.* 75, 3675.

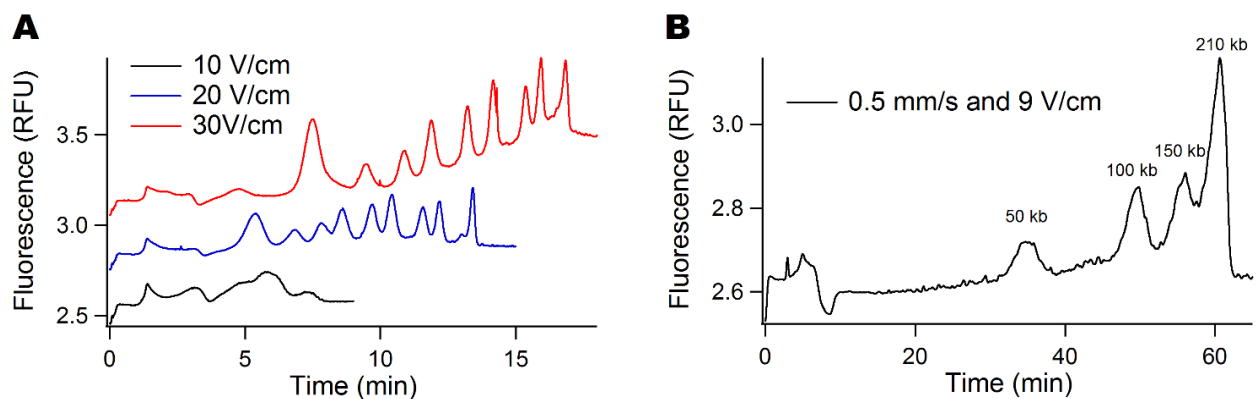
Zischewski, J., Fischer, R., and Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnology Advances* 35, 95–104.

## Technology for DNA isolation coupled to Cas9-assisted targeting for *de novo* sequencing of a 30 kb region in plant genome

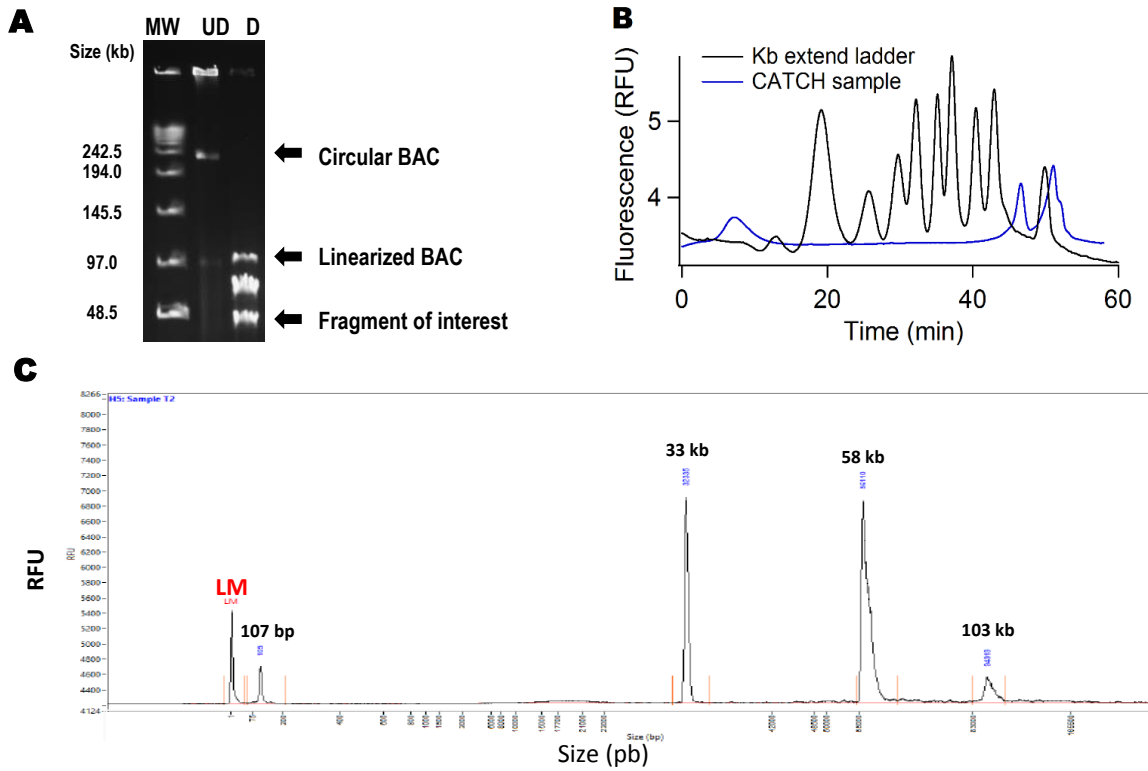
### Supplementary Material:



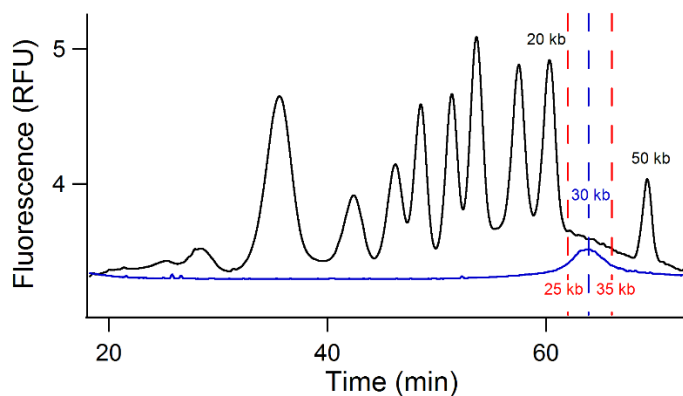
**Supplementary Figure S1:** *M. truncatula* genomic DNA analysis in agarose plugs before (BP) and after (AP) short PFGE purification and analyzed using long PFGE (see Methods section). MW: molecular weight reference with annotated band sizes on the left panel. One half of a gDNA plug is loaded in each well.



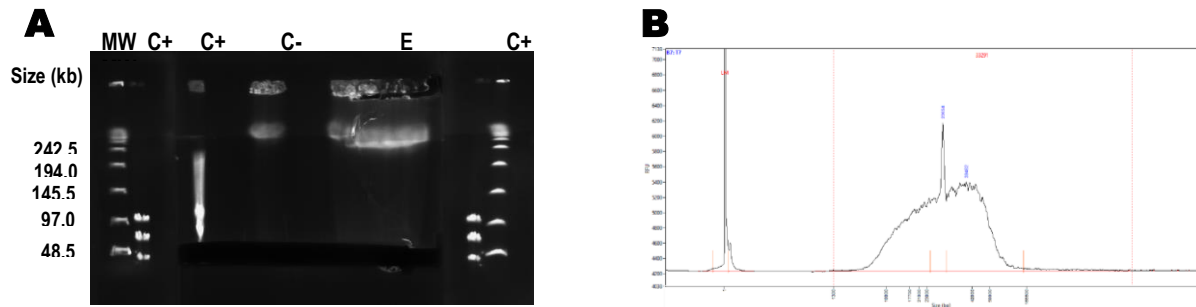
**Supplementary Figure S2:** **A:** The graph presents the separation of the kb extend DNA ladder for three different electric field ranging from 10 to 30 V/cm for a flow rate  $v_m$  of 1 mm/s. **B:** Separation of a custom DNA ladder containing four bands of 50, 100, 150 and 210 kb at 0.5 mm/s and 9 V/cm with the custom-synthesized PVP 43 kDa dissolved at a volumic concentration of 3%. This custom DNA ladder was obtained by the digestion of 2  $\mu$ g of four BAC clones with the FastDigest NotI restriction enzyme (Thermo Scientific Fermentas) for 40 minutes at 37°C.



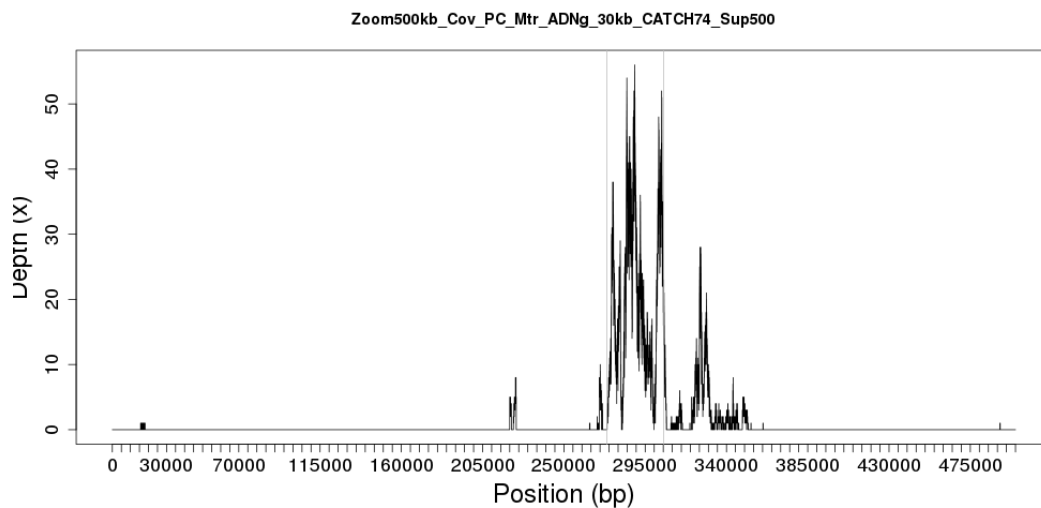
**Supplementary Figure S3:** **A:** PFGE analysis of the DNA BAC clone of 100 kb. The UD lane corresponds to the undigested sample. The D lane is the double digested BAC using Cas9 and 2 sgRNA with one fourth of a plug, *i.e.* 500 ng of DNA. The high MW PFG ladder and the corresponding fragment sizes are shown in the MW lane. **B:** Quantification and sizing of the Cas9 digested BAC clone (Blue) based on the fluorescence ratio and the migration time of the reference kb extend DNA ladder (Black). **C:** FEMTO Pulse® analysis of the Cas9 digested BAC clone.



**Supplementary Figure S4:** Evaluation of the selection window for the purification of the 31 kb fragment. The blue curve corresponds to the sample obtained by purification of the broad gDNA signal obtained by direct electroelution of the blocks (see Fig. 5A in main text). The black curve is the kb extend ladder, and the two red dashed lines correspond to references of 25 and 35 kb.



**Supplementary Figure S5:** **A:** PFGE analysis of *M. truncatula* genomic DNA without Cas9 nor sgRNA (C-), or double digested by Cas9 using the 2 sgRNA (E). The lanes MW and C+ correspond to the high MW PFG ladder with the first bands indicated in the left panel, and Cas9 double digested BAC DNA, respectively. **B:** Visualization of the 31 kb fragment from Cas9 digested gDNA by the FEMTO Pulse® Instrument using PFGE, band excision and electro-elution (Method 2 or 3 in Main text).



**Supplementary Figure S6:** The graph presents the reads depth as a function of the physical position on the first 500 kb of the *Medicago truncatula* chromosome V sequence. The vertical grey lines represent the position of the sgRNA on the reference BAC sequence. The X-axis corresponds to the position in the sequence in bp and the Y-axis to the depth.



**Supplementary Table S1.** Signal to noise ratio (S/N) of the DNA peaks observed in a serial dilution of the kb extend DNA ladder, from 1 to 20 ng/ml. The limit of detection (LOD) is estimated for each single DNA band using the noise level calculated by the Open Lab software and using the concentration threshold of S/N=3 to determine the LOD.

Total DNA ladder concentration (pg/ $\mu$ L)	Signal to noise ratio				Single fragment
	20	10	3	1	LOD (fg/ $\mu$ L)
3 kb	240	139	53	30	20
6 kb	205	120	37	25	10
10 kb	280	160	62	25	13
50 kb	175	92	33	11	20

**Supplementary Table S2.** Evaluation of the precision and the accuracy of DNA sizing over ten analysis of the 1 kb extend DNA ladder from NEB. The mean calculated size over the ten analysis, the standard deviation (SD), the precision and the accuracy of the measures are reported for each fragment of the ladder.

Expected size (bp)	Calculated size (bp)	SD	Precision (% CV)	Accuracy (%)
3000	2947	21.3	1	2
4000	3941	26.4	1	1
5000	4935	37.4	1	1
6000	5900	58	1	2
8000	7955	100	1	1
10000	9930	209	2	1
15000	15530	667.7	4	4
20000	19735	857.8	4	1
50000	51300	1249.2	2	3

**Supplementary Table S3.** Evaluation of the precision and the accuracy of DNA quantification over ten analysis of the 1 kb extend DNA ladder from NEB. The mean calculated mass over the ten analysis, the standard deviation (SD), the precision and the accuracy of the measures are reported for each fragment of the ladder.

Fragment size (kb)	Expected mass (pg)	Calculated mass (pg)	SD	Precision (% CV)	Accuracy (%)
3	100	104.1	1.1	1	4
4	26	27.5	0.3	1	6
5	33	31.9	0.8	3	3
6	40	38.9	1.1	3	3
8	33	32.7	1.2	4	1
10	33	35.0	1.8	5	6
15	45	38.0	1.3	4	16
20	30	33.6	2.5	6	12
50	30	34.0	2.8	9	13

**Supplementary Table S4:** Sizes and concentrations obtained from the FEMTO *Pulse*® analysis of the Cas9 digested BAC clone in bp and pg/ $\mu$ L respectively.

Peak Size (in bp)	Concentration (in pg/ $\mu$ L)
107	1061.5
32821	3174.7
57867	4937
103100	612.8

## 4.1 Conclusion du Chapitre 4

Dans ce chapitre, nous avons réalisé une optimisation de la technique  $\mu$ LAS pour la manipulation des fragments d'ADN jusqu'à 200 kb. Pour cela, nous avons défini les conditions d'actionnement électro-hydrodynamique ainsi qu'une formulation des propriétés viscoélastiques du tampon adéquate. Nous avons atteint une sensibilité de 10 fg/ $\mu$ L sur un fragment de 10 kb en un temps d'analyse de 60 minutes environ. Cette excellente sensibilité rend la technologie  $\mu$ LAS particulièrement adaptée à la détection et à l'isolation de longs fragments d'ADN ciblés par la technique CRISPR/Cas9. Cette méthode d'isolation spécifique, appelée CATCH, est particulièrement intéressante pour l'analyse des génomes complexes. Cependant, la purification du fragment d'intérêt est laborieuse car nécessitant une électrophorèse en champs pulsés. De plus, les précédentes études sur cette méthode CATCH n'ont pas permis d'obtenir un assemblage convaincant de la région ciblée après le séquençage. Nous avons montré dans ce chapitre que la technologie  $\mu$ LAS permettait, en complément d'un gel d'électrophorèse en champs pulsés (PFGE), d'isoler un fragment de 30 kb issu de l'ADN génomique de la plante *Medicago Truncatula*, et d'obtenir un assemblage de la région d'intérêt après un séquençage avec la technologie PacBio. Le facteur d'enrichissement obtenu est en particulier de 892 contre environ 200 pour les précédentes études utilisant uniquement un PFGE.

Nous n'avons toutefois pas réussi à obtenir un assemblage directement à partir d'une isolation  $\mu$ LAS en se dispensant d'un PFGE. En effet, nous n'avons pas observé le fragment d'intérêt après électro-élution du matériel génomique directement du plug d'agarose, malgré des concentrations théoriques au-dessus de la limite de sensibilité de la technologie  $\mu$ LAS. Durant les analyses, nous avons observé la présence d'ADN génomique dégradé, entre 20 et 100 kb, qui masquait vraisemblablement le signal du fragment d'intérêt. Ainsi, la collecte « en aveugle » de la zone d'intérêt n'a pas permis d'obtenir suffisamment de données dans la région d'intérêt pour obtenir un assemblage. De plus, les données de séquençage ont montré une forte contamination de l'échantillon d'intérêt par de l'ADN humain dont nous ne comprenons pas encore l'origine.

## 5. Développement d'une vanne accordable pour la sélection en taille d'ADN

---

Nous avons vu dans le premier chapitre que de nombreuses technologies de génomique de 3<sup>ème</sup> génération utilisent de l'ADN de haut poids moléculaire. Afin d'isoler de l'ADN de qualité suffisante, les protocoles de préparation comprennent souvent une étape de sélection en taille visant à éliminer les petits fragments d'ADN qui limitent les performances des séquenceurs. Nous avons vu au cours du deuxième chapitre que la principale technologie pour effectuer cette sélection en taille était le BluePippin, commercialisé par SageScience. Cet instrument a cependant le défaut de consommer de grandes quantités d'échantillons à cause de faibles rendements et d'avoir des temps de fonctionnement très longs.

La technologie  $\mu$ LAS nous semble particulièrement adaptée à la préparation et à la sélection en taille d'ADN pour les techniques de génomiques de 3<sup>ème</sup> génération. Nous avons ainsi développé une vanne de sélection d'ADN accordable par tension électrique adaptée à l'enrichissement spécifique de fragments de haut poids moléculaire. Afin de confronter les performances de cet outil aux spécifications des techniques de 3<sup>ème</sup> génération, nous avons entrepris une collaboration avec la plateforme de séquençage de Toulouse (GeT-PlaGe) afin de préparer et sélectionner en taille un échantillon pour la technologie Chromium de 10X Genomics. L'objectif de ce projet est de fournir environ 20 ng d'ADN de taille supérieure à 40 kb à partir d'un échantillon génomique de Melon fourni par le GeT-PlaGe.

L'étude suivante, présentée sous la forme d'un article, décrit le principe de fonctionnement de cette vanne de sélection d'ADN, ses performances, ainsi que son optimisation au format multicapillaire afin de traiter des quantités de matière suffisantes. Nous présentons également la préparation de l'échantillon d'ADN de melon et ses caractéristiques avant l'étape de séquençage. Cette étude n'est cependant pas totalement finalisée et, à l'heure d'écrire ces lignes, nous n'avons pas encore les résultats de séquençage.

## **HIGH MOLECULAR WEIGHT DNA PURIFICATION WITH AN ACCORDABLE SELECTION THRESHOLD**

Nicolas MILON<sup>1,2</sup>, Juan Fuentes<sup>1</sup>, Vincent Picot<sup>2</sup>, Laurent Bigot<sup>3</sup>, Géraud Bouwmans<sup>3</sup>, Karen Baudelle<sup>3</sup>, Audrey Boutonnet<sup>2</sup>, Frédéric Ginot<sup>2</sup>, Aurélien Bancaud<sup>1,\*</sup>

<sup>1</sup>*CNRS, LAAS, 7 avenue du colonel Roche, F-31400, Toulouse, France.*

<sup>2</sup>*Picometrics, 478 Rue de la Découverte, 31670 Labège, France*

<sup>3</sup>*Univ. Lille, CNRS, UMR 8523 - PhLAM - Physique des Lasers Atomes et Molécules, F-59000 Lille, France*

Correspondence: [abancaud@laas.fr](mailto:abancaud@laas.fr)

Keywords: Size selection, enrichment,  $\mu$ LAS, HMW DNA, Sequencing

## **Abstract**

DNA purification is an essential step of molecular biology, which is often associated to size selection with the removal of low molecular weight (MW) residues. The production of quality sequencing data in third generation sequencing calls for high MW size selection of fragments of more than 10 kb. This demand is mostly fulfilled using pulsed field gel electrophoresis despite the limitations in terms of labor effort and yield of this technology. Here, we describe the principle and setup of a size-accordable valve for the purification of high MW DNA, which is based on the monitoring of DNA transport in electrohydrodynamic flows. First, we derive a model to predict the size selection threshold as a function of the operating parameters. This model proves to be in excellent agreement with our data for a DNA purification system built on a monicapillary device. We specifically demonstrate the fractionation of samples in the size range 3 to 50 kb with collection yields of 90%, but also show a saturation for 0.2 ng of purified material. We then scale up the device with a multicapillary fiber composed of an array of 61 channels, and demonstrate the purification of 20 ng of genomic DNA with a MW larger than 20 kb. Altogether, our report shows a scalable and versatile solution to high MW DNA purification.

**Introduction:**

DNA purification is an essential step of molecular biology to produce target sequences for downstream analysis or cloning (Borst and Miller, 1990). Purification is often associated to a size selection process, in which low molecular weight (MW) residues up to a variable cut-off are eliminated. In the case of DNA synthesis by PCR, free nucleotides and primers of 20 to 30 bases have to be washed away after the reaction for amplicon isolation. This operation is mostly accomplished using reversible adsorption of DNA to silica matrices (Boom et al., 1990). High MW DNA molecules remain stably bound to silica during the washing steps with organic alcohols, but not nucleotides and oligonucleotides that are removed. A more complex selection is performed for cloning applications (Stadler et al., 2004) because the construction of plasmids relies on the ligation of purified pieces of DNA of 500 bp to 5 kb. Each fragment is generally isolated by band excision after agarose slab gel electrophoresis, a common technique yet labor-intensive and low-yield.

Size selection is also carried out during DNA library preparation before sequencing. Indeed, tightly sized insert libraries and removal of small fragments like adapters or adapter dimers help to optimize second generation sequencing performances (Quail et al., 2012). Most laboratories use solid-phase reversible immobilization on magnetic beads to perform DNA purification and size-selection simultaneously. The size range of the purified material is lower than 1000 bp and often broadly distributed. In the context of third generation sequencing (TGS), the need of high MW DNA selection methods in the range of 10 to 200 kb to produce quality sequencing data (Feehan et al., 2017) is only partially fulfilled by current technologies. The method of choice remains pulsed-field gel electrophoresis (PFGE), yet at the expense of excessively long run times (Gabrieli et al., 2018). It can nevertheless be automated, allowing collection of programmable DNA fractions with a maximum size cut-off of 40 kb (Gong et al., 2017) with a moderate collection yield of ~30% and a run time of 20 h. Due to these limitations, recent developments for targeted retention of DNA molecules of 100 kb or more on magnetic particles have been successfully performed, enabling the collection of 10 µg of genomic DNA (Becker et al., 2016).

Microfluidic technologies appear to offer unique solutions for the manipulation and purification of high MW DNA. For instance, DNA electrophoresis in artificial separation matrices made out of periodic arrays of obstacles etched in glass or silicon demonstrated their relevance for sorting and subsequent collection of DNA molecules of ~100 kb in 15 s (Huang et al., 2002). Probably due to the small cross-section of separation channels and the resulting limited amount of collected material, the exploitation of artificial separation matrices for DNA purification in TGS has not been reported. We recently developed the µLAS technology for the separation, concentration and isolation of DNA fragments in the size range 0.1 to 200 kb (Andriamanampisoa et al., 2018; Milon, 2018). This technology consists in monitoring DNA transport in a microfluidic channel by using simultaneous electro-hydrodynamic actuation (Chami et al., 2018; Ranchon et al., 2016). Because the technology is operated without separation matrices, it is readily adapted to purify a target fragment using its size as selection criterion. For instance, we recently reported

the isolation of a Cas9-digested 30 kb molecule from the bulk of genomic DNA (Milon, 2018), but the small amount of output material of 0.1 ng required a step of whole genome amplification before any analysis by TGS. Yet, the latter step appeared to be associated to the formation of chimera, complicating downstream bioinformatic analysis. Here, we describe the principle of DNA size-accordable valve for the purification and subsequent sequencing of 20 ng of high MW DNA. The first section of this report describes the operating principle of this valve. The second one is an experimental study of the DNA valve in a single channel system with an analysis of the size selection threshold in the range of 0.2 to 50 kb, and of the saturation of this system at ~0.3 ng. Finally, we design a parallelized system based on the same principle with 61 capillaries in photonic crystal geometry, which allows us to produce 20 ng of DNA.



## Material and Methods

### Reagents

Molecular biology grades chemicals were purchased from Sigma-Aldrich (St. Louis, MO). The Quick-Load® 1 kb Extend DNA Ladder (# N3239S), containing 13 bands from 0.5 to 50 kb and the Low Molecular Weight DNA Ladder (# N3233S), containing 11 bands from 25 to 766 bp were purchased from New England Biolabs (Ipswich, MA). Running buffers were composed of 1X Tris-Borate-EDTA (TBE, 89 mM Tris, 89 mM boric acid, 2 mM EDTA) supplemented with 5% 360 kDa Polyvinylpyrrolidone (PVP) or 3% 43 kDa PVP for low and high MW DNA, respectively (Andriamanampisoa et al., 2018; Milon, 2018). The intercalating dye (Picometrics Technologies, France, # 16-BB-DNA1K/01) was used at a 2X final concentration in the buffer solution. All buffer solutions were filtered at 0.22 µm before use.

### Multicapillary fiber fabrication

The multi-capillary system is based on an air/silica assembly obtained using the technology developed to manufacture photonic crystal optical fibers by the so-called stack-and-draw method (Russell, 2003). More precisely, a meter-long silica tube of 25 mm outside diameter (OD) and 15 mm inside diameter (ID) is elongated into several tens of millimeter-sized capillaries. This operation is performed on a drawing tower, a vertical equipment consisting, among all, in a feeding preform unit, a high temperature furnace and a tractor unit. The as-drawn capillaries are then manually assembled together to form an hexagonal stack that is inserted into a tube of 25 mm OD and 19 mm ID. This stack is drawn into a multi-capillary cane using the above-mentioned drawing tower. At this step, attention is paid to the collapsing of the interstitial holes between the circular capillaries while preserving their inner holes. The hole size,  $d$ , and hole-to-whole distance, pitch  $\Lambda$ , of the hexagonal array are adjusted by the drawing ratio whereas  $d/\Lambda$  is mainly governed by the dimensions of the starting materials. A fiber of about 900 µm in diameter and constituted of 61 capillaries of 45.8µm in diameter (standard deviation 0.6) is obtained.

### Instrumentation and capillary assembly

Fused silica monicapillaries were purchased from Polymicro Technologies (Phoenix, US). The monicapillary device is composed of three parts (Fig. 1A-B): 5 cm of 50 µm ID capillary, 4 cm of 330 µm ID capillary acting as an injection chamber of 3.4 µL and another 19 cm of 50 µm ID capillary. Monicapillary capillary devices were assembled manually at Picometrics as previously described (Andriamanampisoa et al., 2018) using a UV-curable acrylate-based glue. Multicapillary devices are composed of a 2 cm-long home-made fiber of 1 mm OD assembled with a vial insert of 300 µL acting as a reservoir (# 4025 GF-625, J.G. Finneran Associates, Vineland, NJ; Fig. 1C-D). The bottom of the vial insert is cut to accommodate the fiber piece that is subsequently glued with the acrylate UV-based glue.

The multicapillary device is operated with a prototype composed of a device holder that can be adjusted in height and a carousel to perform sample fractionation (Figure 1, D). Two electrodes connected to the vial insert and the collection vial are integrated in the prototype.

Monocapillary experiments were carried out with an Agilent 1600CE system equipped with a Zetalif LED 480 nm detector (Picometrics Technologies, France) and modified to install the capillary device with minimum bending by machining the center of the cassette to place the optical head of the detector 7 cm downstream of the concentration junction. In the multicapillary system, the sample is manually injected at the entrance of the fiber and liquids flow by gravity. Constant flow velocity in the device is maintained by regularly refilling the vial insert throughout the experiment. With a multicapillary fiber of 2 cm, the total height of liquid in the device is 5.5 cm, corresponding to a mean flow velocity of 1.2 mm/s (See supplementary for the calculation). Prior to use, every capillary device was rinsed with 1 M NaOH, then with 0.1 M HCl, and finally flushed with de-ionized water. Prior to each experiment, the devices were conditioned with 0.1 M HCl, 1% poly-vinylalcohol (PVA), and flushed with running buffer.

After size selection, purified fractions of low and high MW DNA were characterized with the  $\mu$ LAS separation and titration protocol described in refs (Andriamanampisoa et al., 2018; Milon, 2018). Size accordable selection in monocapillary and subsequent analysis are performed on the same capillary device but used in opposite polarity. The short 5 cm side of the device is used for size selection to speed up the leak phase whereas the 19 cm side is used for the sizing of the fractions.

## **Principle of the DNA valve controlled by an electric field**

In the conventional operation of  $\mu$ LAS (Andriamanampisoa et al., 2018), DNA manipulation is performed in a fluidic system formed by the assembly of two capillaries of different inner diameters of 320 and 50  $\mu$ m (Fig. 1A). The sample is conveyed in the large capillary using a constant pressure source and a counter electrophoretic force associated to an electric field of 10 to 100 kV/m. Transverse viscoelastic forces occur (Ranchon et al., 2016), dragging and trapping DNA molecules at the vicinity of the walls (Chami et al., 2018) where the flow velocity is low. Because transverse viscoelastic forces are much lower in the large capillary, hydrodynamic transport is prevalent upstream of the constriction. Conversely, the fast flow velocity in the narrow capillary triggers a high viscoelastic force associated to a mode of transport dominated by electrophoresis (upper panel of Fig. 1A), and defining a concentration at the constriction. As soon as the electric field decreases, transverse viscoelastic forces decrease and DNA molecules progressively escape from the constriction toward the small capillary according to their MW (Fig. 1B). The DNA valve therefore consists in setting the electric field to a threshold value that enables the passage of molecules up to a given size and retain the high MW fraction. The purified genomic material is finally obtained by turning off the electric field, and collecting the molecules in a fresh vial.

More quantitatively, the transverse viscoelastic force increases linearly with the fluid maximum velocity  $V_{p_0}$ , the electrophoretic velocity  $V_e$ , and the DNA MW according to:

$$F_{VE}(\varepsilon) \sim -K\eta\tau N \frac{l_K}{R} V_{p_0} V_e \frac{\varepsilon}{R} \quad (1)$$

with  $\eta$  the fluid viscosity,  $\tau$  the fluid elastic time constant,  $N$  the number of monomers in the chain of Kuhn length  $l_K$ ,  $R$  the radius of the capillary, and  $\varepsilon$  the distance from the walls.  $K$  is a constant derived from the Maxwell model of viscoelastic fluids. This force is equivalent to an elastic spring, maintaining DNA molecules near the walls. Hence the conditions to stop the molecules are achieved for an electrophoretic velocity  $V_e$  equal to  $\sim 10\%$  of the maximal fluid velocity  $V_{p_0}$  (Ranchon et al., 2016). Notably, the concentration of PVP for high MW DNA separation and concentration is lower than the overlapping concentration (Milon, 2018), suggesting that transverse forces also occur in Newtonian flows (Montes et al., 2018). This effect has been studied in theory, predicting a similar scaling dependence for the transverse force in DNA size, hydrodynamic and electrophoretic velocities (Butler et al., 2007).

The DNA size threshold  $N_c$  corresponds to a unique value of the electric field  $E$ , and equivalently of the electrophoretic velocity defining the mobility  $\mu_0$  as  $V_e = \mu_0 E$ . Using Boltzmann statistics and the force expression Eq. (1) (Chami et al., 2018), we deduce the average position of DNA molecules from the surface:

$$\langle \varepsilon \rangle \cong \sqrt{\frac{k_B T R^2}{K' \eta \tau N l_K V_{p_0} \mu_0 E}} \quad (2)$$

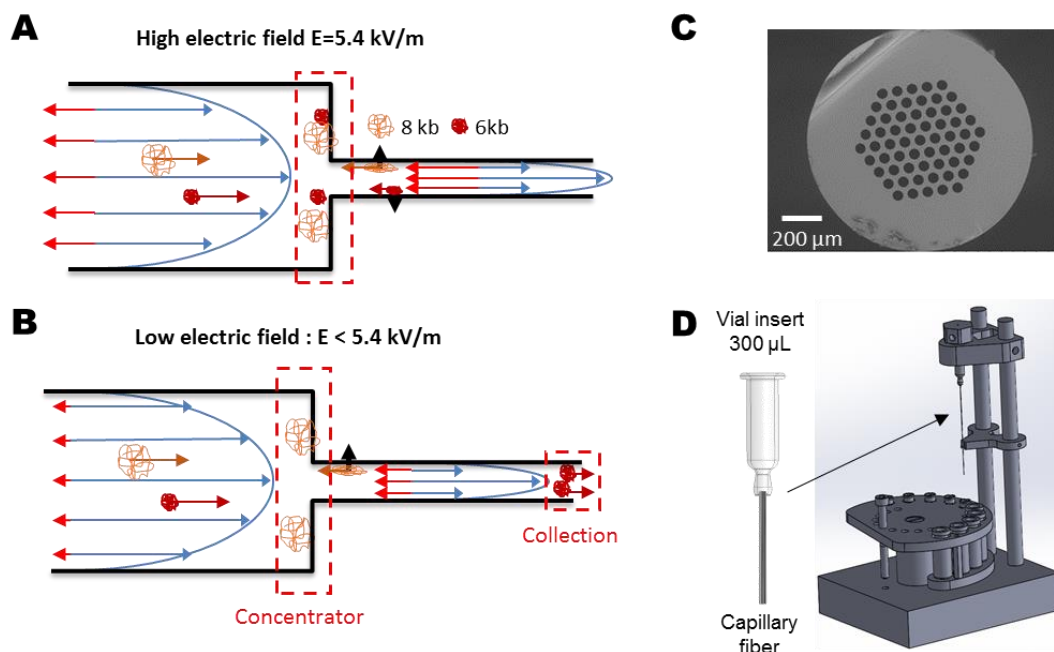
The threshold size  $N_c$  is defined by the electric field such that the velocity in the narrow capillary is null:

$$\frac{2V_{p_0}}{R} \langle \varepsilon \rangle = \mu_0 E \quad (3)$$

Plugging Eq. (2) into Eq. (3), we deduce that:

$$N_c = \frac{k_B T}{K' \eta \tau l_K \mu_0^3} \times \frac{V_{p_0}}{E^3} \quad (4)$$

We thus conclude that size selection can readily be piloted by the electric field by retaining molecules of MW greater than an arbitrary threshold calibrated by the operator (see more below). Interestingly, this prediction applies to the monocapillary system, as described in refs. (Andriamanampisoa et al., 2018; Milon, 2018) but also to the multicapillary system (Fig. 1C-D). Indeed, parallelization is expected to linearly increase the electric current and the flow rate in the device with the same settings for the maximum flow and electrophoretic velocities in each channel of the multicapillary. This prediction is relevant for preparative operations in the context of TGS technologies.



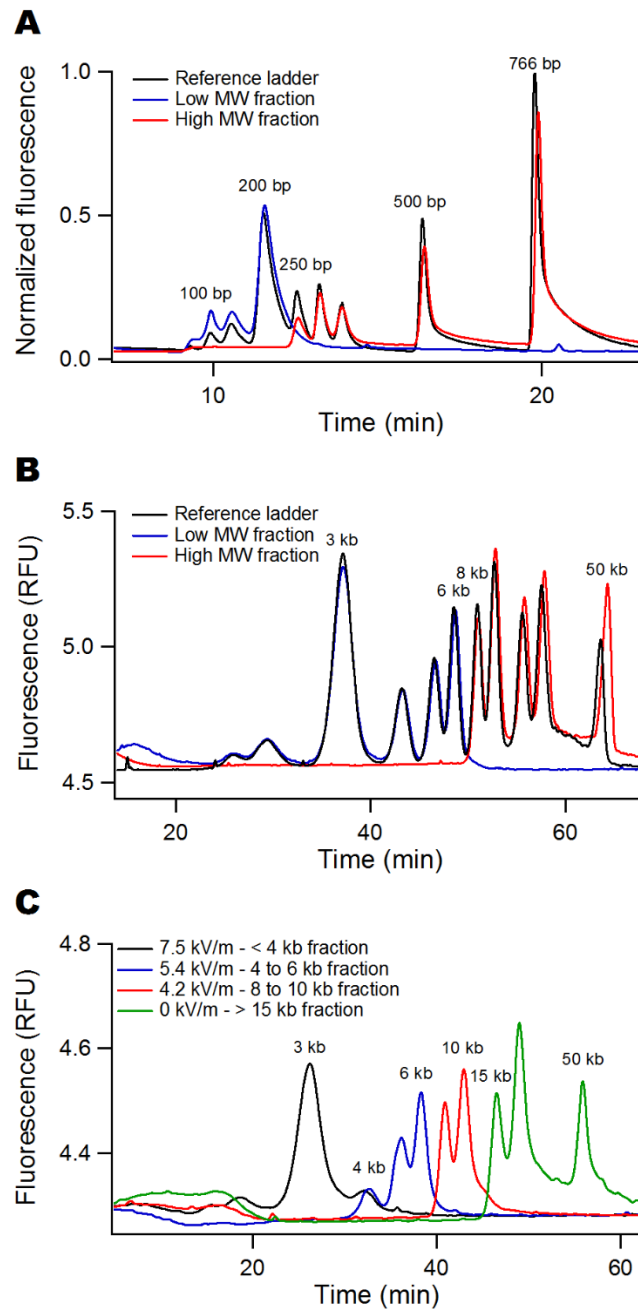
**Figure 34 : Principle of the DNA valve and associated instrument.** (A) The sketch represents the behavior of DNA molecules in a constriction upon application of a high electric field in order to retain high and low MW molecules (in red and orange). (B) The same representation for a reduced electric field to allow the passage of small molecules specifically, which are collected downstream of the small capillary. (C) The electron micrograph shows a multicapillary system with 61 channels of 46  $\mu\text{m}$ . (D) The schematics presents the DNA purification prototype with the multicapillary system.

## Results

### Characterization of the DNA valve in a monocapillary

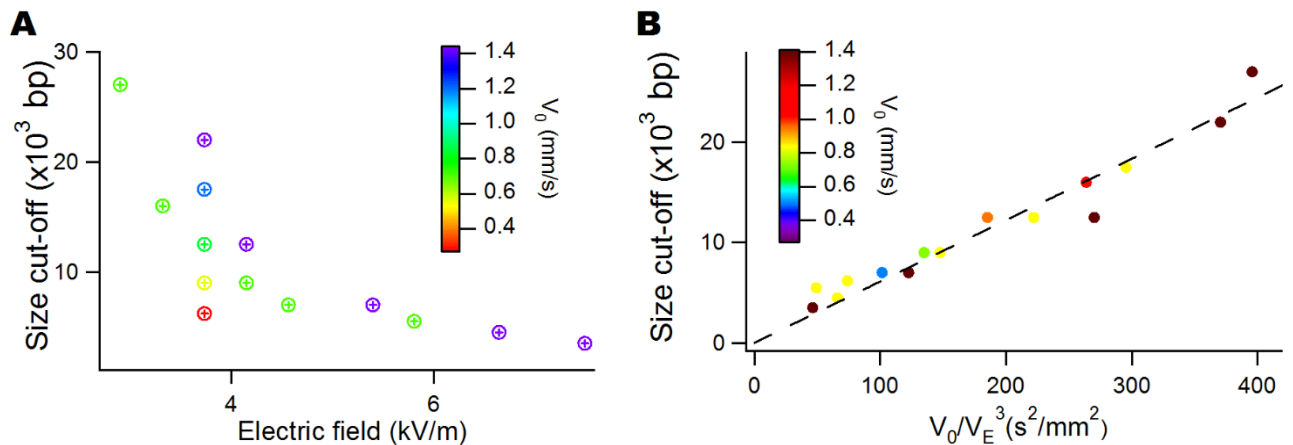
We first characterized the size selection cut-off for DNA molecules of MW spanning 0.2 to 50 kb. We used two different DNA ladders in the low and high MW limits and performed the experiments with two viscoelastic fluid formulations (see Methods). Our experiments consisted in loading 0.1 ng of DNA ladder at the constriction at a given electric field that corresponded to a size selection threshold. After 20 minutes, the electric field was turned off to retrieve the purified fraction in a separate vial. The molecular composition of the leak and purified fractions were subsequently analyzed using conventional DNA size analysis by  $\mu\text{LAS}$ . As an example in Fig. 2A-B, we report the reference ladder chromatogram in black, and superimpose the leak and the purified fractions in blue and red, respectively. The cut-offs of 230 bp (Fig. 2A) and 7 kb (Fig. 2B) were for instance obtained setting the electric field to 55 and 5.4 kV/m, and the flow velocity to 13.6 mm/s and 1.5 mm/s, respectively. The electric field therefore appears to be a versatile control parameter for size-accordable DNA purification. This assertion is further confirmed by fractioning the high MW DNA ladder shown in Fig. 1B into four products with three consecutive settings

for the electric field (Fig. 2C). The final retained fraction plotted in green corresponds to the peaks of 15, 20, and 50 kb, suggesting that high MW DNA can be retained for TGS technologies.



**Figure 2: DNA size accordable valve in a monicapillary.** (A) The chromatogram shows the reference DNA ladder with sizes in the range 100 to 766 bp in black, as well as the leak and purified fractions with a threshold electric field of 55 kV/m that corresponds to a cut-off size of ~230 bp (blue and red curves, respectively). Notably, the concentration phase could be modulated in the range 5 to 35 minutes without any loss in DNA fluorescence signal for the purified fraction (Supplementary Fig. S1). (B) The chromatogram shows the high MW ladder with bands spanning 3 to 50 kb in black with the outputs of a purification process. The threshold electric field is set to 5.4 kV/m, and the threshold size  $N_c$  is 7 kb. (C) The high MW ladder can be sampled in four fractions using a series of three threshold electric fields, as indicated in the caption.

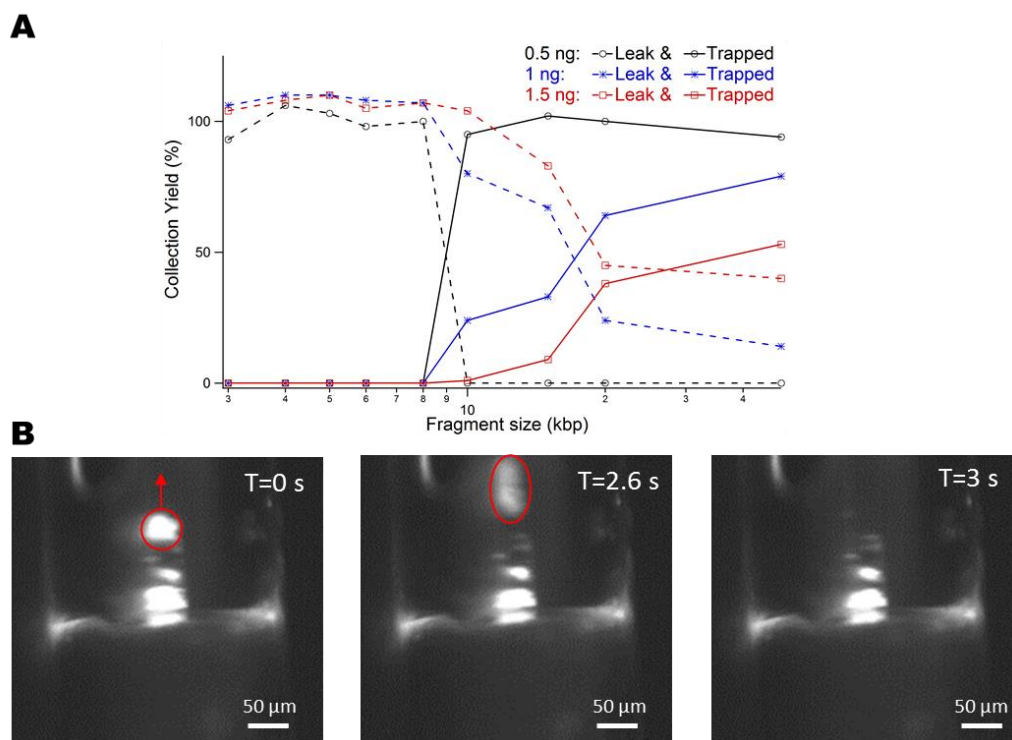
In Fig. 3A, we plot the cut-off as a function of the electric field for various hydrodynamic speed adapted to high MW DNA purification. The cut-off decreases monotonously as the electric field increases, confirming that its value could be precisely adjusted to purify target MW fractions in complex mixtures. More quantitatively, we plotted the size cut-off as a function of  $V_0/V_E^3$  as described by our model in Eq. (4) (Fig. 3B). The resulting data followed a linear response (dashed line in Fig. 3B), in agreement with our model for a set of  $V_0$  spanning 0.3 to 1.4 mm/s. Altogether, we demonstrate that the control of the electric field enables the purification of DNA with an accordable size exclusion cut-off.



**Figure 3: Experimental determination of the cut-off threshold.** (A) The graph shows the size selection cut-off as a function of the electric field for high MW DNA for various flow velocities, as indicated in inset. (B) Same dataset normalized according to our model and Eq. (4).

### Saturation of the DNA valve in a monicapillary device

In order to produce 20 ng of DNA for TGS technologies, we wished to evaluate the maximal amount of DNA that could be purified in a monicapillary device. We therefore performed a series of experiments that consisted in injecting gradual amounts of DNA ladder in the monicapillary system. We used with the same pressure and voltage settings as in Fig. 2B and retained molecules larger than 7 kb, which corresponded to 34% of the whole mass of the ladder. For 0.2 ng of DNA at the constriction (black datasets in Fig. 4A), the retained fraction appeared to be isolated with the expected cut-off and a collection yield higher than 90%. These performances appeared to be progressively degraded for larger amounts of DNA. The blue and red curves in Fig. 4A obtained for 0.3 and 0.5 ng of DNA at the constriction show a much lower collection yield in the retained fraction. Albeit fragments smaller than 8 kb were efficiently rejected from the device, the monicapillary system appeared to be inadequate to process 10 ng of DNA or more. Interestingly, similar experiments carried out with the low MW DNA ladder and a cut-off set at 230 bp indicated a saturation for  $\sim 10$  ng in the constriction, i.e. fifty times more DNA than for high MW selection (data not shown).

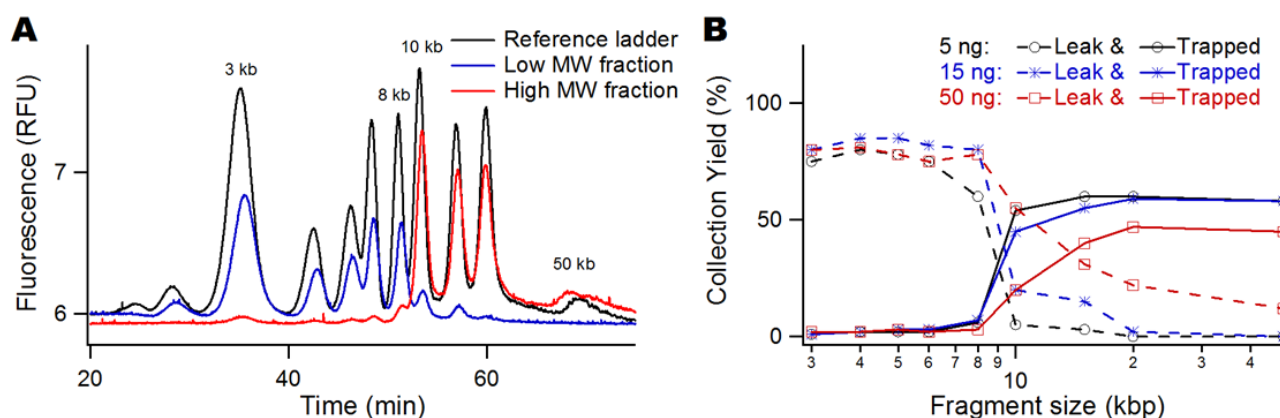


**Figure 4: Saturation of the DNA valve.** (A) The graph shows the yield of DNA collection using the valve selection settings shown in Fig. 2B. The yield is computed by determining the amount of each band in the chromatograms of the leak and purified fractions. (B) The fluorescence micrographs present a time-lapse recording of high MW DNA concentrated at the interface between two capillaries. The bright cluster marked in red is not retained by viscoelastic forces and leaks away from the constriction.

In order to explain the saturation of the capillary system, we recorded the dynamics of DNA at the constriction using fluorescence microscopy (Fig. 4B). We injected 8 ng of 75 kb DNA in the capillary device, setting the electric field and hydrodynamic flow velocity to 5.4 kV/m and 1.5 mm/s, respectively. We observed an accumulation of molecules in the narrow capillary associated to the formation of bright clusters, which could not be stably retained in the concentrator, as shown with the leak of a DNA aggregate from the concentrator in the middle panel of Fig. 4B. We suggest that the large hydrodynamic radius of the DNA cluster leads to the takeover of hydrodynamic over electrophoretic forces. Because the volume of the region where DNA is concentrated represents  $\sim 0.2$  nL, DNA concentration at the junction is estimated to be  $\sim 1$  mg/mL. Oppositely, low MW DNA fragments accumulate in the large ID capillary of 330  $\mu\text{m}$  (Andriamanampisoa et al., and Supplementary Fig. S2), in a volume of  $\sim 9$  nL which represents a concentration at the junction of  $\sim 1.1$  mg/mL, very similar to high MW saturation threshold. Interestingly, this estimate is intermediate between two non-mutually exclusive mechanisms, namely the solubility limit of DNA in the range 10-100 mg/mL (Cleaver and Boyer, 1972) and the threshold of electric field induced DNA aggregation instabilities that build up at concentrations higher than 0.05 mg/mL (Mitnik et al., 1995).

## Upscaling of the technology with multicapillary systems

To increase the amount of DNA processed by our technology, we fabricated a device with a multicapillary fiber containing 61 capillaries of 46  $\mu\text{m}$  I.D. (Fig. 1C). By simple multiplexing of the concentration process by a factor of 60, we expected to purify  $\sim 12$  ng of DNA. Because the multicapillary system could not be integrated in the Capillary Electrophoresis system, a specific prototype was designed (Fig. 1D), and its outlet was connected to a carousel for sample fractionation (see methods). Gravitational forces were sufficient to induce a flow velocity of 1.2 mm/s in the capillaries (see Supplementary Material), meaning that the operating conditions of the monicapillary system could be reproduced without pressure source. During the retention phase, the electric field was set to 4 kV/m during 40 min and was then turned off to collect the purified material. The leak and retained fractions were subsequently characterized with a monicapillary system, showing that molecules larger than  $\sim 9$  kb could be selected although the sharpness of the selection filter was lower than for a monicapillary (Fig. 5A). Nevertheless, this experiment readily showed that the single capillary technology could be efficiently transposed to a multicapillary system with similar electro-hydrodynamic actuation parameters.



**Figure 5: Saturation of the DNA valve.** (A) The chromatogram shows the high MW ladder with bands spanning 3 to 50 kb in black with the outputs of a purification process on the multicapillary system. The threshold electric field is set to 4 kV/m, and the threshold size is 9 kb. (B) The graph shows the yield of DNA collection using the valve selection with the multicapillary system. The yield is computed by determining the amount of each band in the chromatograms of the leak and purified fractions.

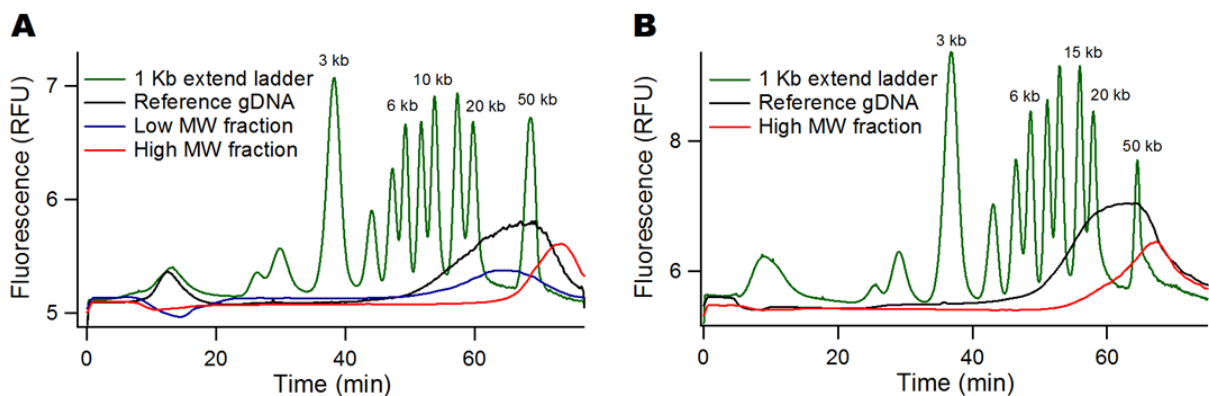
We then evaluated the saturation threshold of the multicapillary device by injecting gradual amounts of the kb extend DNA ladder from 5 to 50 ng, *i.e.* 1.5 ng to 15 ng retained at the constriction, with the same actuation parameters as in Fig. 5A. The collection yield of the small MW fraction remained constant with 75% for the 3 amounts injected, but the collection yield of the high MW fraction remained lower than 55% (Fig. 5B). The selection properties of the filter remained efficient up to 4.5 ng at the constriction, in good agreement with our linear argument of multiplexing. When 50 ng of DNA were injected, high MW fragments of 10 to 50 kb were also found in the leak fraction and the collection yield dropped to 40%,



indicating the saturation of the multicapillary system (Fig. 5B). Altogether, the upscaling of our technology to 61 capillaries allows us to carry out high MW DNA size selection up to 4 ng of material loaded in the constriction.

## High MW DNA selection from genomic DNA samples

The minimal amount of genomic DNA for linked-read sequencing is ~15 ng with fragments of ~40 kb or more. In order to validate our multicapillary system for the size selection of high MW DNA fragments, we purified a melon genomic DNA sample from a large amount of small MW DNA residues resulting from the extraction process. Indeed, size analysis revealed a broad size distribution from 3 to more than 50 kb (black curve in Fig. 6A), as confirmed using FEMTO Pulse analysis (Supplementary Fig. S3). We first established the voltage threshold to isolate DNA molecules larger than 40 kb on the multicapillary device. Using the same flow velocity of 1.2 mm/s, we set the electric field to 2.8 kV/m and collected a broadly distributed peak starting at ~40 kb (red curve in Fig. 6A), whereas the leak fraction was centered at ~25 kb (blue curve in Fig. 6A). This experiment indicates that the fractionation of genomic DNA samples can be performed with our technology. Although we used a low mass input of 2.5 ng, we note that the leak fraction contained residual amounts of high MW molecules, suggesting that saturation of the device occurred. This saturation threshold was evaluated to be 4 ng with molecules from 8 to 50 kb (Fig. 5B). We presume that the size of DNA fragments influences the saturation threshold and that long genomic DNA fragments, as those found in the melon sample, are prone to trigger the saturation.



**Figure 6: Saturation of the DNA valve.** (A) The chromatogram shows the high MW ladder with bands spanning 3 to 50 kb in green, the unpurified melon genomic DNA sample in black with the outputs of the size selection process on the multicapillary system in blue and red. The threshold electric field is set to 2.8 kV/m, and the threshold size is ~50 kb. This calibration experiment is carried out on 2.5 ng of melon sample. (B) The chromatogram shows the same size selection process carried out four times with 100 ng of DNA input in order to purify 15 ng of size selected melon DNA.

We then moved to the preparation of the melon sample for Chromium library preparation. As depicted in Fig. 5B, collection yield of the retained fraction drops when more than 4 ng of DNA are stacked

in the multicapillary device. In order to purify the 15 ng required for Chromium library preparation, we performed the size selection 4 times consecutively by collecting the retained fraction in the same vial. About 100 ng of melon sample were injected in the multicapillary device and the size selection lead to the collection of 25  $\mu$ L of purified sample. The retained fraction containing the high MW melon DNA was then reprocessed in the monicapillary device to assess its size distribution and its concentration (Fig. 6B). Although a size selection occurred, the cut-off was not as sharp and as high as expected. Indeed, we confirmed the removal of DNA fragments lower than 15 kb, however some fragments in the range 15 to 40 kb were not completely removed from the sample. We evaluated the sample concentration to be 0.7 ng/ $\mu$ L meaning that we purified ~17 ng of DNA above ~20kb (see methods). Therefore, we had enough DNA materiel to perform Chromium library preparation.

## Conclusion and perspectives

In this study, we described a size accordable DNA valve based on the  $\mu$ LAS technology. This DNA purification method allows the precise fractionation of a broadly distributed DNA sample from 200 bp to 50 kb. This method is therefore well suited for the removal of small DNA fragments before linked-read sequencing. We evaluated the maximum DNA mass that can be processed with this device and the collection yields to be 200 pg and 90% respectively. Above 200 pg stacked in the device, saturation occurs and high MW DNA fragments leaks from the constriction. Because the chromium technology requires ~15 ng of high MW DNA, we scaled up our DNA valve in a multicapillary format. This cheap and mere device, composed of a fiber containing 61 channels and a vial insert, performs the size-selection on 4 ng of DNA only with gravitational flow. In order to validate the technology, we tackled the purification of a broadly distributed melon DNA sample by isolating 17 ng of DNA of more than 20kb. This sample was then used for the Chromium library preparation and sequenced on an Illumina sequencer.

## Bibliography

- Andriamanampisoa, C.-L., Bancaud, A., Boutonnet-Rodat, A., Didelot, A., Fabre, J., Fina, F., Garlan, F., Garrigou, S., Gaudy, C., and Ginot, F. (2018). BIABooster: Online DNA Concentration and Size Profiling with a Limit of Detection of 10 fg/ $\mu$ L and Application to High-Sensitivity Characterization of Circulating Cell-Free DNA. *Analytical Chemistry* *90*, 3766–3774.
- Becker, L., Steglich, M., Fuchs, S., Werner, G., and Nübel, U. (2016). Comparison of six commercial kits to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. *Scientific Reports* *6*, 28063.
- Boom, R., Sol, C.J., Salimans, M.M., Jansen, C.L., Wertheim-van Dillen, P.M., and van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* *28*, 495–503.
- Borst, M., and Miller, D.M. (1990). DNA isolation and Southern analysis: a clinician's view. *Am. J. Med. Sci.* *299*, 356–360.
- Butler, J.E., Usta, O.B., Kekre, R., and Ladd, A.J.C. (2007). Kinetic theory of a confined polymer driven by an external force and pressure-driven flow. *Physics of Fluids* *19*, 113101.
- Chami, B., Socol, M., Manghi, M., and Bancaud, A. (2018). Modeling of DNA transport in viscoelastic electro-hydrodynamic flows for enhanced size separation. *Soft Matter* *14*, 5069–5079.
- Cleaver, J.E., and Boyer, H.W. (1972). Solubility and dialysis limits of DNA oligonucleotides. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* *262*, 116–124.
- Feehan, J.M., Scheibel, K.E., Bourras, S., Underwood, W., Keller, B., and Somerville, S.C. (2017). Purification of High Molecular Weight Genomic DNA from Powdery Mildew for Long-Read Sequencing. *J Vis Exp*.
- Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y., and Ebenstein, Y. (2018). Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Research* *46*.
- Gong, L., Wong, C.-H., Cheng, W.-C., Tjong, H., Menghi, F., Ngan, C.Y., Liu, E.T., and Wei, C.-L. (2017). Nanopore Sequencing Reveals High-Resolution Structural Variation in the Cancer Genome. *bioRxiv* 209718.
- Huang, L.R., Tegenfeldt, J.O., Kraeft, J.J., Sturm, J.C., Austin, R.H., and Cox, E.C. (2002). A DNA prism for high-speed continuous fractionation of large DNA molecules. *Nat Biotech* *20*, 1048–1051.
- Milon, N. (2018). Technology for DNA isolation coupled to Cas9-assisted targeting for assembly and sequencing of a 30 kb region in plant genome. In Submission.
- Mitnik, L., Heller, C., Prost, J., and Viovy, J.L. (1995). Segregation in DNA solutions induced by electric fields. *Science* *267*, 219–222.
- Montes, R.J., Butler, J.E., and Ladd, A.J. (2018). Trapping DNA with a high throughput microfluidic device. *Electrophoresis*.
- Quail, M.A., Gu, Y., Swerdlow, H., and Mayho, M. (2012). Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation. *Electrophoresis* *33*, 3521–3528.
- Ranchon, H., Malbec, R., Picot, V., Boutonnet, A., Terrapanich, P., Joseph, P., Leïchlé, T., and Bancaud, A. (2016). DNA separation and enrichment using electro-hydrodynamic bidirectional flows in viscoelastic liquids. *Lab Chip* *16*, 1243–1253.

Russell, P. (2003). Photonic crystal fibers. *Science* 299, 358–362.

Stadler, J., Lemmens, R., and Nyhammar, T. (2004). Plasmid DNA purification. *J Gene Med* 6 *Suppl 1*, S54-66.

## HIGH MOLECULAR WEIGHT DNA PURIFICATION WITH AN ACCORDABLE SELECTION THRESHOLD FOR LINKED READ SEQUENCING

### Supplementary Materiel:

#### Flow velocity in the multicapillary device:

The flow rate  $Q$  in the multicapillary device is described by:

$$Q = \frac{\Delta P}{R_h} \quad (1)$$

With  $\Delta P$  the hydrostatic pressure and  $R_h$  the hydraulic resistance of the fiber.

We define  $R_{eq}$  the hydraulic resistance of one channel in the multicapillary fiber. Then, the hydraulic resistance of the 61 channels in parallel is expressed by:

$$\frac{1}{R_h} = \sum_{i=1}^{61} \frac{1}{R_{eqi}} \quad (2)$$

And :

$$R_{eq} = \frac{8\mu L_f}{\pi R^4} \quad (3)$$

With  $\mu$  the dynamic viscosity of the fluid,  $L_f$  the length of the fiber and  $R$  the radius of one channel of the fiber.

We assume that all 61 channels of the fiber are equivalent, then:

$$R_h = \frac{R_{eq}}{61} \quad (4)$$

Hydrostatic pressure is defined by:

$$\Delta P = \rho g L_T \quad (5)$$

With  $\rho$  the volumetric mass of the fluid,  $g$  the standard gravity and  $L_T$  the total height of the device ( $L_T=L_f + L_v$ ).

Then we find :

$$Q = \frac{\rho g L_T 61 \pi R^4}{8 \mu L_f} \quad (6)$$

And :

$$Q = \frac{\rho g 61 \pi R^4}{8 \mu} \left(1 + \frac{L_v}{L_f}\right) \quad (7)$$

We can add a corrective term corresponding to capillary forces in the vial insert and obtain the expression of the flow rate in the fiber:

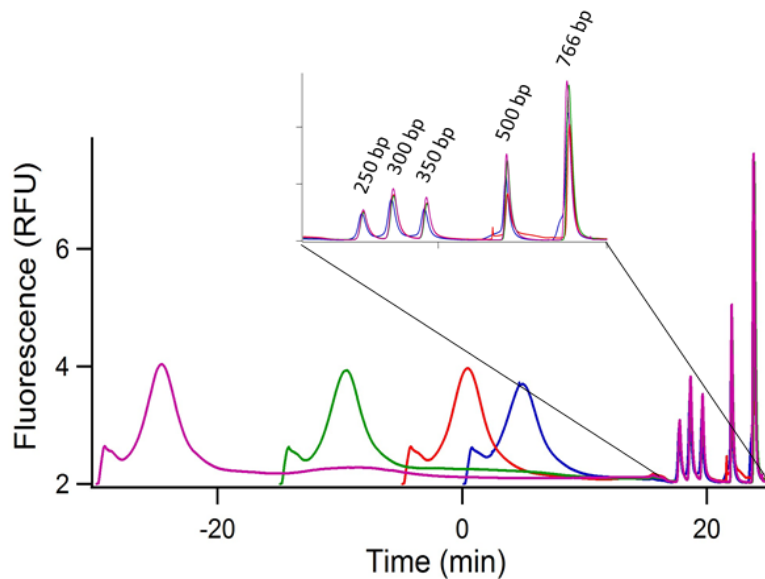
$$Q = \frac{\rho g 61 \pi R^4}{8 \mu} \left( 1 + \frac{L_v}{L_f} - \frac{2 \sigma}{R_v L_f \rho g} \right) \quad (8)$$

For one channel we have:

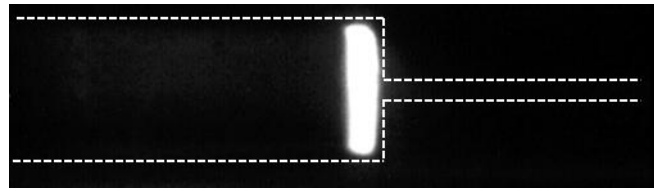
$$Q = \frac{\rho g \pi R^4}{8 \mu} \left( 1 + \frac{L_v}{L_f} - \frac{2 \sigma}{R_v L_f \rho g} \right) \quad (9)$$

We can then find the mean flow velocity in each channel of the fiber:

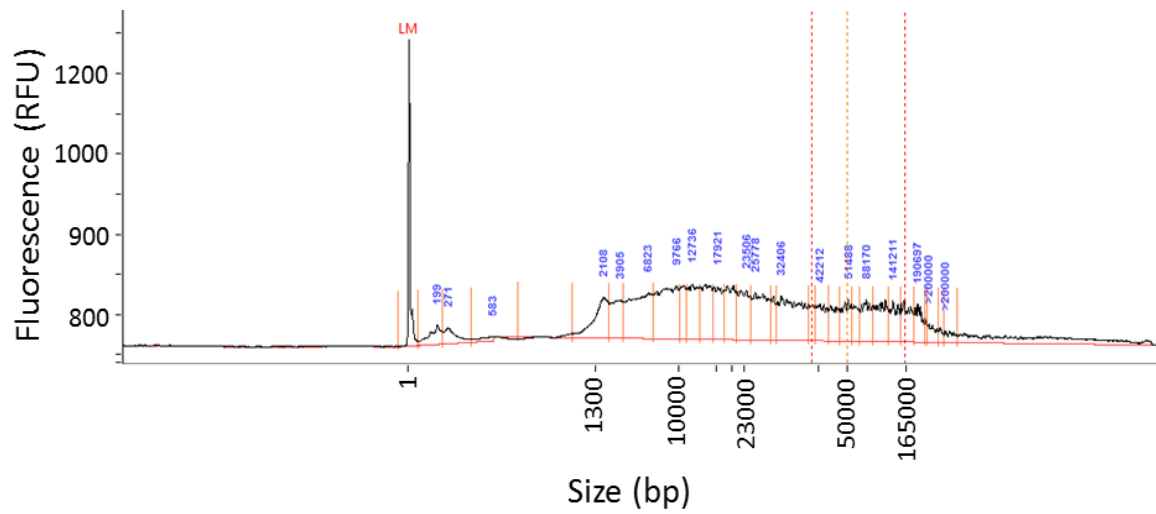
$$V = \frac{\rho g 61 R^2}{8 \mu} \left( 1 + \frac{L_v}{L_f} - \frac{2 \sigma}{R_v L_f \rho g} \right) \quad (10)$$



**Supplementary Figure S1:** The chromatogram shows the concentration of the low MW DNA ladder during 5, 10, 20, and 35 min (blue, red, green, and purple curves, respectively) followed by size separation by decreasing the electric field gradually to 0, as described in (Andriamanampisoa et al., 2018; Milon, 2018). Note that the datasets have been shifted in time according to the duration of the concentration phase, and that the signal intensity traces of the retained bands are similar in the four datasets.



**Supplementary Figure S2:** Fluorescence micrograph of the concentration with low MW DNA actuation parameters at the junction of two capillaries of different inner diameters (ID) of 100 and 20  $\mu\text{m}$ .

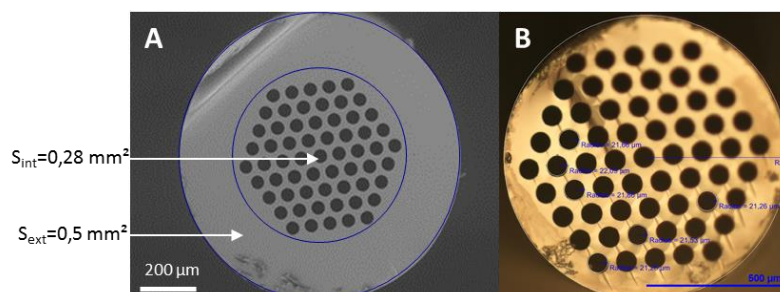


**Supplementary Figure S3:** FEMTO *Pulse*® analysis of the melon genomic DNA sample.

## 5.1 Conclusion du Chapitre 5

Dans ce cinquième chapitre, nous avons montré le développement d'une vanne de sélection d'ADN accordable permettant l'enrichissement sélectif de longs fragments. Nous avons évalué les performances de cette vanne en termes de précision et de quantité de matière qu'il est possible d'isoler. Afin de correspondre aux spécifications des technologies de génomique de 3<sup>ème</sup> génération, nous avons spécifiquement développé un prototype au format multicapillaire afin d'augmenter les masses d'ADN manipulables. Nous avons ensuite purifié 20 ng d'ADN issu d'un génome de melon, et confié cet échantillon au GeT-PlaGe pour un séquençage avec la technologie Chromium de 10X Genomics.

Nous remarquons que la fibre multicapillaire ne peut pas être intégrée dans l'instrument d'électrophorèse capillaire d'Agilent. Nous avons donc construit un prototype permettant d'appliquer pression et tension à travers la fibre. Ce prototype est cependant très manuel et son utilisation demande une certaine expérience. De plus, la pression est contrôlée uniquement par la hauteur d'eau dans le dispositif, ce qui simplifie l'instrumentation mais impacte fortement la reproductibilité des expériences. Il faut par exemple maintenir le niveau de tampon dans le réservoir continuellement durant les 40 minutes d'expérience, à la pipette. Les rendements sur cette fibre multicapillaire ont également chuté par rapport au format mono-capillaire. Une des suppositions est que la surface occupée par les canaux est faible devant la surface du reste de la fibre (Figure 35). Nous faisons ainsi l'hypothèse que l'ADN reste coincé sur cette couronne latérale de la fibre sans être évacuée par les canaux. Nous avons fait fabriquer une seconde fibre avec une protection plus fine, ce qui augmente la surface des canaux accessible pour le traitement des molécules. Nous avons également développé une seconde version du prototype, possédant un plus grand réservoir et moins manuel, qui permettra d'avoir des résultats plus reproductibles. Ces deux améliorations sont actuellement en cours d'évaluation.



**Figure 35 : Image en microscopie électronique à balayage des fibres multicapillaires fabriquées par l'IRCICA. A : Fibre utilisée dans l'étude du chapitre 5. B : Nouvelle fibre, avec une protection plus fine et donc un ratio surface canaux sur surface fibre plus intéressant.**



Bien que très performant sur de faibles quantités d'ADN, inférieures à 4 ng environ, ce format multicapillaire ne semble pas adapté à la préparation d'échantillons pour les techniques de génomique de 3<sup>ème</sup> génération. En effet, les quantités de matière manipulées sont trop proches de la limite inférieure des spécifications des outils du marché de séquençage. A ce jour, il apparaît compliqué d'atteindre 100 ng avec un simple « multiplexage » des canaux dans une fibre multicapillaire.

# Conclusion générale et perspectives

---

## Conclusion

Dans ce manuscrit, nous avons montré que le développement de nouvelles technologies d'analyse génomique est un domaine en plein essor. Nous avons détaillé les principales innovations, dites de 3<sup>ème</sup> génération, en nous attardant sur leurs exigences en termes d'échantillon d'ADN. Ces nouvelles technologies fonctionnent avec de l'ADN de grande taille, supérieur à 50 kb, ce qui nécessite des méthodes de préparation d'échantillon spécifiques. Nous avons décrit l'état de l'art des méthodes d'extraction et de caractérisation d'ADN de haut poids moléculaire et constaté que certaines étapes n'étaient pas traitées de manière satisfaisante avec les outils du marché. L'objectif de cette thèse était de répondre à ces besoins.

D'un point de vue opérationnel, nous nous sommes appuyés sur la technologie  $\mu$ LAS, permettant la concentration et la séparation d'ADN lors de l'application bidirectionnelle d'une électrophorèse et d'un flux laminaire. Nous avons d'abord optimisé les différents paramètres régissant la migration d'ADN avec  $\mu$ LAS afin d'identifier des conditions opératoires favorables à la manipulation de grands fragments d'ADN de 10 à 200 kb. Nous avons ensuite démontré des séparations jusqu'à 200 kb avec des sensibilités de l'ordre de la dizaine de fg/ $\mu$ L.

Ces excellentes performances analytiques nous ont permis d'entreprendre deux collaborations. Une première avec le CNRGV visant à isoler spécifiquement un locus génomique de 30 kb dans le génome de la plante *Medicago Truncatula* découpé avec la technologie CRISPR/Cas9. Nous avons montré qu'en complément d'un gel d'électrophorèse en champs pulsés, la technologie  $\mu$ LAS permettait de d'isoler efficacement ce locus d'intérêt du reste du génome. Le séquençage de ce fragment par la technologie PacBio a permis l'assemblage de cette région génomique de 30kb, une première avec cette méthode de capture. Ces résultats font l'objet d'une publication en cours de soumission chez *Nucleic Acid Research*.

Dans cette étude, nous avons montré la possibilité d'isoler et de séquencer un fragment d'une taille donnée dans un échantillon complexe. Notre approche comporte cependant certaines limites. En effet, nous n'avons jamais observé le fragment d'intérêt directement après l'électroélution du plug d'agarose. Il est donc impossible de collecter le fragment d'intérêt sans connaître au préalable sa taille. Or, l'objectif du projet est de mieux caractériser des régions génomiques inconnues, donc de tailles incertaines. Nous ne sommes pas capables de dire si cette limitation tient à la manipulation et la fragilité des génomes de plantes ou si c'est un phénomène général intervenant lors de l'extraction du

génomique de tout organisme eucaryote. Une des pistes d'amélioration tient à limiter le nombre d'opérations lors des étapes d'extraction et de digestion avec Cas9. Nous pensons donc qu'il faut travailler sur (i) la simplification des procédures de collecte du fragment d'intérêt et (ii) l'augmentation des quantités collectées au sortir du système. Sur le premier volet, la sensibilité de  $\mu$ LAS autorise une caractérisation systématique des échantillons, même avec des concentrations résiduelles, ce qui ouvre la voie à une optimisation rigoureuse de toute la chaîne de purification. Le second objectif consiste à éviter toute étape d'amplification qui engendre la formation de chimères et complexifie l'analyse bio-informatique. L'utilisation du dispositif multicapillaire pourrait être utile, en particulier si la technologie Chromium de 10X Genomics, qui nécessite peu d'ADN, se généralise.

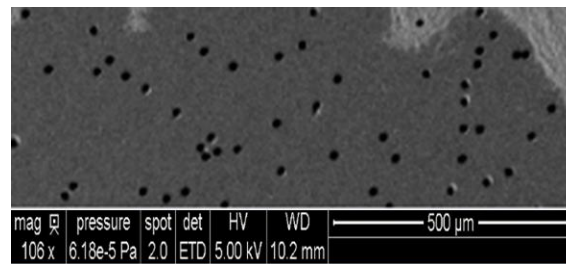
Notre seconde collaboration, avec la plateforme de séquençage de Toulouse (GeT-PlaGe), concerne le développement d'une méthode permettant de sélectionner en taille les fragments d'ADN d'un échantillon complexe afin d'isoler les plus grands fragments pour les séquencer. Les technologies actuelles pour cette sélection souffrent en effet d'inconvénients, en particulier en termes de rendement. Nous avons ainsi développé une vanne accordable, uniquement via la modulation du champ électrique, permettant la sélection d'une taille d'ADN précise dans un échantillon avec une large distribution en taille. Cette vanne fonctionne bien dans une gamme de taille de 200 bp à 50 kb mais pour des quantités de matière inférieures à 200 pg environ. Elle n'est donc pas compatible avec les spécifications des séquenceurs de 3<sup>ème</sup> génération. Pour augmenter les quantités traitées, nous avons développé un prototype fonctionnant avec une fibre multicapillaire pouvant purifier jusqu'à 4 ng d'ADN. Ces quantités étant en adéquation avec la technologie Chromium de 10X Genomics, nous avons sélectionné 17 ng d'ADN de melon au-dessus de 20 kb afin de comparer notre méthode à l'existant. Les différents échantillons sont en cours de séquençage.

Même si les résultats de séquençage sont bons, notre instrument reste limité du point de vue de la quantité de matière traitée. Nous parvenons à correspondre aux spécifications de la technologie de 10X Genomics mais nous sommes très loin des  $\mu$ g requis pour un séquençage Nanopore ou PacBio. Nous avons néanmoins démontré la possibilité de sélectionner en taille un échantillon d'ADN grâce à la technologie  $\mu$ LAS dans des temps d'analyse et avec des rendements très prometteurs.

## **Perspectives**

Nous avons vu que la vanne accordable pour la sélection en taille d'ADN était une approche prometteuse mais que les formats capillaire et multicapillaire étaient limitant en termes de quantités de matière purifiées. Nous réfléchissons donc à un nouveau format de technologie pour purifier de plus grandes quantités d'ADN en utilisant des membranes isopores (Figure 36). Les membranes isopores sont disponibles commercialement et obtenues par gravure de feuille de polymère aux ions

lourds. Chaque pore agirait ainsi comme un canal microfluidique à l'intérieur duquel nous pouvons effectuer une purification de l'ADN grâce au principe de vanne accordable décrit dans le chapitre 5. De nombreuses références de membranes sont disponibles avec des diamètres et des longueurs de pores de 1 à 5  $\mu\text{m}$  et de 20 à 100  $\mu\text{m}$  respectivement. Nous espérons que la forte densité de pore par membrane permette de traiter de plus grandes quantités de matière et ainsi correspondre aux spécifications des séquenceurs de 3<sup>ème</sup> génération.

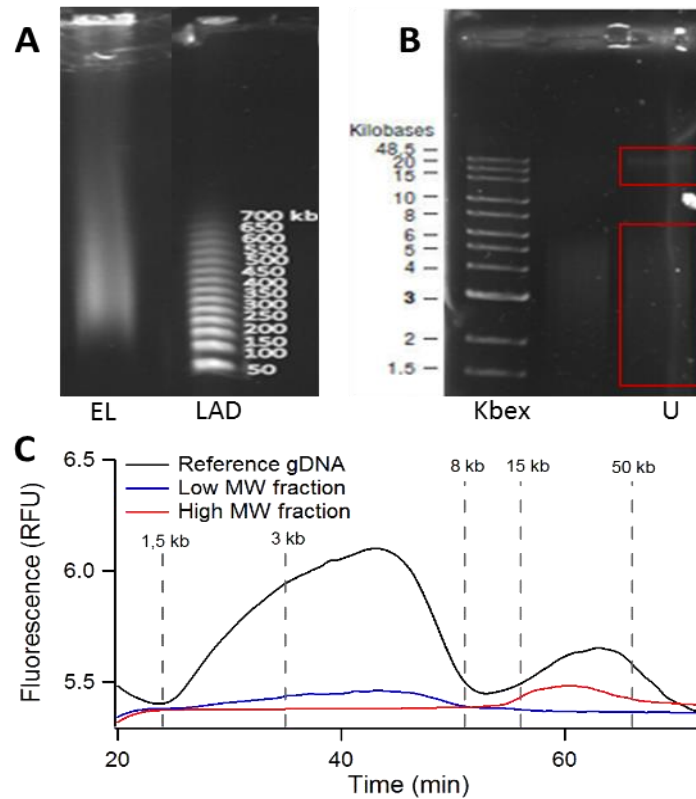


**Figure 36 : Image en microscopie électronique à balayage d'une membrane isopore**

En plus de l'augmentation du débit par l'augmentation du nombre de trous, ce nouveau format présente un avantage supplémentaire par rapport au format capillaire. En effet, les membranes isopores permettent de confiner le champ électrique dans les trous des membranes, c'est-à-dire à l'endroit même où la sélection en taille doit être réalisée, et ainsi de limiter les tensions appliquées. Au contraire, la longueur des capillaires induit de forts voltages et donc des zones d'échauffement importantes.

Cette vanne de sélection d'ADN pourrait également servir à purifier des échantillons après l'étape de lyse cellulaire lors d'une extraction d'ADN. En effet, tous les échantillons d'ADN que nous avons manipulés dans cette thèse étaient déjà purifiés, c'est-à-dire avec peu ou pas de sels et sans protéines. Or, la vanne de sélection  $\mu\text{LAS}$  agit comme un système de dialyse, où seules les molécules d'ADN suffisamment grandes sont retenues dans le concentrateur. Nous pouvons ainsi utiliser cette fonction après l'étape de lyse cellulaire pour éliminer les sels, protéines et contaminants cellulaires tout en réalisant la sélection en taille. Des tests préliminaires ont été effectués sur de l'ADN extrait de cellules SH-SY5Y en culture. Les cellules ont été lysées avec un protocole classique à base de SDS et de protéinase K (voir Annexe C). Une analyse en PFGE de ces échantillons montre que l'ADN est peu dégradé et majoritairement entre 100 kb et 1 Mb (Figure 37, A). Afin de mettre en évidence le bon fonctionnement de la vanne de sélection avec un échantillon non purifié, nous avons fragmenté cet ADN extrait par une exposition aux ultrasons (sonication). L'ADN obtenu présente alors une distribution en taille bimodale avec un pic large entre 1,5 et 8 kb et un second entre 15 et 100 kb (Figure 37, B). Nous avons ainsi effectué une sélection en taille sur cet échantillon non purifié afin de collecter dans un tube la fraction de bas poids moléculaire et dans un autre celle de haut poids

moléculaire. Ces deux fractions ont ensuite été analysées grâce à la méthode d'analyse  $\mu$ LAS décrite en 3.3.3 et comparé à l'échantillon non purifié (Figure 37, C).



**Figure 37 : Sélection en taille avec la vanne accordable sur un échantillon d'ADN non purifié. A :** PFGE d'une extraction de cellules SH-SY5Y en phase liquide. (EL) Extraction liquide. (Lad) Référence de taille **B :** Gel d'électrophorèse de l'extraction A après sonication. (U) Echantillon après sonication. Les rectangles rouges indiquent les traces d'une dégradation. (Kbex) Référence de taille. **C :** Analyse  $\mu$ LAS des collectes après purification. La courbe noire correspond à l'échantillon fragmenté avant purification. Les courbes bleue et rouge correspondent à la fraction de bas poids moléculaire et la fraction de haut poids moléculaire respectivement après la sélection en taille sur le dispositif multicapillaire. Les lignes grises indiquent les tailles de référence.

La sélection en taille a ainsi parfaitement fonctionné sur un échantillon d'ADN non purifié. La caractérisation de la pureté de l'échantillon obtenu reste complexe à l'heure actuelle car les quantités de matière traitées dans cette expérience sont faibles, de l'ordre de 100 pg/ $\mu$ L. Or, nous avons vu au chapitre 2 que le seul instrument permettant d'évaluer la pureté d'un échantillon était le Nanodrop, dont la limite de sensibilité est de l'ordre de 1 ng/ $\mu$ L (voir 2.3.1). Si le développement des membranes isopores est prometteur, nous serons en mesure de mieux caractériser l'élimination des contaminants cellulaire. Nous pouvons toutefois envisager le développement d'un instrument qui permettrait d'extraire l'ADN directement en phase liquide depuis le matériel biologique. Puis ce lysat serait purifié et sélectionné en taille dans le même instrument par la vanne de sélection  $\mu$ LAS et accélérerait ainsi la préparation des échantillons pour le séquençage.

Le développement d'un tel instrument permettrait également de simplifier, d'automatiser et de standardiser les protocoles actuels d'extraction d'ADN circulant dans le plasma sanguin<sup>193</sup>. L'analyse de l'ADN circulant dans le sang est un domaine en plein essor et il semble être un biomarqueur prometteur pour différentes pathologies<sup>194</sup>. La technologie de vanne de sélection  $\mu$ LAS semble bien adaptée à l'extraction de cet ADN circulant, d'autant que des études récentes montrent que la taille de l'ADN circulant est un élément important à considérer pour les analyses personnalisées<sup>195</sup>.

D'un point de vue analytique, enfin, il semble intéressant d'étendre encore la gamme d'ADN à séparer avec  $\mu$ LAS, au-delà de 200 kb. Afin d'améliorer la résolution des séparations, il est nécessaire d'augmenter les différences de mobilité entre les fragments tout en limitant leur dispersion. Un modèle prédictif et prospectif décrivant la migration  $\mu$ LAS des fragments d'ADN est en développement au LAAS. Ce modèle devrait s'attacher à décrire le transport des molécules mais également leur dispersion. Nos séparations ont la particularité de plaquer les molécules d'ADN aux parois ce qui modifie les phénomènes de dispersion classique dans un écoulement de Poiseuille<sup>196</sup>. Ce modèle permettra également de faciliter le choix des conditions d'actionnement (pression, tension, tampon) pour une séparation optimale dans une gamme de taille donnée. Il serait également un outil de choix pour l'évaluation en taille d'un échantillon inconnu, actuellement faite par comparaison des temps de migration à des fragments calibrés.



## Références

---

1. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **74**, 5463–5467 (1977).
2. Karger, B. L. & Guttman, A. DNA sequencing by CE. *ELECTROPHORESIS* **30**, S196–S202 (2009).
3. Wetterstrand, K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: <https://www.genome.gov/27541954/dna-sequencing-costs-data/>.
4. Sedlazeck, F. J., Lee, H., Darby, C. A. & Schatz, M. C. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics* **1** (2018).
5. van Daelen, R. A., Jonkers, J. J. & Zabel, P. Preparation of megabase-sized tomato DNA and separation of large restriction fragments by field inversion gel electrophoresis (FIGE). *Plant molecular biology* **12**, 341–352 (1989).
6. Ranchon, H. *et al.* DNA separation and enrichment using electro-hydrodynamic bidirectional flows in viscoelastic liquids. *Lab on a Chip* **16**, 1243–1253 (2016).
7. Bancaud, A. & Ranchon, H. Procédé de séparation de molécules et cellules biologiques en solution. (2014).
8. Bancaud, A., Ranchon, H., Leichle, T. & Pattamon, T. Procédé et dispositif de concentration de molécules ou objets dissous en solution. (2016).
9. Malbec, R. Système microfluidique  $\mu$ LAS pour l'analyse de l'ADN résiduel: Application au diagnostic de la maladie de Huntington et à l'analyse de l'ADN circulant dans le sang. (INSA de Toulouse, 2018).
10. Andriamanampisoa, C.-L. *et al.* BIABooster: Online DNA Concentration and Size Profiling with a Limit of Detection of 10 fg/ $\mu$ L and Application to High-Sensitivity Characterization of Circulating Cell-Free DNA. *Analytical Chemistry* **90**, 3766–3774 (2018).
11. Hagemann, I. S. Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. in *Clinical Genomics* 3–19 (Elsevier, 2015). doi:10.1016/B978-0-12-404748-8.00001-0



12. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends in Genetics* **30**, 418–426 (2014).
13. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Archives of disease in childhood - Education & practice edition* **98**, 236–238 (2013).
14. McGinn, S. & Gut, I. G. DNA sequencing – spanning the generations. *New Biotechnology* **30**, 366–372 (2013).
15. Whitmore, John. NGS market SHare. Available at: <https://www.openpr.com/news/492470/Next-Generation-Sequencing-Market-Assessment-2017-2021.html>.
16. Illumina. NovaSeq platform. *Illumina website* Available at: <https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html>.
17. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research* **40**, e72–e72 (2012).
18. Li, R. *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20**, 265–272 (2010).
19. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics* **14**, 125–138 (2013).
20. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The Third Revolution in Sequencing Technology. *Trends in Genetics* **34**, 666–681 (2018).
21. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Human Molecular Genetics* **19**, R227–R240 (2010).
22. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333–351 (2016).
23. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
24. Levene, M. J. *et al.* Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations.pdf. *Science* **299**, 682–686 (2003).

25. Schadt, E. E. *et al.* Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Research* **23**, 129–141 (2013).
26. Chaisson, M. J. P., Wilson, R. K. & Eichler, E. E. Genetic variation and the de novo assembly of human genomes. *Nature Reviews Genetics* **16**, 627–640 (2015).
27. Detter, J. C. *et al.* Nucleic acid sequencing for characterizing infectious and/or novel agents in complex samples. in *Biological Identification* 3–53 (Elsevier, 2014). doi:10.1533/9780857099167.1.3
28. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **13**, 278–289 (2015).
29. Chaisson, M. J. P. *et al.* Resolving the complexity of the human genome using single-molecule sequencing. *Nature* **517**, 608–611 (2015).
30. Seo, J.-S. *et al.* De novo assembly and phasing of a Korean human genome. *Nature* **538**, 243–247 (2016).
31. Ardui, S., Ameer, A., Vermeesch, J. R. & Hestand, M. S. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research* **46**, 2159–2168 (2018).
32. Korlach, J. *et al.* Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences* **105**, 1176–1181 (2008).
33. Fang, H. *et al.* Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome medicine* **6**, 89 (2014).
34. Smith, D. E., Perkins, T. T. & Chu, S. Dynamical scaling of DNA diffusion coefficients. *Macromolecules* **29**, 1372–1373 (1996).
35. Han, J. & Craighead, H. G. Separation of long DNA molecules in a microfabricated entropic trap array. *Science* **288**, 1026–1029 (2000).

36. Larkin, J., Henley, R. Y., Jadhav, V., Korlach, J. & Wanunu, M. Length-independent DNA packing into nanopore zero-mode waveguides for low-input DNA sequencing. *Nature Nanotechnology* **12**, 1169–1175 (2017).
37. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, (2016).
38. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* **4**, 265–270 (2009).
39. Rang, F. J., Kloosterman, W. P. & de Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology* **19**, (2018).
40. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
41. Castro-Wallace, S. L. *et al.* Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Scientific Reports* **7**, (2017).
42. Zhao, Y. *et al.* Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nature Nanotechnology* **9**, 466–473 (2014).
43. Kolmogorov, M., Kennedy, E., Dong, Z., Timp, G. & Pevzner, P. A. Single-molecule protein identification by sub-nanopore sensors. *PLOS Computational Biology* **13**, e1005356 (2017).
44. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* **36**, 338–345 (2018).
45. Jain, M. *et al.* Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology* **36**, 321–323 (2018).
46. An, N., Fleming, A. M., White, H. S. & Burrows, C. J. Nanopore Detection of 8-Oxoguanine in the Human Telomere Repeat Sequence. *ACS Nano* **9**, 4296–4307 (2015).
47. Jain, M. *et al.* release and analysis of R9.0 chemistry [version 1; referees. (2017).
48. Travers, K., Otto, G., Turner, S., Heiner, C. & Ma, C. Compositions and methods for nucleic acid sequencing. (2008).

49. Tyler, A. D. *et al.* Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports* **8**, (2018).
50. Pollack, M. G., Pamula, V. K., Srinivasan, V. & Eckhardt, A. E. Applications of electrowetting-based digital microfluidics in clinical diagnostics. *Expert Review of Molecular Diagnostics* **11**, 393–407 (2011).
51. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201 (2015).
52. Marks, P., Garcia, S. & Chruch, D. Resolving the Full Spectrum of Human genome using linked-reads.pdf. *BioRxiv* (2017). doi:<http://dx.doi.org/10.1101/230946>
53. Kuleshov, V., Snyder, M. P. & Batzoglou, S. Genome assembly from synthetic long read clouds. *Bioinformatics* **32**, i216–i224 (2016).
54. Zheng, G. X. Y. *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology* **34**, 303–311 (2016).
55. Porubsky, D. *et al.* Dense and accurate whole-chromosome haplotyping of individual genomes. *Nature Communications* **8**, (2017).
56. Wong, K. H. Y., Levy-Sakin, M. & Kwok, P.-Y. De novo human genome assemblies reveal spectrum of alternative haplotypes in diverse populations. *Nature Communications* **9**, (2018).
57. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Research* **27**, 757–767 (2017).
58. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
59. Bansal, V. A computational method for estimating the PCR duplication rate in DNA and RNA-seq experiments. *BMC Bioinformatics* **18**, (2017).
60. Bouchez, O. & Kuchly, C. Implementation and Evaluation of 10X Genomics Chromium technology.
61. Ramnarain, S. P., Huff, E. J. & Wangf, Y.-K. Ordered Restriction Maps of *Saccharomyces cerevisiae* Chromosomes Constructed by Optical Mapping. *Biol* **10**, 6417 (1990).

62. Cao, H., Tegenfeldt, J. O., Austin, R. H. & Chou, S. Y. Gradient nanostructures for interfacing microfluidics and nanofluidics. *Applied Physics Letters* **81**, 3058–3060 (2002).
63. Tegenfeldt, J. O. *et al.* The dynamics of genomic-length DNA molecules in 100-nm channels. *Proceedings of the National Academy of Sciences* **101**, 10979–10983 (2004).
64. Sheats, J., Reifenberger, J. G., Cao, H. & Dorfman, K. D. Measurements of DNA barcode label separations in nanochannels from time-series data. *Biomicrofluidics* **9**, 064119 (2015).
65. Zohar, H. & Muller, S. J. Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. *Nanoscale* **3**, 3027 (2011).
66. Teague, B. *et al.* High-resolution human genome structure by single-molecule analysis. *Proceedings of the National Academy of Sciences* (2010).
67. Lam, E. T. *et al.* Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nature Biotechnology* **30**, 771–776 (2012).
68. Larson, J. W. *et al.* Single DNA molecule stretching in sudden mixed shear and elongational microflows. *Lab Chip* **6**, 1187–1199 (2006).
69. Herrick, J. & Bensimon, A. Introduction to molecular combing: genomics, DNA replication, and cancer. in *DNA replication* 71–101 (Springer, 2009).
70. Michaeli, Y. & Ebenstein, Y. Channeling DNA for optical mapping. *Nature Biotechnology* **30**, 762–763 (2012).
71. Fang Lim, S. *et al.* DNA methylation profiling in nanochannels. *Biomicrofluidics* **5**, 034106 (2011).
72. Matsuoka, T. *et al.* Nanoscale Squeezing in Elastomeric Nanochannels for Single Chromatin Linearization. *Nano Letters* **12**, 6480–6484 (2012).
73. Jiao, W.-B. & Schneeberger, K. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36**, 64–70 (2017).
74. McCaffrey, J. *et al.* CRISPR-CAS9 D10A nickase target-specific fluorescent labeling of double strand DNA for whole genome mapping and structural variation analysis. *Nucleic Acids Research* **44**, e11–e11 (2016).

75. Zhang, D. *et al.* CRISPR-bind: a simple, custom CRISPR/dCas9-mediated labeling of genomic DNA for mapping in nanochannel arrays. *bioRxiv* 371518 (2018).
76. Lee, H. *et al.* Third-generation sequencing and the future of genomics. *BioRxiv* 048603 (2016).
77. Mayjonade, B. *et al.* Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *BioTechniques* **61**, (2016).
78. Shehadul Islam, M., Aryasomayajula, A. & Selvaganapathy, P. A Review on Macroscale and Microscale Cell Lysis Methods. *Micromachines* **8**, 83 (2017).
79. Kirby, K. New method for the isolation of deoxyribonucleic acids; evidence on the nature of bonds between deoxyribonucleic acid and protein.pdf. *Biochemical Journal* **66**, 495–504 (1957).
80. Green, M. R. & Sambrook, J. Isolation of High-Molecular-Weight DNA from Mammalian Blood Using Proteinase K and Phenol. *Cold Spring Harbor Protocols* **2017**, pdb.prot093492 (2017).
81. Chomczynski, P. & Sacchi, N. The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nature Protocols* **1**, 581–585 (2006).
82. Payne, A., Holmes, N., Rakyan, V. & Loose, M. Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *bioRxiv* 312256 (2018).
83. Weissensteiner, M. H. *et al.* Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Research* **27**, 697–708 (2017).
84. Formenti, G. *et al.* SMRT long-read sequencing and Direct Label and Stain optical maps allow the generation of a high-quality genome assembly for the European barn swallow (*Hirundo rustica rustica*). *bioRxiv* 374512 (2018).
85. Mariano, D. C. B. *et al.* Whole-genome optical mapping reveals a mis-assembly between two rRNA operons of *Corynebacterium pseudotuberculosis* strain 1002. *BMC Genomics* **17**, (2016).

86. Healey, A., Furtado, A., Cooper, T. & Henry, R. J. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant methods* **10**, 21 (2014).
87. Akerman, B. & Tuite, E. Single- and double-strand photocleavage of DNA by YO, YOYO and TOTO. *Nucleic acids research* **24**, 1080–1090 (1996).
88. Garger, S. J., Griffith, O. M. & Grill, L. K. Rapid purification of plasmid DNA by a single centrifugation in a two-step cesium chloride-ethidium bromide gradient. *Biochemical and biophysical research communications* **117**, 835–842 (1983).
89. Price, C. W., Leslie, D. C. & Landers, J. P. Nucleic acid extraction techniques and application to the microchip. *Lab on a Chip* **9**, 2484 (2009).
90. Vogelstein, B. & Gillespie, D. Preparative and analytical purification of DNA from agarose. *Proceedings of the National Academy of Sciences* **76**, 615–619 (1979).
91. Langrau, N., Oelmiiller, U., Weber, M., Fuhrmann, G. & Schorr, J. Fuel cell membrane humidification. (2000).
92. Christel, L. A., Petersen, K., McMillan, W. & Northrup, M. A. Rapid, automated nucleic acid probe assays using silicon microstructures for nucleic acid concentration. *Journal of biomechanical engineering* **121**, 22–27 (1999).
93. Ferrarini, M. *et al.* An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC genomics* **14**, 670 (2013).
94. Bordelon, H., Russ, P. K., Wright, D. W. & Haselton, F. R. A Magnetic Bead-Based Method for Concentrating DNA from Human Urine for Downstream Detection. *PLoS ONE* **8**, e68369 (2013).
95. Berensmeier, S. Magnetic particles for the separation and purification of nucleic acids. *Applied Microbiology and Biotechnology* **73**, 495–504 (2006).
96. Zook, J. M. *et al.* Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data* **3**, 160025 (2016).

97. Gong, L. *et al.* Nanopore Sequencing Reveals High-Resolution Structural Variation in the Cancer Genome. *bioRxiv* 209718 (2017).
98. Zhang, Y. *et al.* A Simple Thermoplastic Substrate Containing Hierarchical Silica Lamellae for High-Molecular-Weight DNA Extraction. *Advanced Materials* **28**, 10630–10636 (2016).
99. Nanobind High Quality DNA/RNA Sample Preparation. *Circulomics* Available at: <https://www.circulomics.com/nanobind>.
100. Heiner, C., Kim, M., Burke, J. M., Bjornson, K. & Liu, K. J. Single Chromosomal Genome Assemblies on the Sequel System with Circulomics High Molecular Weight DNA Extraction for Microbes. (2018).
101. Workman, R., Liu, K. & Timp, W. High Molecular Weight DNA Extraction from Recalcitrant Plant Species for Third Generation Sequencing.pdf. (2018).
102. Oblath, E. A., Henley, W. H., Alarie, J. P. & Ramsey, J. M. A microfluidic chip integrating DNA extraction and real-time PCR for the detection of bacteria in saliva. *Lab on a Chip* **13**, 1325 (2013).
103. Kim, J., Johnson, M., Hill, P. & Gale, B. K. Microfluidic sample preparation: cell lysis and nucleic acid purification. *Integrative Biology* **1**, 574 (2009).
104. Benítez, J. J. *et al.* Microfluidic extraction, stretching and analysis of human chromosomal DNA from single cells. *Lab on a Chip* **12**, 4848 (2012).
105. Snider, A., Nilsson, M., Dupal, M., Toloue, M. & Tripathi, A. A Microfluidics Workflow for Sample Preparation for Next-Generation DNA Sequencing. *SLAS TECHNOLOGY: Translating Life Sciences Innovation* 2472630318796133 (2018).
106. Kim, S. *et al.* High-throughput automated microfluidic sample preparation for accurate microbial genomics. *Nature Communications* **8**, 13919 (2017).
107. Bienvenue, J. M., Duncalf, N., Marchiarullo, D., Ferrance, J. P. & Landers, J. P. Microchip-Based Cell Lysis and DNA Extraction from Sperm Cells for Application to Forensic Analysis. *Journal of Forensic Sciences* **51**, 266–273 (2006).



108. Breadmore, M. C. *et al.* Microchip-Based Purification of DNA from Biological Samples. *Analytical Chemistry* **75**, 1880–1886 (2003).
109. Zhang, X., Wu, X., Peng, R. & Li, D. Electromagnetically controlled microfluidic chip for DNA extraction. *Measurement* **75**, 23–28 (2015).
110. Lehmann, U., Vandevyver, C., Parashar, V. K. & Gijs, M. A. M. Droplet-Based DNA Purification in a Magnetic Lab-on-a-Chip. *Angewandte Chemie International Edition* **45**, 3062–3067 (2006).
111. Reedy, C. R. *et al.* Solid phase extraction of DNA from biological samples in a post-based, high surface area poly(methyl methacrylate) (PMMA) microdevice. *Lab on a Chip* **11**, 1603 (2011).
112. Chen, Z. & Dorfman, K. D. Relationship between frequency and deflection angle in the DNA prism. *Physical Review E* **87**, 012723 (2013).
113. Overhauser, J. Encapsulation of cells in agarose beads. in *Pulsed-Field Gel Electrophoresis* 129–134 (Springer, 1992).
114. Nair, S., Karim, R., Cardosa, M. J., Ismail, G. & Pang, T. Convenient and versatile DNA extraction using agarose plugs for ribotyping of problematic bacterial species. *Journal of microbiological methods* **38**, 63–67 (1999).
115. Zhang, M. *et al.* Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nature Protocols* **7**, 467–478 (2012).
116. Teo, A. S. M., Verzotto, D., Yao, F., Nagarajan, N. & Hillmer, A. M. Single-molecule optical genome mapping of a human HapMap and a colorectal cancer cell line. *GigaScience* **4**, (2015).
117. Gaiero, P. *et al.* Intact DNA purified from flow-sorted nuclei unlocks the potential of next-generation genome mapping and assembly in *Solanum* species. *MethodsX* **5**, 328–336 (2018).
118. Marziali, A., Pel, J., Bizzotto, D. & Whitehead, L. A. Novel electrophoresis mechanism based on synchronous alternating drag perturbation. *ELECTROPHORESIS* **26**, 82–90 (2005).
119. Pel, J. *et al.* Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules. *Proceedings of the National Academy of Sciences* **106**, 14796–14801 (2009).

120. Maydan, J., Thomas, M., Tabanfar, L., Mai, L. & Marziali, A. Electrophoretic High Molecular Weight DNA Purification for Optical Mapping. (2009).
121. Datema, E. *et al.* The megabase-sized fungal genome of *Rhizoctonia solani* assembled from nanopore reads only. *bioRxiv* 084772 (2016).
122. Persat, A., Marshall, L. A. & Santiago, J. G. Purification of Nucleic Acids from Whole Blood Using Isotachopheresis. *Analytical Chemistry* **81**, 9507–9511 (2009).
123. Becker, L., Steglich, M., Fuchs, S., Werner, G. & Nübel, U. Comparison of six commercial kits to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. *Scientific Reports* **6**, (2016).
124. Uyaguari-Diaz, M. I. *et al.* Automated Gel Size Selection to Improve the Quality of Next-generation Sequencing Libraries Prepared from Environmental Water Samples. *Journal of Visualized Experiments* (2015). doi:10.3791/52685
125. Ali, N., Rampazzo, R. de C. P., Costa, A. D. T. & Krieger, M. A. Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics. *BioMed Research International* **2017**, 1–13 (2017).
126. Quail, M. A., Swerdlow, H. & Turner, D. J. Improved Protocols for the Illumina Genome Analyzer Sequencing System. in *Current Protocols in Human Genetics* (eds. Haines, J. L. et al.) (John Wiley & Sons, Inc., 2009). doi:10.1002/0471142905.hg1802s62
127. Quail, M. A., Gu, Y., Swerdlow, H. & Mayho, M. Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation: Nucleic Acids. *ELECTROPHORESIS* **33**, 3521–3528 (2012).
128. NIMBUS Select spécifications. *Hamilton website* Available at: <https://www.hamiltoncompany.com/automated-liquid-handling/assay-ready-workstations/nimbus-select#specifications>.
129. PacBlo & SageScience. 20 kb Template Preparation Using BluePippin™ Size-Selection System. (2016).
130. 10X Genomics & SageScience. DNA size selection protocol.pdf. (2016).

131. Oxford Nanopore Technologies. Sequencing library preparation for MinION and PromethION.pdf. (2017).
132. Cretu Stancu, M. *et al.* Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications* **8**, (2017).
133. Portmann, A.-C. *et al.* A Validation Approach of an End-to-End Whole Genome Sequencing Workflow for Source Tracking of *Listeria monocytogenes* and *Salmonella enterica*. *Frontiers in Microbiology* **9**, (2018).
134. Cartwright, J., Anderson, K., Longworth, J., Lobb, P. & David, J. Highly sensitive detection of mutations in CHO cell recombinant DNA using multi-parallel single molecule real-time DNA sequencing.pdf. *Biotechnology and Bioengineering* **115**, 1485–1498 (2018).
135. Tan, M. H. *et al.* Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *GigaScience* **7**, (2018).
136. Liu, Q., Chang, S., Hartman, G. L. & Domier, L. L. Assembly and annotation of a draft genome sequence for *Glycine latifolia* , a perennial wild relative of soybean. *The Plant Journal* **95**, 71–85 (2018).
137. Boesenberg-Smith, K. A., Pessaraki, M. M. & Wolk, D. M. Assessment of DNA yield and purity: an overlooked detail of PCR troubleshooting. *Clinical Microbiology Newsletter* **34**, 1–6 (2012).
138. Desjardins, P. & Conklin, D. NanoDrop Microvolume Quantitation of Nucleic Acids. *Journal of Visualized Experiments* (2010). doi:10.3791/2565
139. Green, M. R. & Sambrook, J. Isolation and Quantification of DNA. *Cold Spring Harbor Protocols* **2018**, pdb.top093336 (2018).
140. Mansfield, E. S. *et al.* Nucleic acid detection using non-radioactive labelling methods. *Molecular and cellular probes* **9**, 145–156 (1995).
141. Hilario, E. End labeling procedures. *Molecular biotechnology* **28**, 77–80 (2004).
142. Rengarajan, K., Cristol, S. M., Mehta, M. & Nickerson, J. M. Quantifying DNA concentrations using fluorometry: a comparison of fluorophores. *Mol Vis* **8**, 416–421 (2002).

143. Piehler, J. *et al.* Label-free monitoring of DNA–ligand interactions. *Analytical biochemistry* **249**, 94–102 (1997).
144. Qubit™ dsDNA HS Assay Kit. Available at: <http://www.thermofisher.com/order/catalog/product/Q32851>.
145. Lerman, L. S. & Frisch, H. L. Why does the electrophoretic mobility of DNA in gels vary with the length of the molecule? *Biopolymers: Original Research on Biomolecules* **21**, 995–997 (1982).
146. Viovy, J.-L. Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms. *Reviews of Modern Physics* **72**, 813 (2000).
147. Lee, P. Y., Costumbrado, J., Hsu, C.-Y. & Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments* (2012). doi:10.3791/3923
148. Schwartz, D. C. & Cantor, C. R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *cell* **37**, 67–75 (1984).
149. Birren, B. & Lai, E. Rapid pulsed field separation of DNA molecules up to 250 kb. *Nucleic acids research* **22**, 5366–5370 (1994).
150. Magnúsdóttir, S., Isambert, H., Heller, C. & Viovy, J.-L. Electrohydrodynamically Induced Aggregation During Constant and Pulsed Field Capillary Electrophoresis of DNA. *Biopolymers* **49**, 385–401 (1999).
151. Kim, Y. & Morris, M. D. Ultrafast high resolution separation of large DNA fragments by pulsed-field capillary electrophoresis. *Electrophoresis* **17**, 152–160 (1996).
152. Heller, C., Magnúsdóttir, S. & Viovy, J.-L. Robust field inversion capillary electrophoretic separation of long DNA fragments. *Capillary Electrophoresis of Nucleic Acids: Volume I: Introduction to the Capillary Electrophoresis of Nucleic Acids* 293–305 (2001).
153. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).
154. Dorfman, K. D. DNA electrophoresis in microfabricated devices. *Reviews of Modern Physics* **82**, 2903–2947 (2010).

155. Doyle, P. S., Bibette, J., Bancaud, A. & Viovy, J.-L. Self-assembled magnetic matrices for DNA separation chips. *Science* **295**, 2237–2237 (2002).
156. Bakajin, O. *et al.* Separation of 100-kilobase DNA molecules in 10 seconds. *Analytical chemistry* **73**, 6053–6056 (2001).
157. Dorfman, K. D., King, S. B., Olson, D. W., Thomas, J. D. P. & Tree, D. R. Beyond Gel Electrophoresis: Microfluidic Separations, Fluorescence Burst Analysis, and DNA Stretching. *Chemical Reviews* **113**, 2584–2667 (2013).
158. Huang, L. R. *et al.* A DNA prism for high-speed continuous fractionation of large DNA molecules. *Nature Biotechnology* **20**, 1048–1051 (2002).
159. Gumuscu, B., Bomer, J. G., de Boer, H. L., van den Berg, A. & Eijkel, J. C. T. Exploiting biased reptation for continuous flow preparative DNA fractionation in a versatile microfluidic platform. *Microsystems & Nanoengineering* **3**, 17001 (2017).
160. Superose 6 Increase small-scale size exclusion chromatography columns. Available at: <https://www.gelifesciences.com/en/es/shop/chromatography/prepacked-columns/size-exclusion/superose-6-increase-small-scale-size-exclusion-chromatography-columns-p-05803>.
161. Ranchon, H. Développement d'outils analytiques par et pour la microfluidique: caractérisation d'écoulements d'objets dissous et intégration d'un système de séparation sans matrice de biomolécules. (Université de Toulouse, Université Toulouse III-Paul Sabatier, 2013).
162. Barron, A. E., Blanch, H. W. & Soane, D. S. A transient entanglement coupling mechanism for DNA separation by capillary electrophoresis in ultradilute polymer solutions. *Electrophoresis* **15**, 597–615 (1994).
163. Wang, X. *et al.* Free Solution Hydrodynamic Separation of DNA Fragments from 75 to 106 000 Base Pairs in A Single Run. *Journal of the American Chemical Society* **132**, 40–41 (2010).
164. Zheng, J. & Yeung, E. S. Mechanism for the Separation of Large Molecules Based on Radial Migration in Capillary Electrophoresis. *Analytical Chemistry* **75**, 3675–3680 (2003).

165. Graessley, W. W. & Edwards, S. F. Entanglement interactions in polymers and the chain contour concentration.pdf. *Polymer* **22**, 1329–1334 (1981).
166. McFarlane, N. L., Wagner, N. J., Kaler, E. W. & Lynch, M. L. Poly(ethylene oxide) (PEO) and Poly(vinyl pyrrolidone) (PVP) Induce Different Changes in the Colloid Stability of Nanoparticles. *Langmuir* **26**, 13823–13830 (2010).
167. Leshansky, A. M., Bransky, A., Korin, N. & Dinnar, U. Tunable Nonlinear Viscoelastic ‘Focusing’ in a Microfluidic Device. *Physical Review Letters* **98**, (2007).
168. Seo, K. W., Byeon, H. J., Huh, H. K. & Lee, S. J. Particle migration and single-line particle focusing in microscale pipe flow of viscoelastic fluids. *RSC Adv.* **4**, 3512–3520 (2014).
169. Young Kim, J., Won Ahn, S., Sik Lee, S. & Min Kim, J. Lateral migration and focusing of colloidal particles and DNA molecules under viscoelastic flow. *Lab on a Chip* **12**, 2807 (2012).
170. D’Avino, G. & Maffettone, P. L. Particle dynamics in viscoelastic liquids. *Journal of Non-Newtonian Fluid Mechanics* **215**, 80–104 (2015).
171. D’Avino, G., Greco, F. & Maffettone, P. L. Particle Migration due to Viscoelasticity of the Suspending Liquid and Its Relevance in Microfluidic Devices. *Annual Review of Fluid Mechanics* **49**, 341–360 (2017).
172. Chami, B., Socol, M., Manghi, M. & Bancaud, A. Modeling of DNA transport in viscoelastic electro-hydrodynamic flows for enhanced size separation. *Soft Matter* (2018). doi:10.1039/C8SM00611C
173. Guinaudeau, A., Mazières, S., Wilson, D. J. & Destarac, M. Aqueous RAFT/MADIX polymerisation of N-vinyl pyrrolidone at ambient temperature. *Polym. Chem.* **3**, 81–84 (2012).
174. Kang, K., Lee, S. S., Hyun, K., Lee, S. J. & Kim, J. M. DNA-based highly tunable particle focuser. *Nature Communications* **4**, (2013).
175. Butler, J. E., Usta, O. B., Kekre, R. & Ladd, A. J. C. Kinetic theory of a confined polymer driven by an external force and pressure-driven flow. *Physics of Fluids* **19**, 113101 (2007).

176. Liu, L., Guo, Z., Huang, Z., Zhuang, J. & Yang, W. Size-selective separation of DNA fragments by using lysine-functionalized silica particles. *Scientific Reports* **6**, (2016).
177. Simpson, S. L., Quirino, J. P. & Terabe, S. On-line sample preconcentration in capillary electrophoresis. *Journal of Chromatography A* **1184**, 504–541 (2008).
178. Dawod, M. & Chung, D. S. High-sensitivity capillary and microchip electrophoresis using electrokinetic supercharging. *Journal of Separation Science* **34**, 2790–2799 (2011).
179. Xu, Z., Esumi, T., Ikuta, N. & Hirokawa, T. High-sensitive analysis of DNA fragments by capillary gel electrophoresis using transient isotachopheresis preconcentration and fluorescence detection. *Journal of Chromatography A* **1216**, 3602–3605 (2009).
180. Friedrich, S. M., Burke, J. M., Liu, K. J., Ivory, C. F. & Wang, T.-H. Molecular rheotaxis directs DNA migration and concentration against a pressure-driven flow. *Nature Communications* **8**, (2017).
181. Arca, M., Ladd, A. J. C. & Butler, J. E. Electro-hydrodynamic concentration of genomic length DNA. *Soft Matter* **12**, 6975–6984 (2016).
182. Zheng, J. & Yeung, E. S. Anomalous Radial Migration of Single DNA Molecules in Capillary Electrophoresis. *Analytical Chemistry* **74**, 4536–4547 (2002).
183. Montes, R. J., Butler, J. E. & Ladd, A. J. C. Trapping DNA with a high throughput microfluidic device. *ELECTROPHORESIS* (2018). doi:10.1002/elps.201800287
184. Kulkarni, P. & Frommolt, P. Challenges in the Setup of Large-scale Next-Generation Sequencing Analysis Workflows. *Computational and Structural Biotechnology Journal* **15**, 471–477 (2017).
185. Chu, J., Mohamadi, H., Warren, R. L., Yang, C. & Birol, I. Innovations and challenges in detecting long read overlaps: an evaluation of the state-of-the-art. *Bioinformatics* btw811 (2016). doi:10.1093/bioinformatics/btw811
186. Claros, M. G. *et al.* Why Assembling Plant Genome Sequences Is So Challenging. *Biology* **1**, 439–459 (2012).

187. Schatz, M. C., Witkowski, J. & McCombie, W. R. Current challenges in de novo plant genome sequencing and assembly. *Genome biology* **13**, 243 (2012).
188. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**, 182–189 (2009).
189. Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E. & Hendrickson, C. L. Overview of Target Enrichment Strategies: Overview of Target Enrichment Strategies. in *Current Protocols in Molecular Biology* (eds. Ausubel, F. M. *et al.*) 7.21.1-7.21.23 (John Wiley & Sons, Inc., 2015). doi:10.1002/0471142727.mb0721s112
190. Loomis, E. W. *et al.* Sequencing the unsequenceable: Expanded CGG-repeat alleles of the fragile X gene. *Genome Research* **23**, 121–128 (2013).
191. Jia, H., Guo, Y., Zhao, W. & Wang, K. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. *Scientific Reports* **4**, (2015).
192. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.pdf. *Science* **337**, 816–821 (2012).
193. El Messaoudi, S., Rolet, F., Mouliere, F. & Thierry, A. R. Circulating cell free DNA: preanalytical considerations. *Clinica Chimica Acta* **424**, 222–230 (2013).
194. Malbec, R. *et al.* Microfluidics for minute DNA sample analysis: open challenges for genetic testing of cell-free circulating DNA in blood plasma. *Micro and Nano Engineering* (2018). doi:10.1016/j.mne.2018.10.003
195. Aarthy, R., Mani, S., Velusami, S., Sundarsingh, S. & Rajkumar, T. Role of Circulating Cell-Free DNA in Cancers. *Molecular Diagnosis & Therapy* **19**, 339–350 (2015).
196. Bancaud, A., Wagner, G., Dorfman, K. D. & Viovy, J.-L. Measurement of the surface concentration for bioassay kinetics in microchannels. *Analytical chemistry* **77**, 833–839 (2005).



## Références

# ANNEXES

---

## A. Fabrication des puces microfluidiques

Les puces microfluidiques sont composées de canaux gravés dans un substrat de silicium encapsulé par un capot de verre. Elles sont fabriquées dans la salle blanche du LAAS qui fournit un environnement contrôlé. Ce protocole de fabrication est issu des thèses de Rémi Malbec et de Hubert Ranchon<sup>9,161</sup>.

### A.1. Photolithographie optique

La photolithographie est la première étape du processus de fabrication. Un substrat de silicium (dopé p, de 4 pouces de diamètre et  $\sim 500 \mu\text{m}$  d'épaisseur) est nettoyé par un traitement piranhas (bain en solution d'eau oxygénée  $\text{H}_2\text{O}_2$  et acide sulfurique  $\text{H}_2\text{SO}_4$  à 50/50 en volume) et du HDMS (Hexamethyldisilazane) puis est vaporisé pour promouvoir l'adhésion de la résine. Une résine positive ECI 3027 est ensuite enduite sur la surface du substrat à graver et durcie par un recuit. L'épaisseur de la couche de résine dépend de la profondeur des canaux à graver. La résine est insolée à l'aide d'un masque en verre avec des motifs en chrome qui fonctionne comme un pochoir. Les zones de la résine positive exposées aux UV (car non protégées par le chrome de masque) sont polymérisées. Après un nouveau recuit, la résine est développée par dissolution des zones non insolées par un solvant. Les zones du substrat à nu correspondent aux motifs qui vont être gravés. Cette technique convient pour obtenir des résolutions de l'ordre du micron.

### A.2. Gravure du silicium

L'étape de gravure permet de reporter les motifs exposés dans le silicium. Nous utilisons une gravure RIE (Reactive Ion Etching) pour obtenir des parois droites et un fond sans aspérités. Cette technique de gravure consiste à bombarder la surface du matériau de gaz ionisés. La vitesse de gravure est différente pour le silicium ou la résine et la gravure est dite sélective. Le temps de gravure est ajusté de manière à obtenir la profondeur de canal désirée tandis que le reste du substrat est protégé par la résine. Les restes de résine sont retirés par nettoyage au plasma  $\text{O}_2$  et la profondeur des canaux est vérifiée à l'aide d'un profilomètre mécanique.

### A.3. Perçage des puces

L'étape de perçage du silicium permet de créer l'entrée et la sortie du canal qui seront connectées aux réservoirs. Un film sec est laminé sur la surface du substrat pour protéger les motifs des projections. Le perçage est réalisé par sablage. Cette technique consiste à projeter des grains d'oxyde d'aluminium par air comprimé pour traverser l'épaisseur du substrat ( $\sim 0,5\text{mm}$ ) par abrasion, et percer des trous

d'environ 1 mm de diamètre. Le film sec est ensuite retiré par dissolution dans un solvant sous sonication.

#### A.4. Oxydation des puces

L'étape d'oxydation des puces permet d'isoler électriquement les canaux par croissance thermique d'une couche d'oxyde humide ( $\text{SiO}_2$ ) d'environ 400nm à la surface du substrat. L'épaisseur de cette couche est contrôlée par ellipsométrie et la profondeur des canaux n'est pas modifiée car la croissance est homogène sur toute la surface du substrat.

#### A.5. Encapsulation des puces

Les canaux sont encapsulés par soudure anodique d'un capot de verre dopé au sodium de 170  $\mu\text{m}$  d'épaisseur. En appliquant une tension (entre 600 et 800 V) à haute température (370°C), le passage du courant entraîne la migration des ions sodium à la surface du verre et la formation de liaisons covalentes (Si-O-Si) entre le verre et le silicium pour sceller les canaux. Après découpage, les puces microfluidiques sont prêtes à l'emploi.

## B. Calcul des résistances hydraulique et électrique du dispositif capillaire

### B.1. Vitesse d'écoulement dans le dispositif capillaire

On définit la résistance hydraulique  $R_h$  subi par un fluide de viscosité dynamique  $\mu$  lors d'un écoulement dans une canalisation circulaire de rayon  $R$ , de longueur  $L$  par :

$$R_h = \frac{8\mu L}{\pi R^4} \quad (6)$$

Pour un assemblage de capillaire, il suffit de sommer les résistances hydrauliques de chaque capillaire composant l'assemblage pour obtenir la résistance hydraulique du système  $R_T$  :

$$R_T = \sum_{i=1}^n R_{hi} \quad (7)$$

On peut ensuite relier la résistance hydraulique de l'assemblage de capillaire à la différence de pression appliquée  $\Delta P$  et au débit  $Q$  par :

$$R_T = \frac{\Delta P}{Q} \quad (8)$$

On peut ainsi calculer la vitesse d'écoulement moyenne  $V_m$  dans la canalisation en fonction des caractéristiques de la canalisation, de la pression appliquée et de la viscosité du fluide :

$$V_m = \frac{\Delta P R^2}{8\mu L} \quad (9)$$

### B.2. Champ électrique dans le dispositif capillaire

La résistance électrique d'un fluide dans un canal s'écrit :

$$R_e = \rho * \frac{L}{S} \quad (10)$$

Avec  $\rho$  la résistivité du fluide,  $L$  la longueur du canal et  $S$  sa section.

De même que pour le calcul de la vitesse d'écoulement, la résistance électrique totale  $R_{eT}$  s'écrit :

$$R_{eT} = \sum_{i=1}^n R_{ei} \quad (11)$$

La formule qui lie tension et courant dans le fluide est similaire à un système électrique :

$$U = R_{eT}I \quad (11)$$

Avec U la tension appliquée et I l'intensité du courant.

L'intensité de courant dans un système en série permet d'écrire :

$$I = \frac{\Delta U_T}{R_{eT}} = \frac{\Delta U_1}{R_{e1}} = \frac{\Delta U_2}{R_{e2}} = \frac{\Delta U_3}{R_{e3}} \quad (12)$$

D'où :

$$\begin{aligned} \Delta U_1 &= \frac{R_{e1}}{R_{eT}} * \Delta U_T \\ \Delta U_2 &= \frac{R_{e2}}{R_{eT}} * \Delta U_T \\ \Delta U_3 &= \frac{R_{e3}}{R_{eT}} * \Delta U_T \end{aligned} \quad (13)$$

Avec les résistivités  $\rho$  qui s'annulent dans le rapport des résistances.

Ainsi, les champs électriques E se déduisent de la tension dans chaque partie :

$$\begin{aligned} E_1 &= \frac{\Delta U_1}{L_1} \\ E_2 &= \frac{\Delta U_2}{L_2} \\ E_3 &= \frac{\Delta U_3}{L_3} \end{aligned} \quad (14)$$

On peut ensuite déduire la vitesse électrophorétique  $V_e$  à partir de E et  $M_e$  la mobilité électrophorétique :

$$V_e = M_e * E \quad (15)$$

## C. Protocole d'extraction en phase liquide

La lignée cellulaire humaine SH-SY5Y (neuroblastes issus de la moelle osseuse) a été utilisée pour l'extraction de l'ADN génomique. La culture cellulaire a été réalisée sur le milieu DMEM + GlutaMAX (Dulbecco's Modified Eagle Medium, Life technologies) supplémenté avec 10 % de sérum de veau foetal (SVF) et 1X PS (Pénicilline et Streptomycine). Les conditions de culture ont été fixées à 37 °C, dans une atmosphère contrôlée à 5 % de CO<sub>2</sub>. Des flacons de culture T25 ont été utilisés comme support d'adhésion des cellules.

Pour la lyse, la solution cellulaire est mélangée à un ratio 1:1 avec un tampon de lyse composé de TE 1X, SDS 2 %, protéinase k à 0,2 mg/ml, DTT 1%, PEO (poly-ethylene oxide) à 0,6% (m/v) et de RNase à 0,2 mg/ml. Ensuite les tubes ont été placés à 45°C pendant 60 minutes. L'ADN obtenu a été fragmenté par des sonication (Vibra cell, Bioblock scientific) à une amplitude 20 % pendant 1 seconde.