



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

M. ETIENNE LIAIS

le mardi 2 décembre 2014

Titre :

IDENTIFICATION ET CARACTERISATION DE VIRUS AVIAIRES PAR
DES APPROCHES DE SEQUENCAGE A HAUT DEBIT

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Ecole Nationale Vétérinaire de Toulouse (ENV TOULOUSE)

Directeur(s) de Thèse :

M. JEAN-LUC GUERIN

MME MARIETTE DUCATEZ

Rapporteurs :

M. ANTOINE TOUZE, UNIVERSITE DE TOURS

M. BERNARD DELMAS, INRA JOUY EN JOSAS

Membre(s) du jury :

M. ALAIN VIGNAL, INRA TOULOUSE, Président

M. JEAN-LUC GUERIN, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Membre

Mme GENEVIÈVE LIBEAU, CIRAD MONTPELLIER, Membre

Mme MARIETTE DUCATEZ, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Membre

Remerciements

Je voudrais remercier toutes les personnes ayant rendu ce projet personnel possible. Tout d'abord, les enseignants de l'université François Rabelais de Tours qui ont contribué (certains plus que d'autres il faut bien l'avouer) à ma volonté de prolonger mon parcours universitaire. Bien évidemment j'aimerais remercier également Jean-Luc (dit « Rouanne Loucos ») et Mariette de m'avoir choisi parmi les candidats (même si ils s'en sont mordus les doigts par la suite...). Non seulement ils m'ont permis de réaliser ce projet mais en plus ils ont toujours su être présents, et m'ont soutenu face à l'adversité (Dieu sait qu'il y en a eu !). Je suis devenu un grand... du moins un peu plus.

Je remercie les habitants et ex-habitants du poulailler (accessoirement mes co-bureaux) qui ont su me supporter, m'aider et qui ont rendu ce passage (éclair sur l'échelle d'une vie de chercheur) dans l'équipe IHAP si agréable au quotidien ! Leur soutien a été sans faille et partager un bout de chemin avec eux est tout simplement (pas toujours facile) un réel plaisir ! Je parle bien évidemment de Guillaume C, Guillaume LL (le marocain), Angélique, Charley, Charlotte et tous les « p'tits cocos » qui sont passés par là (étudiants M2 / stagiaires BTS / thésards véto notamment Florian, Quentin, Renaud et Jean-Benoît pour ne citer qu'eux). Un grand merci au FDJ, avec qui j'ai pu partager des marathons, bar à thon... et autres rassemblements scientifiques et festifs !

Que serait le labo sans le trio infernal Brigitte, Josy et Sylvie... Merci à vous d'avoir été toujours aussi joyeuses, amicales et pleines de vie ! Toujours un petit mot pour remonter le moral ou juste pour faire plaisir... c'était un vrai bonheur ! Bien sûr il n'y avait pas que l'ambiance, vous cherchez quelque chose ? Elles vous le trouveront ! Merci aussi à Martine qui m'a aidé lors de mes débuts dans le « labo des Martines ». Même si elles ne liront probablement pas cette thèse, je remercie aussi nos chères secrétaires qui ne sont jamais loin et qui ont toujours le sourire et les réponses aux questions administratives.

Je tiens aussi à remercier Stéphane et Christelle qui étaient toujours là pour répondre à mes questions (que ce soit scientifique ou non...), pour leur bonne humeur, leur gentillesse et leur savoir qu'ils ont partagé avec moi. Car oui, il y avait du boulot...

Comment ne pas remercier aussi le randonneur de l'extrême, le coureur unijambiste, l'expert kayakiste, le possédé du démon de la danse, le vampire des vacances, la terreur des PAINS AU CHOCOLAT (chocolatine pour les sudistes), le très jeune mais néanmoins redouté maître Aïkidoka... je veux bien sur parler de Vincent ! Je vais passer l'aspect scientifique (on

s'en fout un peu hein..) et aller droit au but, merci pour ton soutien morale, ton aide majeure dans l'extermination des viennoiseries, et ta bonne humeur qui nous a tous conquis !

Je remercie aussi la blonde d'immuno, qui nous a fait vivre Dallas mais avec qui chaque moment a été très agréable (sans mauvais jeu de mot...) et l'auteur à succès Benoît qui est aussi professeur d'anglais durant son temps libre. Même sans être de l'équipe, leur bonne humeur et leur gaîté ont rendu joviaux les jours les plus maussades.

Je voudrais aussi remercier Emily qui m'a supporté moi et mon caractère plutôt aléatoire...Elle ne m'a pas seulement accompagné durant ces 3 années de thèse mais elle m'a aussi épaulé pour les bons et les mauvais moments. Je voudrais aussi remercier ceux qui n'ont pas forcément été là physiquement mais qui m'ont soutenu durant cette belle expérience qu'est la thèse... mes camarades et amis de promo... mes amis tout court d'ailleurs, Alexis (le dompteur de pangolin) et Gwen (l'artiste photographe) qui sont tous deux proches de la rédaction, Adi, Mac, Tony, Moud, JB... enfin tout ce petit monde.

Je me répète mais encore un grand merci à Jean-Luc, Mariette et Christelle qui « ce çond arachez les chevoeux sure se manuscript »... ainsi que tous ceux que j'ai oublié volontairement ou non d'ailleurs !

INTRODUCTION.....	1
I. PRÉSENTATION DU CONTEXTE GÉNÉRAL DE LA THÈSE.....	2
II. MÉTHODES DE DIAGNOSTIC AVEC A PRIORI.....	4
A. Réactions de polymérase en chaîne.....	5
1. Principe de l'hybridation.....	5
2. Réaction de polymérase en chaîne (PCR).....	5
B. PCR quantitative ou en temps réel.....	9
1. Agent intercalant.....	11
2. Sonde.....	11
3. PCR compétitive.....	16
C. Techniques Sérologiques.....	17
1. ELISA.....	17
2. Autres tests sérologiques.....	21
D. Puces à ADN (Microarrays).....	21
III. MÉTHODES DE DIAGNOSTIC SANS A PRIORI.....	24
A. Isolement.....	24
1. Isolement sur culture cellulaire.....	24
2. Isolement sur œuf (ovoculture).....	25
B. Microscopie électronique.....	27
C. Séquençage.....	30
1. Historique du séquençage Classique.....	30
(1) Méthode Gilbert-Maxam.....	32
(2) Séquençage Sanger.....	32
(3) Le pyroséquençage.....	34
2. Haut débit.....	37
(1) 454 Roche.....	37
a. Historique et principe.....	37
(2) Illumina.....	41
a. Historique et principe.....	42

(3) Life technologie	48
a. Principe et méthode	49
3. Outils d'analyses	55
4. Applications du séquençage à haut débit.....	57
(1) Séquençage de génomes complets.....	60
(2) Métagénomique.....	61
(3) Identification et découverte de pathogènes.....	64
IV. MALADIES VIRALES AVIAIRES	66
A. Principales Maladies respiratoires aviaires	66
B. Principaux virus entériques aviaires	70
1. Les entérites à virus ADN	70
2. Les entérites à virus ARN	71
RÉSULTATS.....	73
I. L'AGENT DE LA MALADIE FOUROYANTE DE LA PINTADE : UN NOUVEAU CORONAVIRUS.....	74
A. Introduction	74
B. Articles.....	81
II. APPLICATION DES APPROCHES DE SÉQUENÇAGE À HAUT DÉBIT SANS A PRIORI À LA DÉTECTION ET	
CARACTÉRISATION DE VIRUS RESPIRATOIRES	101
A. Introduction	101
B. Article	108
DISCUSSION GÉNÉRALE.....	130
I. PERSPECTIVES DU DIAGNOSTIC PAR SÉQUENÇAGE À HAUT DÉBIT	131
II. PERSPECTIVES DE L'IDENTIFICATION DU CORONAVIRUS DE LA PINTADE COMME AGENT ÉTIOLOGIQUE DE	
LA MALADIE FOUROYANTE.....	133
III. LE DIAGNOSTIC À HAUT DÉBIT: QUEL AVENIR?	136
RÉFÉRENCES.....	139

Introduction

I. Présentation du contexte général de la thèse

Les maladies infectieuses ont toujours été une cause majeure de mortalité chez les populations humaine et animale. Tuant plus que les guerres, les maladies infectieuses sont même de plus en plus présentes (Figure 1) (1).

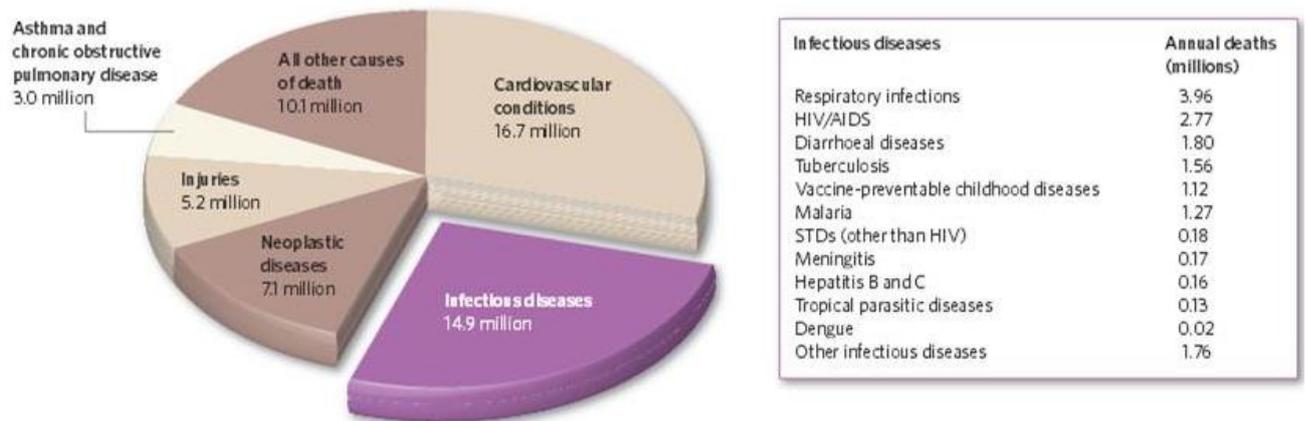


Figure 1 : Principales causes de mortalité dans le monde

The challenge of emerging and re-emerging infectious diseases
David M. Morens, Gregory K. Folkers & Anthony S. Fauci
Nature **430**, 242-249(8 July 2004)
doi:10.1038/nature02759

Les mécanismes d'adaptation de certains pathogènes comme les virus de la grippe ou du SIDA avec leurs taux de mutation élevés, les échanges de matériel génétique chez les bactéries ou encore le manque de connaissances et de moyens de traitement efficaces comme en ce moment pour la crise sanitaire autour de la ré-émergence du virus Ebola en Afrique, leur permettent de faire des carnages sur les populations (1). Les maladies émergentes ou ré-émergentes sont nombreuses et n'épargnent aucun continent (Figure 2). Pour pallier à cela, les scientifiques cherchent toujours à aller plus loin dans le diagnostic, la découverte de nouveaux pathogènes et de nouveaux systèmes de lutte (vaccins, nouvelles thérapies, etc) afin de mieux prémunir les populations contre ces fléaux.

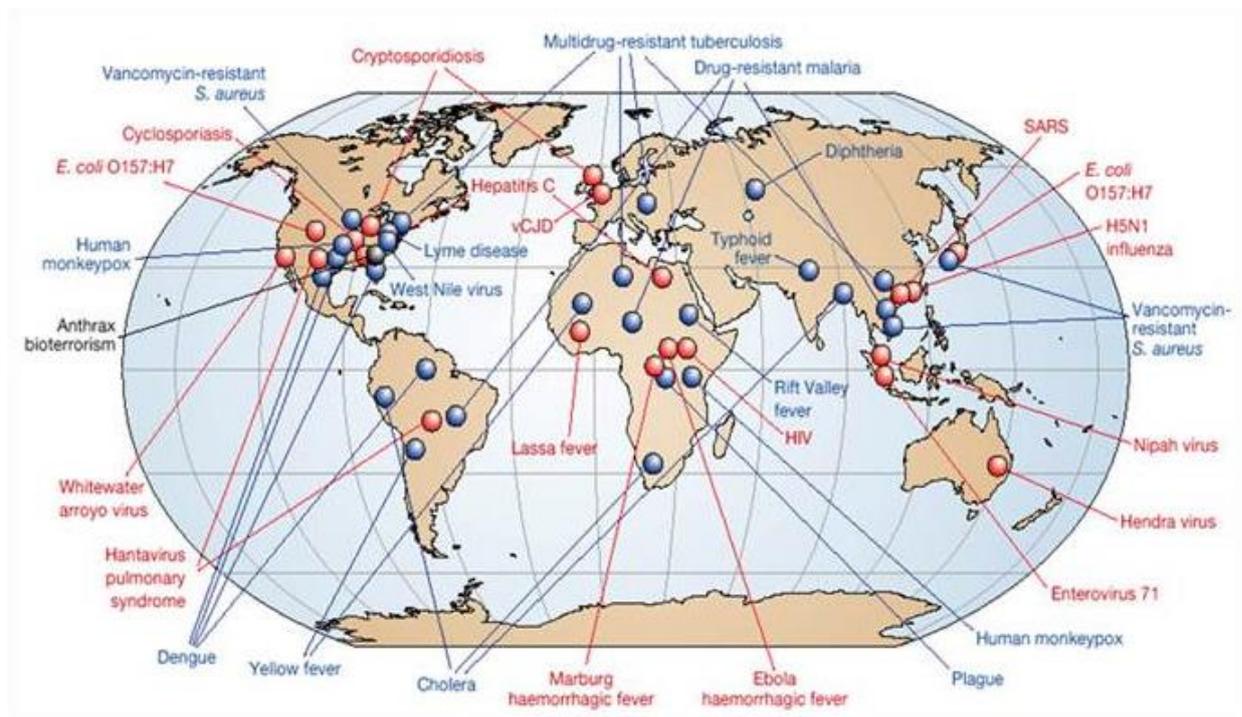


Figure 2: Représentation des émergences et ré-émergences de maladies infectieuses dans le monde.

En bleu les ré-émergences, en rouge les émergences des maladies infectieuses

The challenge of emerging and re-emerging infectious diseases
 David M. Morens, Gregory K. Folkers & Anthony S. Fauci
Nature **430**, 242-249 (8 July 2004)
[doi:10.1038/nature02759](https://doi.org/10.1038/nature02759)

En médecine humaine comme en médecine vétérinaire, les maladies infectieuses tiennent donc un rôle de premier plan et à ce titre, il est primordial de pouvoir identifier les agents pathogènes responsables de ces maladies. Plusieurs enjeux apparaissent rapidement lorsque l'on parle de recherche d'agents pathogènes : la rapidité, robustesse et répétabilité de la technique employée. De plus l'identification de pathogènes émergents, qu'ils soient connus ou non, est très important. En effet l'identification des agents permet d'améliorer nos connaissances et ainsi nous permet de mieux nous prémunir contre de futures ré-émergences : traitements anti-viral/fongique/bactérien, vaccination ou encore sérothérapie. Dans ce but, de nombreuses techniques ont été développées au fil des ans afin de rendre le diagnostic des pathogènes toujours plus précis et plus rapide. .

Les méthodes de diagnostic peuvent être classées en deux grandes catégories :

- Les méthodes avec *a priori*
- Les méthodes sans *a priori*

On parle de technique de diagnostic (de détection) avec *a priori* lorsque l'on cible un/des agent(s) pathogène(s) déjà connu(s) et caractérisé(s). On connaît donc la cible et on cherche à déterminer si l'agent est présent ou non dans l'échantillon biologique.

Les techniques dites sans *a priori* n'ont quant à elles aucune cible prédéterminée. Ainsi lors d'un diagnostic, ces méthodes permettent une recherche exhaustive des pathogènes s'affranchissant d'une mauvaise hypothèse clinique ou bien d'un agent inconnu. En effet un symptôme peut avoir plusieurs origines comme par exemple pour les entérites qui peuvent être aussi bien virales (rotavirus, astrovirus, reovirus, adenovirus de type 1 ou coronavirus (2–4)) que bactériennes (5,6).

Dans le cadre de cette thèse, nous avons eu pour but l'identification et la caractérisation de virus. Même si les deux notions sont très liées et peuvent se comprendre de différentes manières, nous allons considérer ici que l'identification précède la caractérisation d'un agent étiologique. L'objectif étant de déterminer l'intérêt et l'efficacité du séquençage à très haut débit pour la détection d'agents pathogènes chez les espèces aviaires. L'identification des cibles est une étape préliminaire pour la mise au point d'un test de diagnostic sensible et pertinent, tel que la réaction de polymérase en chaîne quantitative à haut débit (PCRq haut débit), qui sera présentée dans l'introduction suivante.

II. Méthodes de diagnostic avec *a priori*

Depuis de nombreuses années, les méthodes de diagnostic se sont terriblement améliorées notamment grâce aux avancées techniques les rendant plus rapides et efficaces. Il existe plusieurs méthodes de diagnostic différentes, regroupées dans deux groupes : les méthodes ciblées et les méthodes non ciblées. Une recherche dite avec *a priori*, ciblera un ou plusieurs pathogènes qui répondra à l'hypothèse posée lors de l'observation des signes cliniques. Le résultat obtenu sera alors positif ou négatif.

Il existe de nombreuses méthodes avec *a priori* de détection d'antigène disponibles en virologie clinique : les tests rapides, l'immunofluorescence, les réactions de polymérase en chaîne, etc, mais nous avons choisi de présenter spécifiquement les réactions de polymérase en chaîne classique et quantitative.

A. Réactions de polymérase en chaîne

1. Principe de l'hybridation

L'hybridation moléculaire désigne les associations qui peuvent survenir entre des acides nucléiques simples brins complémentaires conduisant à la formation de complexes comme par exemple d'acides nucléiques double brin. Ce complexe est assuré par des liaisons hydrogènes spécifiques ; dans le cas d'acide nucléique, on parlera de complémentarité de bases : Adénine (A) – Thymine (T) (ou Uracile (U)) ou Cytosine (C) – Guanine (G). Les liaisons entre la cytosine et la guanine sont plus fortes puisqu'elles comportent trois liaisons hydrogènes alors que les liaisons A – T en ont deux.

Le principe d'hybridation est la source de nombreuses techniques de biologie moléculaire faisant intervenir au moins un brin cible et un brin connu (sonde, amorce).

Dans le contexte d'une recherche de pathogène, il est important d'avoir des notions sur la physiopathologie de celui-ci. Tous les virus/bactéries ne possèdent pas le même tropisme. On retrouve donc un pathogène entérique dans l'appareil digestif et un pathogène respiratoire dans l'appareil respiratoire. De plus, certains pathogènes ne sont pas excrétés de façon continue dans l'hôte, il est donc nécessaire d'effectuer les prélèvements dans la bonne « fenêtre » temporelle. Ainsi les sites et timing de prélèvement sont des points cruciaux lors d'une recherche de pathogène.

2. Réaction de polymérase en chaîne (PCR)

La réaction de polymérase en chaîne (PCR) a été inventée en 1983 par Kary Mullis, invention récompensée par un prix Nobel en 1993. Cette technique permet l'amplification d'un fragment d'ADN connu. Le principe de la PCR est basé sur les propriétés d'une enzyme, l'ADN polymérase, qui a le pouvoir de synthétiser à partir d'une amorce un brin complémentaire. L'ADN matriciel sera répliqué en un très grand nombre de copies. On peut donc parler de clonage moléculaire, puisque seule la cible des amorces sera amplifiée. Désormais incontournable en biologie moléculaire, la PCR possède un nombre d'applications considérable.

Une PCR va mettre en jeu plusieurs acteurs, tout d'abord la matrice à amplifier, habituellement sous forme d'ADN double-brin. Ensuite les amorces qui sont de petits brins d'ADN ayant la capacité de reconnaître et de se fixer par complémentarité à un brin dit

complémentaire. Elles jouent un rôle crucial, elles sont connues et choisies (« dessinées ») de façon à encadrer la séquence à amplifier. L'enzyme sur laquelle repose cette technique d'amplification est l'ADN polymérase (Taq polymérase) qui utilise les quatre nucléotides (dGTP, dCTP, dATP, dTTP), appelés désoxyribonucléotides triphosphates (dNTPs) lorsqu'ils sont regroupés.

Les étapes d'un cycle de PCR sont schématisées dans la figure suivante (Figure 3):

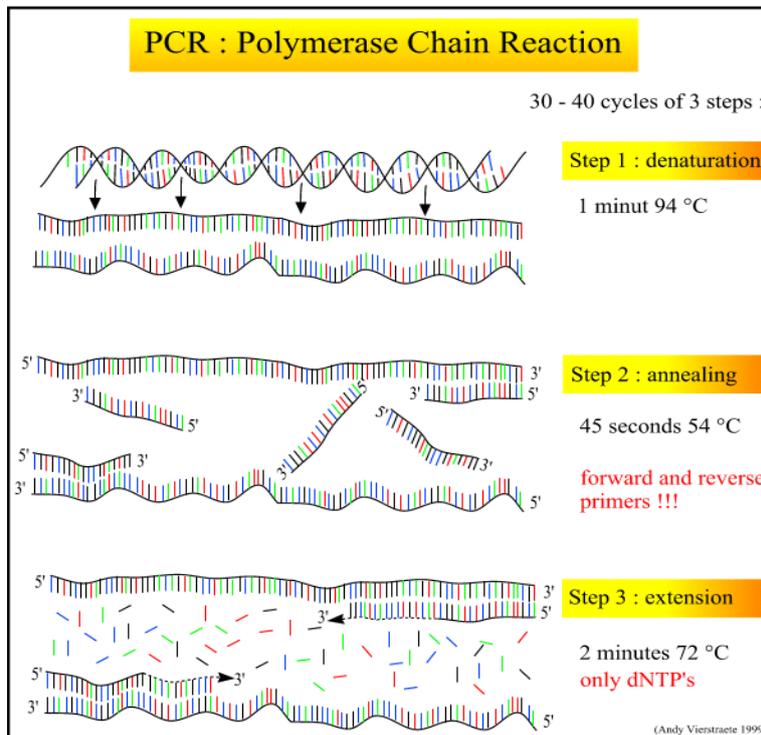


Figure 3 : Schématisation du principe de la PCR
<http://users.ugent.be/~avierst r/principles/pcr.html>

- La dénaturation

Elle consiste à chauffer, généralement à 94°C, les doubles brins d'ADN entraînant leur ouverture. Les liaisons hydrogènes reliant les deux brins d'ADN sont rompues par la température, les ADN double brin donnent donc des ADN simple brin. Cette étape est aussi utile pour dénaturer les structures secondaires (exemple : boucle). On parle alors d'ADN dénaturé.

- L'hybridation

Il s'agit de l'étape durant laquelle les amorces vont reconnaître et se fixer à leur brin complémentaire. La température d'hybridation varie en fonction de la composition nucléique des amorces. En effet, plus les amorces sont riches en guanines et cytosines, plus leur température d'hybridation (T_m) va être élevée. Cette température est généralement

comprise entre 50°C et 60°C, et plus la température est élevée, plus l'hybridation des amorces sur leur brin complémentaire est sélective et donc spécifique.

- L'élongation

Cette troisième étape correspond à la synthèse des brins d'ADN complémentaire par l'ADN polymérase. A cette température la Taq polymérase (thermorésistante) s'attache aux ADN simple brin amorcés et active la réplication. Cette enzyme va utiliser les dNTPs présents dans le milieu réactionnel pour synthétiser le brin complémentaire. L'élongation se fait généralement à 72°C correspondant à la température optimum d'efficacité de l'enzyme. Cette température peut varier en fonction de l'enzyme, en effet la température optimum d'efficacité est de 68°C pour certaines enzymes et peut descendre jusqu'à 60°C pour d'autres. En général, les protocoles recommandent l'ajout d'une étape d'élongation final à 72°C après les cycles d'amplifications surtout avec des séquences d'intérêt de grande taille.

Ces trois étapes sont répétées entre 30 à 45 fois ce qui permet une amplification exponentielle de la matrice de départ, puisqu'une fois l'élongation finie les brins générés deviennent à leur tour une cible pour le cycle suivant. A la fin des cycles d'amplification, une visualisation est possible grâce à des agents intercalants tel que le Bromure d'ethidium (BET) ou le « Sybr® safe ». Ces agents intercalants se glissent entre les doubles brins d'ADN et sont fluorescents lors de l'exposition aux Ultra-violets ou à la lumière bleue. Le produit de PCR est soumis à une migration sur gel d'agarose par électrophorèse. Cette électrophorèse va permettre de séparer les acides nucléiques en fonction de leur poids moléculaire.

La PCR est un outil de diagnostic de routine très utilisé, seul (avec visualisation après migration), ou suivi d'une technique complémentaire de caractérisation de produits amplifiés, tel que le séquençage. Malgré la standardisation de cette méthode, il existe toujours des limites. Le risque de contamination, la présence d'inhibiteurs de PCR, la dégradation des acides nucléiques en cas de mauvais conditionnement, et surtout le risque de mutations au niveau des sites de fixation des amorces sont autant d'inconvénients auxquels il faut faire attention.

La PCR est un outil très puissant pour reconnaître des cibles dont on connaît la séquence. En effet, les amorces utilisées doivent s'hybrider sur la matrice. Cependant, il est possible d'utiliser des amorces « dégénérées » afin de leur permettre une certaine flexibilité. Une base est dite dégénérée lorsqu'elle possède la faculté d'avoir plusieurs correspondances. En

effet, une base dégénérée à la capacité de s’hybrider par complémentarité sur plusieurs bases. Ainsi l’association de plusieurs bases dégénérées dans une amorce permet une hybridation à plus large spectre. Il existe plusieurs possibilités de codage de base dégénérée (Tableau 1).

Tableau 1 : représentation des possibilités de correspondance des bases dégénérées.

Base dégénérée	Correspondance			
R	G	A		
K	G	T		
S	G	C		
W	A	T		
M	A	C		
Y	T	C		
D	G	A	T	
V	G	A	C	
B	G	T	C	
H	A	T	C	
N	G	T	C	A

Cette particularité permet aux amorces de s’hybrider sur des fragments d’ADN ayant une séquence relativement proche comme deux virus d’un même genre ou porteurs de quelques mutations. Ainsi en ciblant des séquences virales conservées il est possible de détecter des variants génétiquement proches (7). Grâce à l’utilisation d’amorces dégénérées, la PCR a permis l’identification de nombreux virus (8,9). Un outil facilitant le design d’amorces dégénérées dans un but d’identification de nouveaux virus a été créé (CODEHOP PCR (7)). Cet outil peut être utilisé pour la découverte de nouveaux virus mais également pour la détection d’un grand nombre de souches différentes d’un même virus (ou variant), que des amorces « classiques » (plus spécifiques) n’auraient pas reconnues (10).

Après la détection d’un agent pathogène par PCR, il est possible de le sous-typer et même de le caractériser au niveau moléculaire. Pour certains pathogènes on peut utiliser des amorces spécifiques d’un Genre, d’un sous-type, d’un variant et donc mieux caractériser la souche avec une ou plusieurs PCRs supplémentaires. Des profils de restriction étaient également très utilisés par le passé pour répondre aux mêmes questions. Aujourd’hui il est fréquent de séquencer un produit de PCR pour le comparer aux bases de données et faire des analyses phylogénétiques permettant de mieux caractériser une souche pathogène. Ces méthodes de

caractérisation post-identification sont également avec *a priori* même si le séquençage ne l'est pas toujours, raison pour laquelle la technique est détaillée plus tard dans cette introduction.

B. PCR quantitative ou en temps réel

La PCR en temps réel est une méthode qui utilise le même principe que la PCR classique, à la différence que l'amplification est quantifiée tout au long de la réaction et non pas uniquement à la fin, donc en temps réel. La quantité d'ADN est quantifiée à chaque cycle d'amplification grâce à un marqueur fluorescent dont l'intensité du signal est proportionnelle à la quantité d'amplicons produits. Ce système s'appuie sur l'excitation par un laser d'un fluorochrome, présent dans le mix de PCR. L'intensité de fluorescence est lue par une caméra numérique, elle-même liée à un logiciel d'analyse. Cela permet de réaliser des dosages quantitatifs de l'ADN d'intérêt, par exemple pour déterminer une charge virale. Il existe deux types de quantifications : la quantification absolue faisant intervenir une gamme étalon et la quantification relative demandant la comparaison entre un gène de ménage et l'ADN à doser. Il est important de définir la limite de sensibilité inférieure de la technique. Ce point correspond à la quantité d'ADN minimale détectable par l'utilisation de la méthode. Plus l'échantillon de départ est concentré en ADN cible et plus le nombre de cycles nécessaires pour que le signal fluorescent dépasse significativement le bruit de fond sera faible. Ce point est communément appelé Ct (« cycle threshold ») et se situe en début de phase exponentielle (Figure 4). Le suivi de la fluorescence au cours des cycles d'amplification fera apparaître 3 phases :

- Bruit de fond : la quantité d'ADN amplifié est inférieure au seuil de détection et n'émet pas une fluorescence suffisante pour être visualisée.
- Phase exponentielle : la quantité d'ADN amplifié permet une fluorescence suffisante pour être enregistrée. La fluorescence sera proportionnelle à la quantité d'ADN généré.
- Phase de plateau : la réaction n'est plus exponentielle due à un élément limitant (souvent l'ADN polymérase).

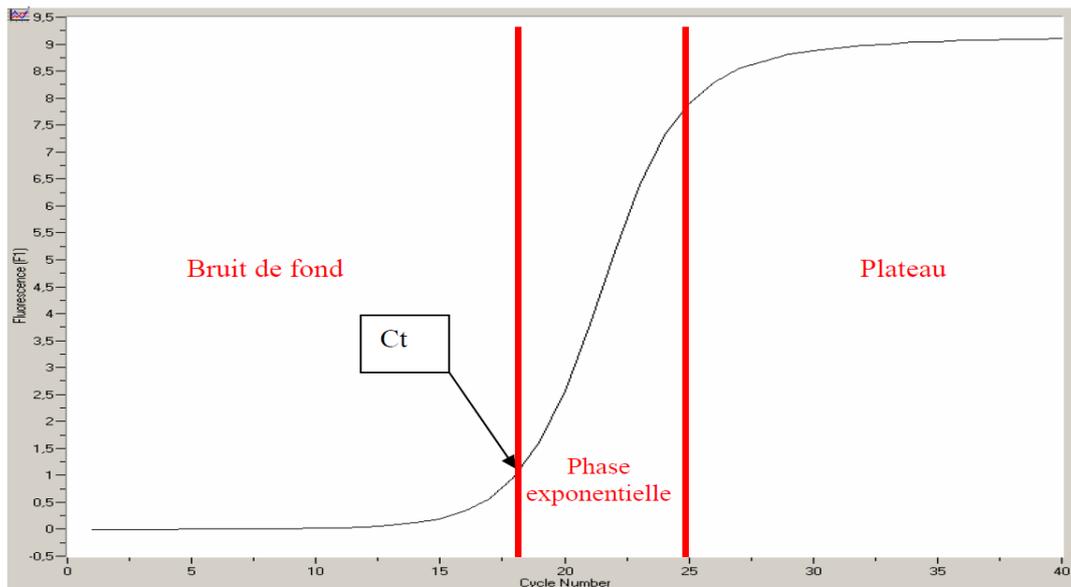


Figure 4 : Représentation des différentes phases lors d'une PCRq. Nombre de cycles sur l'axe des abscisses et quantité de fluorescence sur l'axe des ordonnées.

Pour quantifier un échantillon, il est souvent nécessaire d'établir une courbe standard. Cette courbe correspond à la valeur de Ct en fonction de la concentration en ADN d'une gamme étalon. Le Ct d'un échantillon inconnu sera reporté sur la courbe standard et traduit en concentration d'ADN (Figure5).

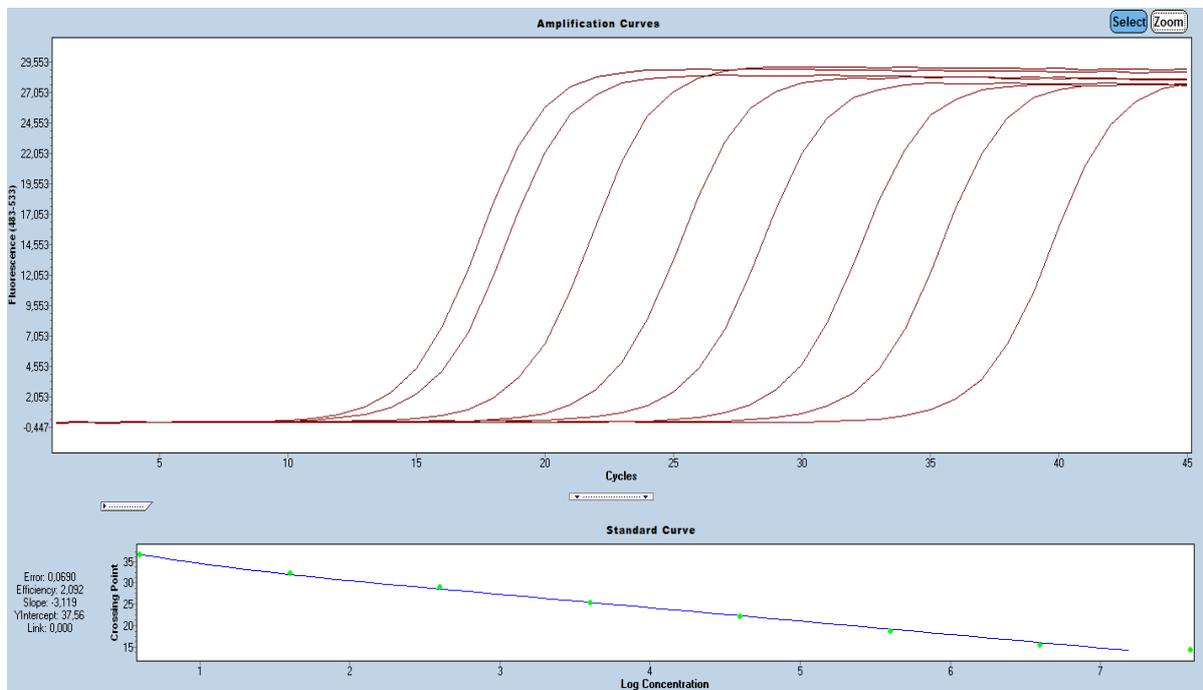


Figure 5 : Représentation d'une gamme étalon d'une PCR quantitative et de sa courbe standard.

Il existe plusieurs méthodes s'appuyant sur deux types de marqueurs de fluorescence : les agents intercalants et les sondes. Il existe aujourd'hui des technologies permettant de réaliser ce type de PCRq en très grand nombre simultanément : elle est dite « à haut débit », comme par exemple le système BioMark™ (<http://www.integragen.com/165-biomark-fluidigm.htm>).

1. Agent intercalant

Il existe plusieurs marqueurs pour mesurer la fluorescence, le plus commun est le Sybr® green. C'est un agent intercalant, il n'intervient donc pas dans la spécificité de la réaction. Par conséquent la spécificité d'une réaction repose sur celle de ses amorces. La pureté des produits amplifiés est vérifiée en réalisant une courbe de fusion. Cette courbe correspond au relevé de la fluorescence lors d'une montée en température progressive. On obtiendra alors un seul pic sur la dérivée primaire de cette courbe correspondant à la température de fusion (aussi appelé Tm) si tous les amplicons ont la même séquence. Si les pics sont décalés par rapport au Tm attendu, ou qu'il y a plusieurs pics sur la courbe de fusion alors cela signifie le produit de PCR n'est pas spécifique. Malgré un coût d'utilisation faible, une facilité d'emploi, cette technique comporte tout de même quelques limites. Par exemple les agents intercalants détectent tous les ADNs double brins (dimères d'amorces, amplicons non spécifiques...) et pas seulement l'amplicon cible. Ceci peut poser un problème de spécificité, généralement corrigé par l'analyse de la courbe de dissociation.

2. Sonde

Il existe plusieurs systèmes utilisant des sondes : *Taqman*®, *Fluorescent Resonance Energy Transfert* (FRET), *Molecular Beacons* et *Scorpion*.

La technologie utilisant la sonde Taqman® est basée sur l'hybridation d'une sonde d'acides nucléiques agrémentée d'un fluorochrome émetteur (reporter) d'une part et d'un fluorochrome suppresseur (quencher) d'autre part. L'hydrolyse de la sonde lors de l'élongation sépare le reporter du quencher, permettant ainsi la mesure de la fluorescence, qui est proportionnelle à la quantité de sonde dégradée (Figure 6).

L'utilisation de sondes Taqman® permet une augmentation de la spécificité par rapport au Sybr® green et rend possible un multiplexage de sondes portant des fluorochromes différents. Néanmoins cette méthode reste très coûteuse et manque de souplesse lorsque les souches sont variables : les amorces et la sonde risquent de ne pas s'hybrider au bon endroit en cas de mutation ou si elles sont situées dans une région de séquences répétées.

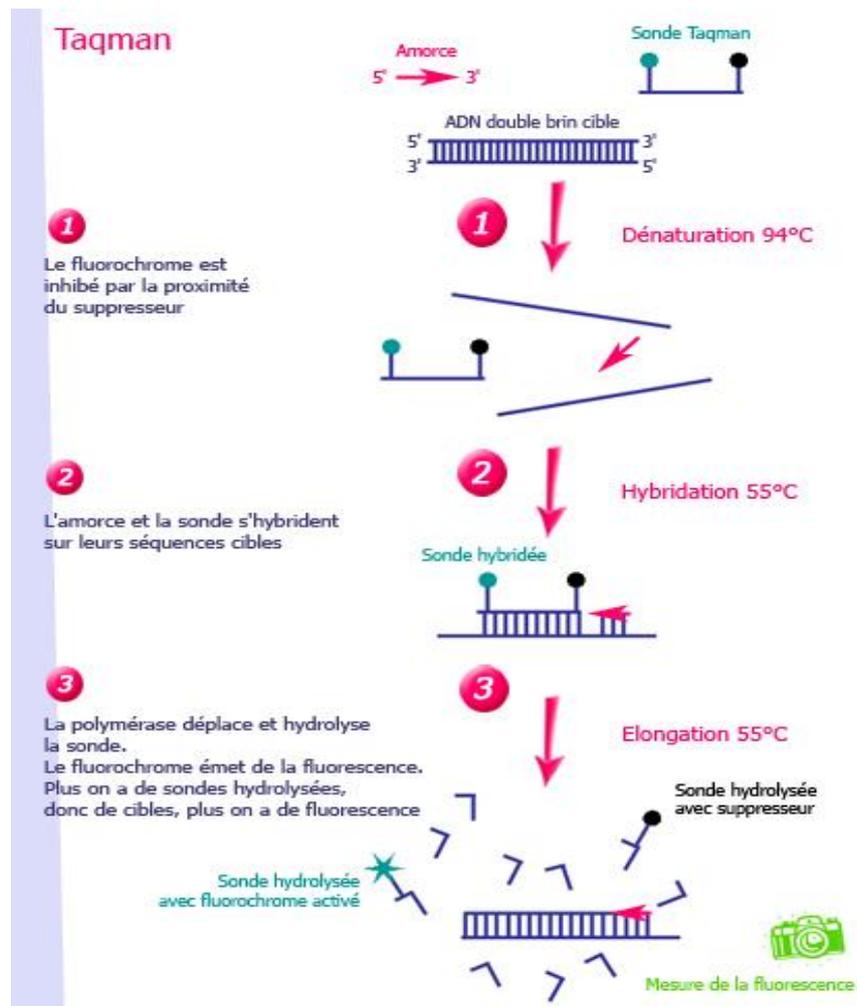


Figure 6 : principe de l'hybridation d'une sonde Taqman.

<http://www.ilm.pf/PCRtempsreel>

Le système FRET fonctionne quant à lui avec deux sondes, une dite porteuse de fluorescence et l'autre dite émettrice. La sonde porteuse de fluorescence n'émet qu'un faible bruit de fond lorsqu'elle n'est pas hybridée et proche de la sonde émettrice. Les deux sondes s'hybrident sur une cible distante que de quelques nucléotides (pas plus de 10 pour que la sonde émettrice joue son rôle) ce qui entraîne un transfert d'énergie de la sonde porteuse vers la sonde émettrice. Ce transfert permet l'émission de fluorescence proportionnelle à la quantité d'ADN synthétisée (Figure 7).

Comme les sondes Taqman®, le système FRET comporte les avantages de l'utilisation des sondes, une très bonne sensibilité et une capacité de multiplexage plus importante encore. Cependant son prix est plus élevé encore que le système taqman® et certaines mutations peuvent déstabiliser le système.

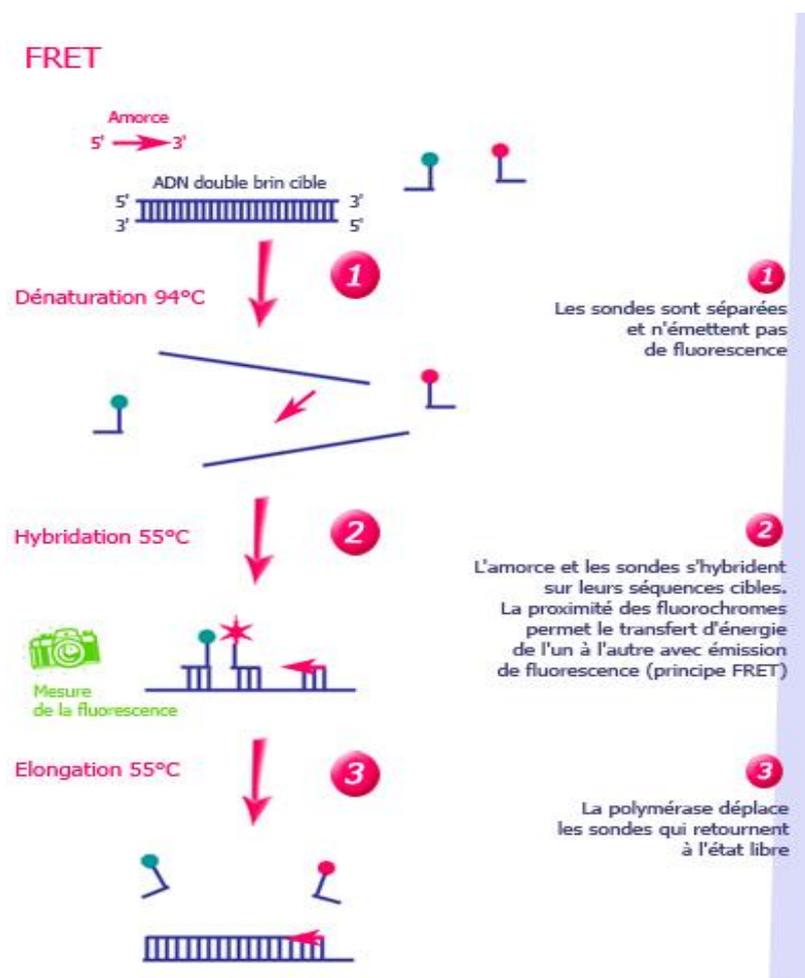


Figure 7 : principe d'hybridation d'une sonde FRET
<http://www.ilm.pf/PCRtempsreel>

Le *Molecular Beacons* utilise le même principe que la sonde Taqman®, à la différence que les deux fluorochromes sont portés à chaque extrémité d'une sonde en épingle à cheveux. Lors de l'étape de dénaturation la sonde se linéarise permettant sa fixation sur l'ADN cible lors de l'étape d'hybridation. Le quencher (suppresseur) n'étant plus à distance suffisante pour jouer son rôle, le reporter pourra émettre de la fluorescence. Durant l'étape d'élongation, la polymérase libère la sonde qui retrouve son état initial et éteint le signal lumineux (Figure 8).

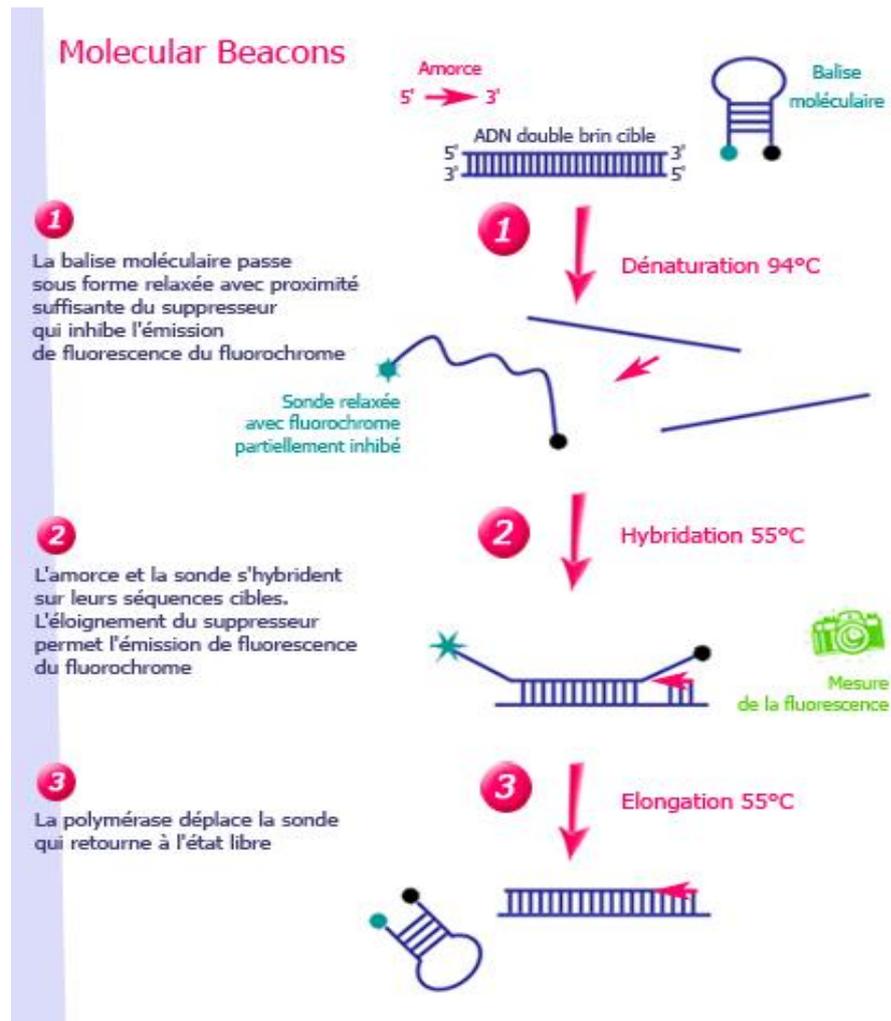


Figure 8 : principe de l'hybridation d'une sonde *Molecular Beacons*
<http://www.ilm.pf/PCRtempsreel>

Le *Scorpion* est quant à lui une variante des *molecular beacons* ayant une molécule d'hexéthylène glycol (HEG) à la suite du quencher. L'HEG porte une amorce qui va se fixer sur la séquence cible afin de permettre la polymérisation du brin complémentaire. Après dénaturation de ce brin néoformé, l'amorce en épingle linéarisée comporte une molécule d'HEG et un brin complémentaire néoformé d'un côté et une partie sonde avec à l'autre extrémité le fluorochrome. Le brin néoformé s'hybride sur la sonde linéarisée lors de ce deuxième cycle éloignant ainsi le quencher et le reporter permettant la lecture de la fluorescence générée (Figure 9).

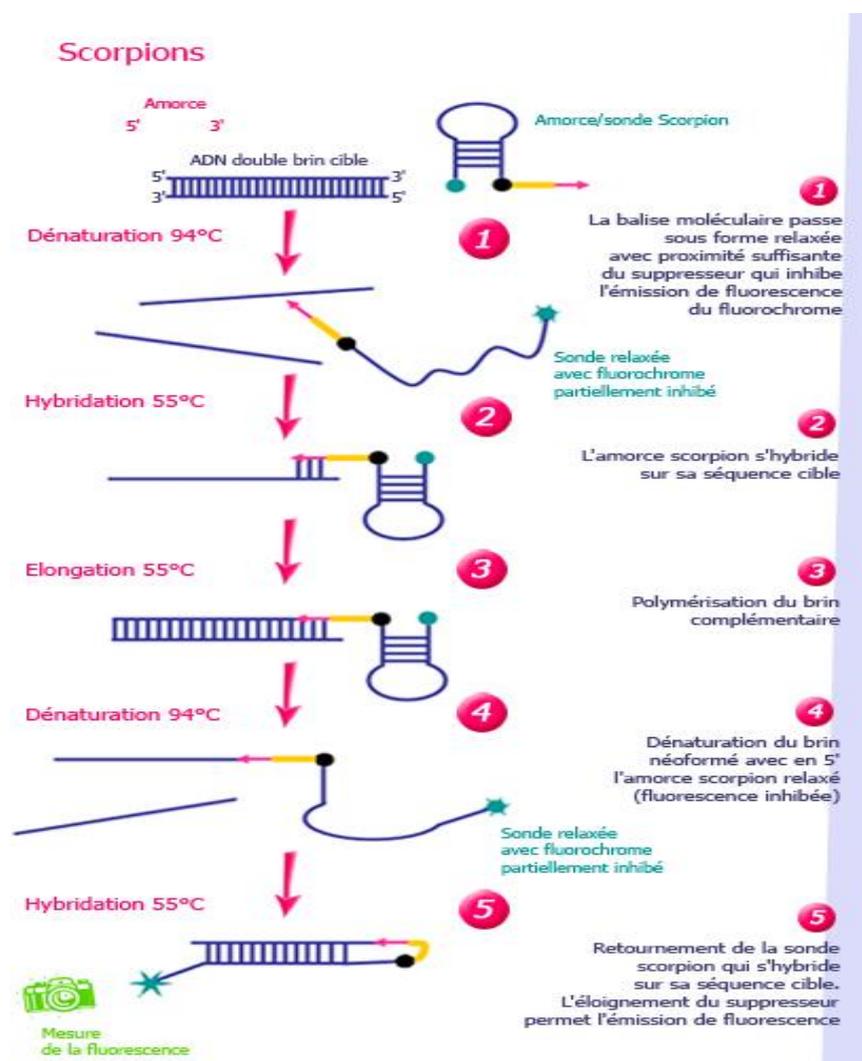


Figure 9 : principe de l'hybridation d'une sonde *Scorpions*
<http://www.ilm.pf/PCRtempsreel>

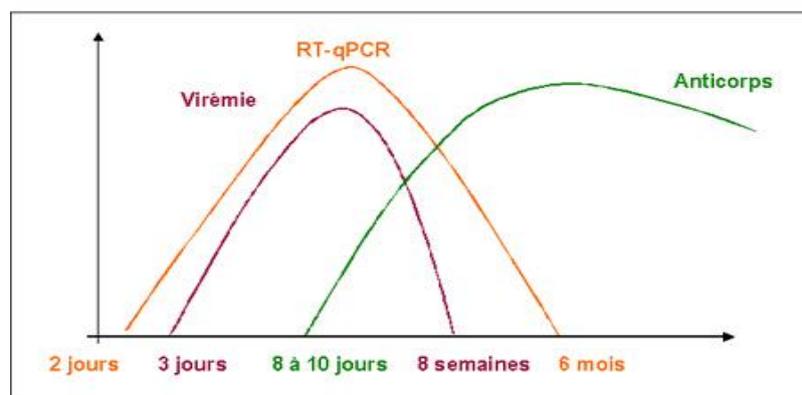
Ces deux dernières sondes possèdent des avantages et des inconvénients communs. En effet elles possèdent toutes deux une grande spécificité qui leur permet d'être utilisées pour la détection de SNP (single nucléotide polymorphism), mais la sonde *Scorpion* est plus rapide, plus efficace et elle permet également une augmentation de l'intensité de la fluorescence. Ces deux sondes sont néanmoins beaucoup plus chères et leur conception est très délicate.

3. PCR compétitive

La PCR compétitive est une méthode où l'on mesure les quantités relatives d'ADN grâce à des standards. Pour cela on ajoute une quantité connue d'ADN portant la séquence reconnue par les amorces à l'échantillon. Les deux cibles sont donc amplifiées avec le même jeu d'amorces, de par cette compétition le ratio des deux produits d'amplifications est équivalent au ratio de la cible recherchée et de l'ADN compétiteur. Puisque la quantité d'ADN compétiteur est connue, il est possible d'établir la quantité d'ADN cible dans l'échantillon.

C. Techniques Sérologiques

Lors d'une recherche sérologique le timing des prélèvements est d'une importance capitale. En effet, lors d'une infection les antigènes (Ag) ne sont retrouvés qu'après répllication de l'agent sur le ou les sites d'infection (suivant le tropisme du pathogène). Les anticorps (Ac) sont quant à eux produits après une période d'incubation, laissant le temps à l'organisme de reconnaître l'Ag et de le combattre : le temps de séroconvertir. Suivant le type d'Ac (Immunoglobuline M (IgM) ou Immunoglobuline G (IgG)) le temps d'apparition et de maintien des Ac peut varier (Figure 10).



D'après S. Zientara et al., 2009

Figure 10 : Cinétique des paramètres biologiques suite à une infection par des virus de la fièvre catarrhale.

<http://www.fcoinfo.fr/spip.php?article373>

1. ELISA

Inventé en 1971 par Peter Perlmann et Eva Engvall de l'université de Stockholm, le test ELISA (enzyme-linked immunosorbent assay) est une technique enzymatique de recherche d'Ac ou d'Ag. Cette méthode est basée sur un système de clé-serrure et la capacité de reconnaissance spécifique des anticorps. Il existe plusieurs types d'ELISA : ELISA direct, indirect, sandwich et compétition.

Le test ELISA direct est le plus simple, il nécessite un Ac et un Ag. Le principe est de fixer (ou « coater ») à un support solide (plaque 96 puits) un Ac ou un Ag (suivant la recherche) et de mettre en contact un Ag ou un Ac marqué avec une enzyme (Ag si la plaque est « coté » avec des Ac et Ac si la plaque est coté avec des Ag). Après avoir éliminé les Ac/Ag non fixés, un substrat est ajouté dans chaque puits permettant la lecture de la réaction par fluorescence

ou colorimétrie (suivant la nature de l'enzyme et du substrat). Cette technique comporte l'avantage d'être rapide et d'éviter les réactions croisées entre plusieurs Ac. Mais elle implique le marquage individuel des Ac/Ag et le signal étant moins amplifié, la lecture est moins sensible.

L'ELISA indirect comporte une étape supplémentaire qui implique un Ac supplémentaire. Le premier Ac est appelé primaire et le second est nommé secondaire. Pour cette technique l'Ag « coaté » sur le support est incubé avec l'Ac recherché dit Ac primaire. Après élimination des éléments non fixés, un Ac secondaire marqué est ajouté dans les puits. L'étape de révélation avec le substrat et de lecture du résultat est similaire à l'ELISA direct. Ce type d'ELISA est plus sensible et moins cher que l'ELISA direct. De plus les Ac secondaires reconnaissent le fragment Fc des Ac grâce au déterminant isotypique, conservé entre différents Ac primaires. Un même Ac secondaire peut donc être utilisé dans différentes situations (avec différents Ac primaires) (Figure 11).

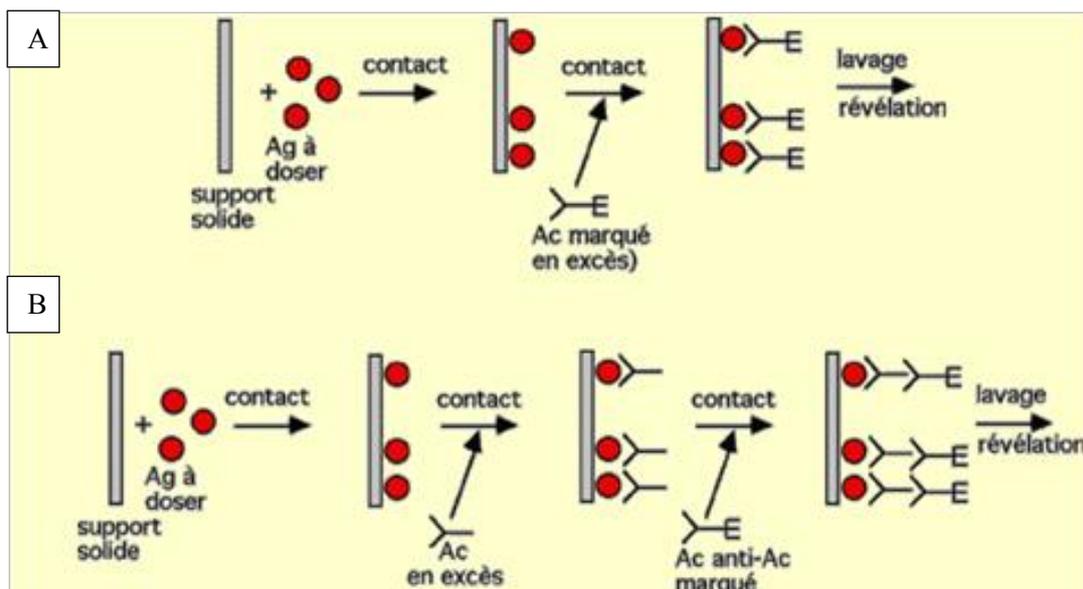


Figure 11 : Représentation du principe des tests ELISA direct et indirect. A/ ELISA direct. B/ ELISA indirect

<http://www.djamiatic.net/immuno/Chapitre%203.html>

Le type d'ELISA en compétition est défini par la mise en concurrence de deux Ac ou deux Ag, l'un étant marqué et l'autre correspondant à l'échantillon. Pour cela la plaque est coâtée avec des Ag ou Ac et les deux Ac ou Ag sont incubés ensemble dans les puits. Contrairement aux ELISA direct et indirect, l'intensité du signal est inversement proportionnelle à la quantité d'Ac ou Ag dosé. En effet, plus il y a d'Ac (ou d'Ag) non marqué (provenant de l'échantillon à doser) et moins l'Ac (ou l'Ag) marqué peut se fixer. Inversement, plus la quantité d'Ac (ou d'Ag) non marqué est faible et plus son concurrent (l'Ac ou l'Ag de marqué) se fixe, permettant une intensité de signal plus importante (Figure 12).

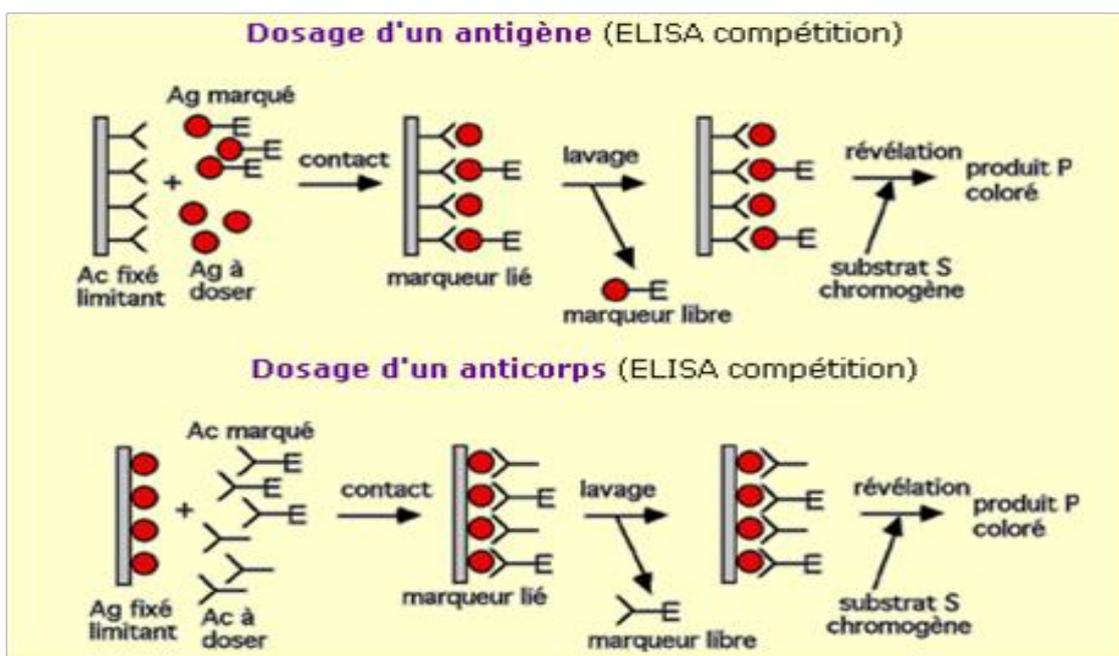


Figure 12 : Schématisation du principe d'un test ELISA en compétition.

<http://www.djamiatic.net/immuno/Chapitre%203.html>

L'ELISA sandwich est une variante moins courante des tests ELISA. Elle permet de quantifier les Ag entre deux couches d'Ac : les Ac de capture et les Ac de détection. Pour utiliser ce type d'ELISA, il faut que l'Ag à doser porte au moins deux épitopes antigéniques capable de fixer des Ac. Les Ac de capture et de détection peuvent être monoclonaux (ne reconnaissant qu'un seul épitope antigénique) ou polyclonaux (reconnaissant plusieurs épitopes antigéniques différents). Les Ac monoclonaux permettent une détection et une quantification plus spécifique que les Ac polyclonaux qui eux reconnaissent plusieurs épitopes différents. Le gros avantage de ce système est qu'il accepte des échantillons très complexes et ne demande pas de purification. L'Ac de capture sert d'étape de purification.

Cependant l'optimisation de ce type d'ELISA peut être difficile puisqu'il faut éviter toute réaction croisée entre les différents acteurs. L'ELISA sandwich peut être « direct » ou « indirect », les différentes étapes de cette méthode de sandwich sont les suivantes :

- une quantité connue d'Ac de capture est fixée à la plaque.
- Après avoir lavé et bloqué les éventuelles liaisons libres, l'Ag est ajouté.
- Après élimination des Ag non fixés, l'Ac primaire est ajouté (Ac primaire marqué si on souhaite réaliser un ELISA sandwich simple).
- Après élimination des Ac primaires non fixés, l'Ac secondaire (s'il s'agit d'un ELISA sandwich indirect) marqué est incubé.
- L'étape de lecture est semblable aux tests ELISA précédemment décrits (lavage, ajout du substrat et lecture après excitation du complexe enzyme-substrat.)

Le nombre d'Ac rend ce test ELISA très spécifique, et l'Ac de capture permet de s'affranchir d'une étape de purification. La possibilité d'employer une technique directe ou indirecte rend ce type d'ELISA très sensible mais également très flexible (Figure 13).

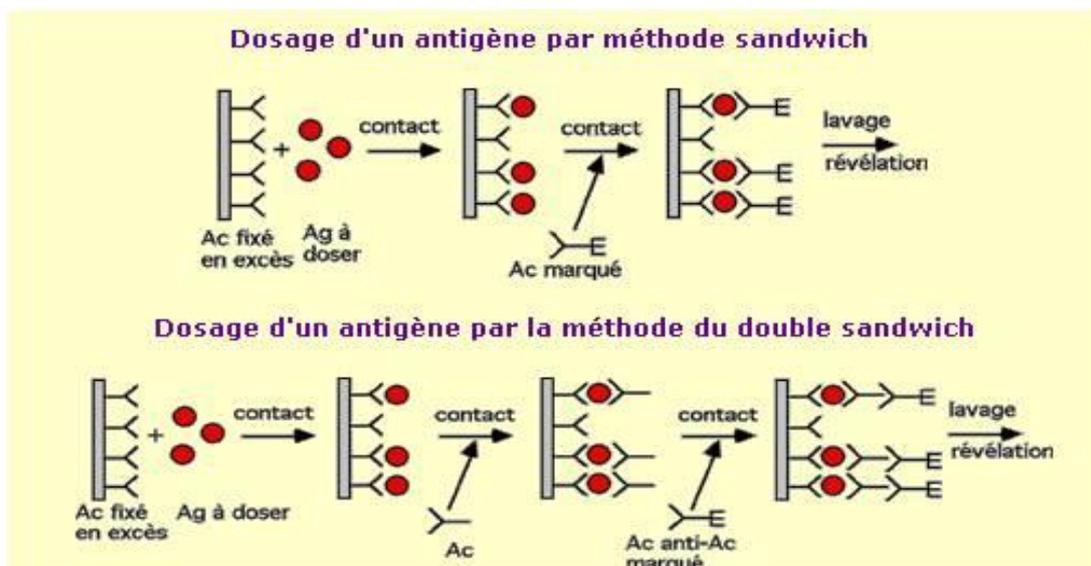


Figure 13 : Représentation des méthodes d'ELISA sandwich. La méthode sandwich correspond à une technique directe et la méthode du double sandwich à une technique indirecte.

<http://www.djamiatic.net/immuno/Chapitre%203.html>

2. Autres tests sérologiques

Il existe un grand nombre de tests basés sur la détection d'anticorps. Chaque test utilise les propriétés de ces anticorps recherchés. Comme le principe d'agglutination, qui permet aux anticorps de s'agglomérer autour des antigènes portés par un support comme des billes de latex (recherche du virus de la rubéole). Ou encore l'inhibition de l'hémagglutination, en effet certains virus comme le virus influenza sont capables d'hémagglutiner les hématies. Les virus une fois neutralisés par les anticorps, les hématies pourront sédimenter, ce qui permet de vérifier la présence ou non de ces anticorps dans un échantillon de sérum. La séroneutralisation est une autre technique très utilisée en diagnostic. Le principe de cette méthode repose sur l'inhibition de la réplication d'un virus. Pour cela, le virus est mis en présence d'un sérum (que l'on souhaite analyser). Si on observe une inhibition alors on peut conclure à la présence d'Ac neutralisants dans le sérum étudié. Pour cette technique l'identification du pathogène est nécessaire.

D. Puces à ADN (Microarrays)

Cette méthode de détection est basée sur les travaux d'Edwin Southern, Kalim Mir et Mikhail Schepinov (11). Ces chercheurs ont montré que deux acides nucléiques complémentaires peuvent se séparer et s'associer de façon réversible à condition d'être dans certaines conditions (Température et concentration saline du milieu optimales). La réaction est réalisée sur un support solide de petite taille tel qu'une lame en verre ou en nylon, ce qui différencie cette technique des dots blots. Sur ce support sont déposées des séquences d'ADN de petite taille ou des oligonucléotides synthétiques (entre 50 et 70-mer) aussi appelés « sondes ». Le choix des sondes est fait de sorte qu'il n'existe qu'une seule et unique cible pouvant s'y fixer. Le support porteur des sondes est ensuite hybridé avec les ADNs extraits des échantillons que l'on souhaite analyser. Ces ADNs aussi nommés « cibles » (éventuellement obtenus par transcription inverse d'ARNs cibles), sont marqués avec des fluorochromes, souvent de la famille des cyanines (Cyanine3 de couleur verte et Cyanine5 de couleur rouge). L'acquisition des signaux d'hybridation est réalisée à l'aide d'un scanner de fluorescence. L'intérêt d'utiliser deux fluorochromes de couleurs différentes est de permettre la distinction des signaux de deux souches distinctes lors d'une même hybridation. Après l'acquisition du signal, une analyse va permettre la quantification

correspondant au niveau d'expression de la cible dans l'échantillon déposé sur la puce (Figure 14).

Le principe de la puce est basé sur la détection de séquences connues, cette technique robuste permet la détection et la surveillance de virus de façon aisée. En 2002, Wang et ses collaborateurs ont montré qu'il était possible non seulement d'identifier et distinguer les serotypes viraux par les hybridations sur la puce de chacun de ces virus, mais également de détecter des virus qui n'étaient pas spécifiquement marqués sur la puce en sélectionnant des marqueurs de famille très conservés (12). Cette caractéristique place cette technique à la frontière des détections avec et sans *a priori*. Cette puce fait partie de la catégorie des puces nommées Panviral microarray assay et est appelée *Virochip*. Elle permettait lors de sa première publication le criblage d'une centaine de virus et en 2011 porte 36.000 sondes dérivées de plus de 1.500 séquences virales (13). La *Virochip* a été utilisée de nombreuses fois, tout d'abord pour valider ses performances dans la détection de virus présents chez des personnes présentant des signes cliniques de détresse respiratoire (14). Cette puce est capable de mettre en évidence simultanément de nombreux virus présents dans un seul échantillon. Ces résultats sont plus sensibles, plus précis que les RT-PCR spécifiques ou l'isolement sur culture cellulaire et a permis de mettre en évidence la présence de sous populations virales non détectables par des méthodes classiques (15–17). Elle a même récemment été utilisée sur des échantillons porcins et a permis l'identification de nombreux virus (18).

Les puces à ADN peuvent donc être considérées à la fois comme une méthode avec *a priori*, puisqu'elles s'appuient sur une technique d'hybridation de sonde spécifique. Mais également comme une technique sans *a priori* quant au grand nombre de virus différents pouvant être criblés lors d'une même réaction.

Schématisation du cheminement pour l'utilisation d'une puce :

<http://www.imm.cnrs.fr/transcriptome/spip.php?rubrique11>

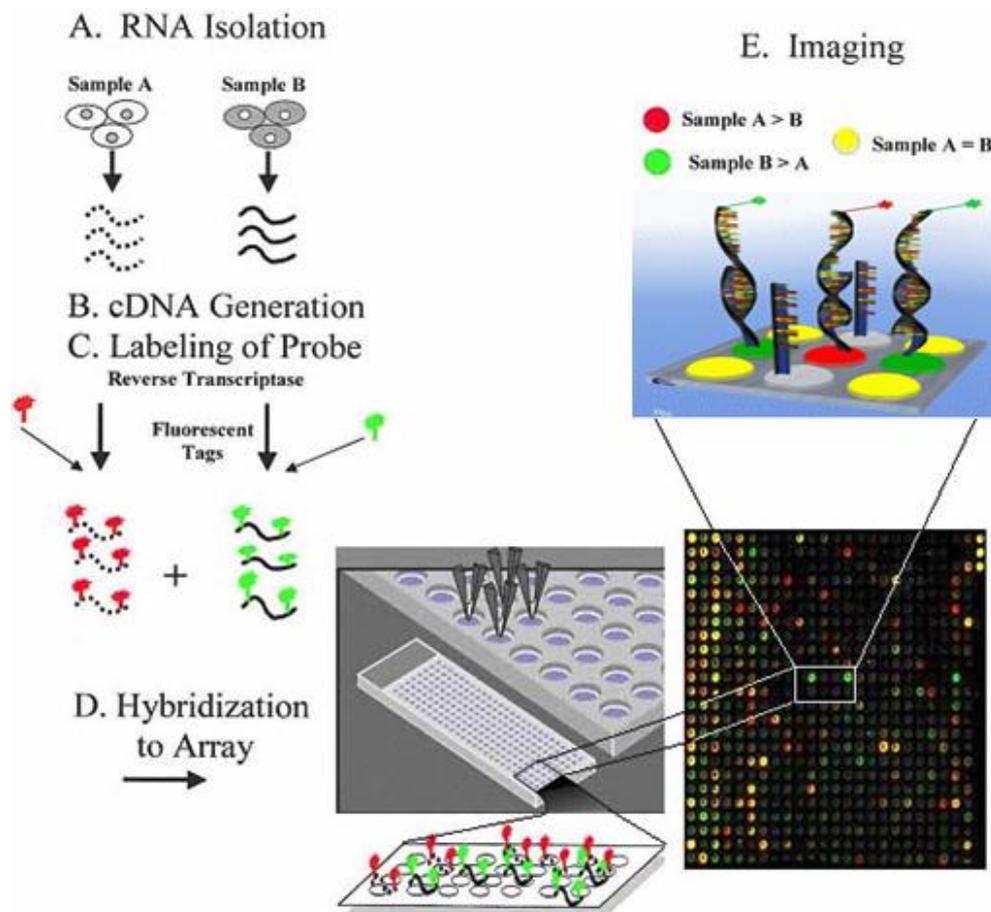


Figure 14 : Cheminement pour l'utilisation d'une puce à ADN. A. Extraction du matériel biologique. B. création des ADNc puis marquage de ces échantillons. D. Dépôt et Hybridation des ADNc marqués sur le support porteur des sondes. E. Analyse des résultats

III. Méthodes de diagnostic sans *a priori*

Les limites des méthodes de diagnostic avec *a priori* sont dues à l'approche qui sélectionne les cibles et qui ne laisse pas de place à l'inconnu. Pour effacer ces contraintes posées par les limites de ces techniques avec *a priori*, des nouvelles méthodes ne nécessitant aucun préjugé sur les agents ont vu le jour. De cette façon, les limites du diagnostic ont été repoussées permettant d'identifier de nombreux nouveaux pathogènes. Il existe un grand nombre de méthodes sans *a priori* qui sont apparues au fil des ans :

- Les isollements sur système cellulaire ou sur œuf
- La microscopie
- Le séquençage à haut débit

A. Isolement

En virologie on appelle isolement la mise à « l'écart » d'un virus. Cette mise à part implique l'élimination, autant que possible des autres acteurs considérés alors comme contaminants. Les techniques d'isolement sont des méthodes de diagnostic direct sans *a priori*. En effet contrairement aux méthodes avec *a priori* comme la PCR où il est indispensable de cibler un pathogène, les méthodes d'isolement *in vivo* ou *in vitro* ne nécessitent pas de connaître les pathogènes cibles. Ainsi un même système (exemple : une lignée cellulaire comme les cellules Véro) peut permettre à différents pathogènes de se répliquer quelle que soit leur l'origine. Cependant certaines limites peuvent poser problème notamment lié à la barrière d'espèce. En effet le but est d'infecter un système cellulaire simple ou complexe avec un pathogène pour qu'il s'y reproduise. Le système doit donc être adapté au pathogène, la majorité des virus aviaires ne pouvant pas se développer sur des cellules humaines par exemple.

1. Isolement sur culture cellulaire

La culture cellulaire est une technique couramment utilisée en laboratoire qui consiste à maintenir en vie *in vitro* des cellules. Les cellules couramment utilisées sont soit des lignées immortalisées soit des cultures de cellules primaires (créées à partir d'organes ou de tissus). L'intérêt majeur des cellules immortalisées est la possibilité de les multiplier de manière quasi illimitée. Il existe plusieurs méthodes pour immortaliser des cellules : utiliser des cellules cancéreuses ; transformer des cellules avec un vecteur (par exemple un virus) les rendant immortelles ; modifier les cellules par génie génétique. Suivant le type cellulaire employé le protocole de culture peut varier. Certaines cellules se cultivent en effet en suspension alors que d'autres sont adhérentes. De plus le milieu et les paramètres physico-chimiques telles que la température ou le pourcentage d'oxygène et de dioxyde de carbone (CO₂) peuvent varier.

L'intérêt majeur de la culture cellulaire lors d'une recherche d'agent inconnu est sa capacité à amplifier certains pathogènes viraux ce qui augmente les chances de caractérisation par d'autres techniques. L'avantage de la culture cellulaire sur méthode classique de PCR lors d'un diagnostic est qu'il n'est pas nécessaire de connaître la séquence du pathogène pour l'identifier, ce qui permet de s'affranchir de potentielles mutations ou mauvais diagnostic clinique. Cependant cette technique prend plus de temps qu'une simple PCR. Un autre avantage qu'offre une telle méthode est la possibilité d'apprécier certaines propriétés biologiques du pathogène comme ses effets cytopathiques. Cette étape d'isolement sur système cellulaire et souvent couplée à de la microscopie afin d'appréhender les caractères morphologiques de l'agent ou des agents en présence.

2. Isolement sur œuf (ovoculture)

Pour isoler ou amplifier un pathogène il est primordial de trouver le support adéquat. Parmi les méthodes d'isolement connues, l'ovoculture ou culture sur œuf embryonné est une technique très utilisée particulièrement pour les virus aviaires. Cette méthode de culture a permis la découverte de virus mais aussi simplement l'amplification de virions permettant leur caractérisation.

Le principe de cette technique est basé sur l'inoculation de fluide biologique contenant le pathogène à identifier dans l'œuf (Figure 15) dans le liquide amniotique, allantoïque, sur la membrane chorio-allantoïdienne, dans l'embryon, en intraveineuse ou dans le vitellus (jaune d'œuf).

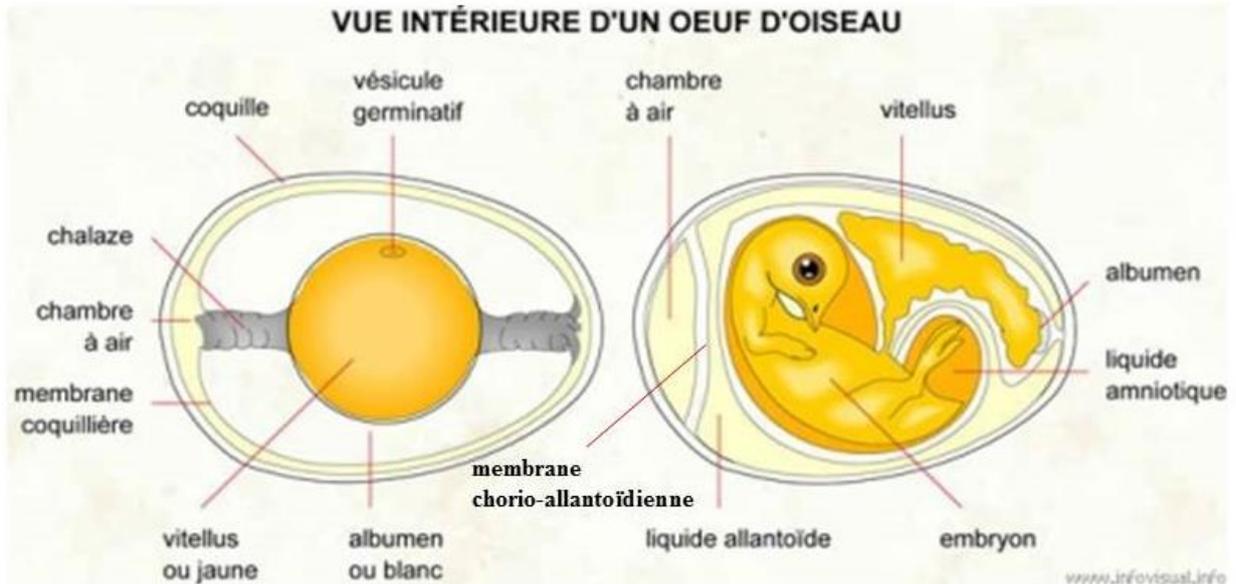


Figure 15 : schématisation des différentes parties d'un œuf embryonné. Représentation de la méthode d'injection dans le liquide allantoïque.

Tous les virus aviaires ne se développent pas de la même façon, dans la même cavité et n'induisent pas non plus les mêmes symptômes sur l'embryon. Ainsi des virus comme le virus Influenza ou le virus de la maladie de Newcastle se développent dans le liquide allantoïque et peuvent provoquer la mort de l'embryon. Alors que les poxvirus se répliquent sur la membrane chorio-allantoïdienne et induisent des lésions blanchâtres à la surface de la membrane. Les coronavirus de la dinde quant à eux se multiplient dans le liquide amniotique d'œuf de dinde. Outre la voie d'inoculation, l'origine de l'œuf est en effet un facteur pouvant influencer sur l'amplification et l'isolement du virus ou non. De cette façon il est nécessaire de choisir l'œuf en fonction du virus si il s'agit d'un virus.

B. Microscopie électronique

La microscopie a longtemps été une des seules possibilités d'identifier un agent lorsque les méthodes traditionnelles avec *a priori* n'avaient pas abouti. En effet, la visualisation de la structure du pathogène a longtemps permis de caractériser l'agent étiologique.

Les premiers prototypes de microscopes électroniques ont été conçus dans les années 1930 par des ingénieurs allemands, notamment Ernst Ruska nobélisé pour ces travaux en 1986. Grâce à deux compagnies industrielles, Siemens et Halske, la microscopie électronique s'est développée et la construction du premier microscope électronique a eu lieu entre les années 1937 et 1938 (Figure 16a (19)). Cette technologie est basée sur l'émission d'un faisceau d'électrons qui passent à travers l'échantillon. Le mode d'enregistrement de l'image a pris plusieurs formes au cours du temps et des avancées technologiques. De nos jours, l'image est enregistrée directement par un capteur d'une caméra CCD (Charge-Coupled Device) transmettant l'information en temps réel sur un ordinateur permettant ainsi la visualisation de l'image (Figure 16b).

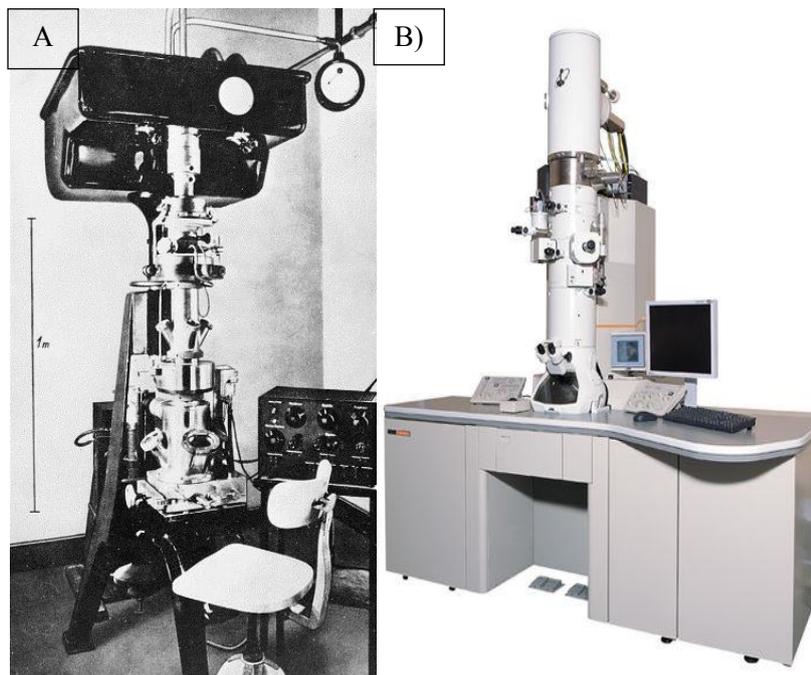


Figure 16 :

A/ Microscope électronique à haute résolution de 1938, de Siemens et Halske.
Kruger et al. 2000.

B/ Microscope à transmission électronique

<http://www.directindustry.fr/prod/jeol/microscopes-electroniques-transmission-tem-20754-695685.html>

Cet outil est très rapidement devenu indispensable et a permis la découverte et la visualisation d'un grand nombre de pathogènes qui étaient à l'époque impossible à voir. En effet, les techniques de microscopie de l'époque ne permettaient pas de voir les particules de moins de 300nm. Dans le premier rapport présentant une structure virale, Helmut Ruska (le frère du physicien Ernst Ruska) présente l'importance de la microscopie électronique en recherche en virologie (20). Le premier virus visualisé fut le virus de la mosaïque du tabac (Figure 17)(21). Les deux frères Ruska ont permis le développement de ce domaine et Helmut participa à la publication de près d'une vingtaine de rapports de microscopie, tant sur la structure de bactéries, de parasites que de virus (22) notamment des poxvirus (Figure 18). Un grand nombre de virus ont été identifiés et observés par cette méthode de microscopie comme les bactériophages en 1942 (23), de nombreux virus intestinaux (24–27) ou encore les virus des Hépatites A et B (28,29) ou le virus de l'immunodéficience acquise (30) (http://www.afd-ld.org/~fdp_viro/content.php?page=historique).

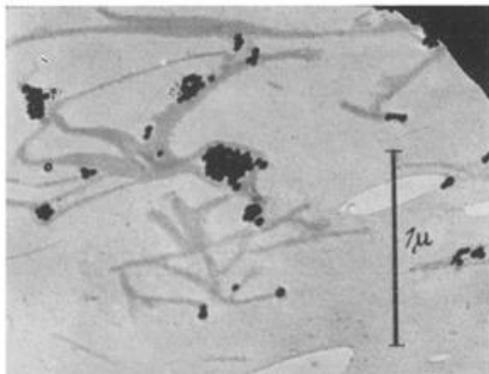


Figure 17 : Première visualisation d'un virus par microscopie électronique. Virus de la mosaïque du tabac.
Kausche et al. 1938.

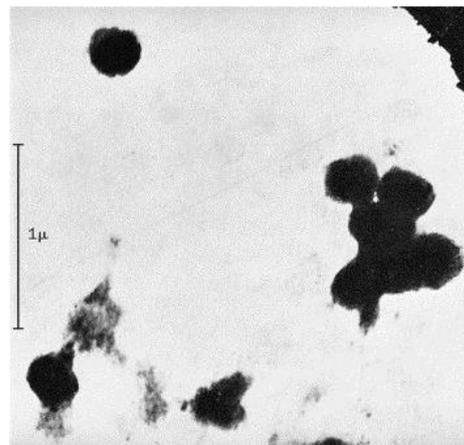


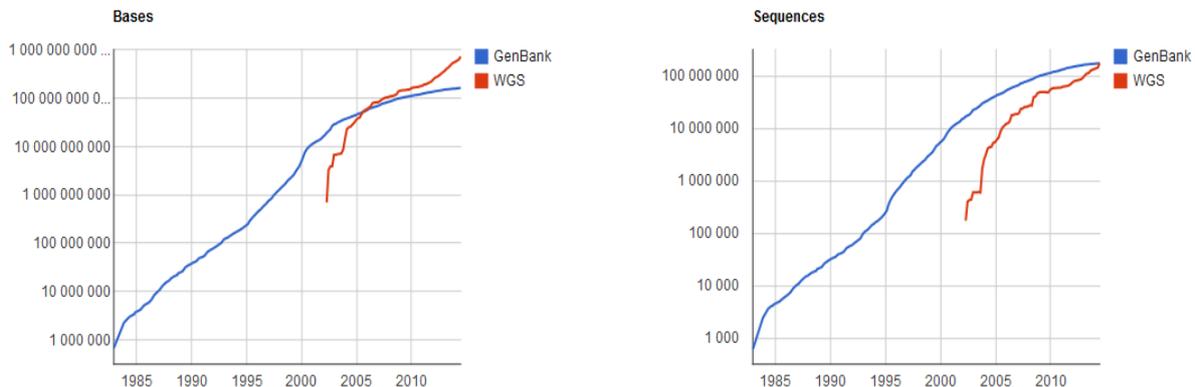
Figure 18 : photo d'un poxvirus par microscopie électronique: le Virus de la vaccine
Kruger et al. 2000.

La microscopie électronique a permis de nombreuses avancées en virologie classique : en 1948 elle a permis de différencier deux poxvirus (31), elle a également permis le début des études de relation entre le virus et son hôte (32,33). La possibilité de visualiser les virus a permis de faire des études de structure et de morphologie virale, ce qui a permis d'établir des classifications de virus basées sur leur morphologie (34,35). La rapidité de cette technique la place en première ligne notamment pour la surveillance de pathogènes viraux lors d'une potentielle attaque terroriste comme le montre un rapport du « Center for Disease Control and Prevention » (centre pour le contrôle et la prévention des maladies, CDC) (36) ou bien pour un diagnostic rapide de maladies infectieuses (37). Ainsi, le microscope à transmission électronique fut essentiel dans l'identification d'un nouveau morbillivirus responsable d'infection respiratoire mortelle en Australie en 1995 (38), pour le virus Nipah découvert dans des élevage porcin en 1999 en Malaisie (39) et plus récemment en 2003 pour l'identification de l'agent étiologique du syndrome aigu sévère respiratoire (SARS)(40).

C. Séquençage

De nos jours, le séquençage est devenu incontournable dans le monde de la science et a permis l'enrichissement des banques de données (Figure 19).

Growth of GenBank and WGS



<http://www.ncbi.nlm.nih.gov/gate1.inist.1>

Figure 19 : Enrichissement de banques de données (GenBank et Whole Genome Shotgun) en fonction des années.

1. Historique du séquençage Classique

Le séquençage que l'on connaît aujourd'hui est basé sur deux découvertes récompensées par le prix Albert Lasker de recherche médicale (*Albert Lasker Medical Research Award*) en 1979 et le prix Nobel de chimie de 1980. Ces deux techniques de séquençage sont connues sous le nom de méthode de Gilbert-Maxam et méthode Sanger. Les prémices du séquençage sont apparues fin des années 70 avec deux études visant à déterminer les séquences d'un fragment d'ADN, dont la structure avait été découverte une vingtaine d'années auparavant (Découverte de la structure de l'ADN 1953 (41)). Ces deux équipes ont travaillé sur des méthodes très distinctes l'une de l'autre, mais avec un but commun : mettre au point le séquençage. De nombreuses avancées ont permis de faire évoluer les méthodes de séquençage pour arriver aux méthodes connues de nos jours (Figure 20).

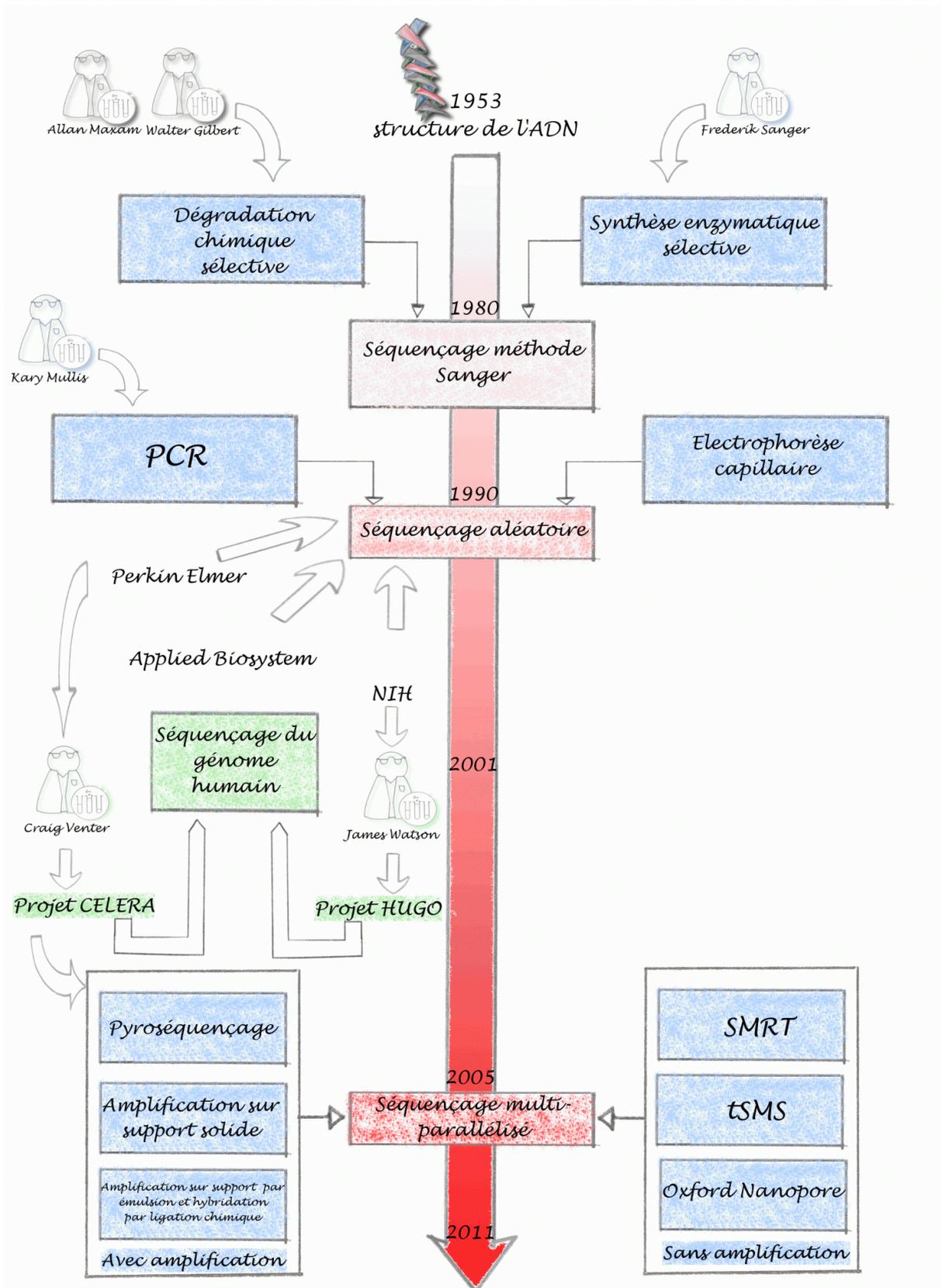


Figure 20 : Historique schématique du séquençage d'acide nucléique
<http://www.biorigami.com/wp-content/uploads/2011/07/sequencingSCHEME1-e1309507597221.qif>

(1) Méthode Gilbert-Maxam

Allan Maxam et Walter Gilbert ont développé une technique (*The Gilbert-Maxam method*) de dégradation chimique ciblant sélectivement les différentes bases (Adénine, Guanine, Cytosine, et Thymines) entraînant un clivage partiel. Leur méthode utilise de l'ADN double brin ayant été marqué à son extrémité avec un marqueur radioactif, le Phosphore 32 (P^{32}). Ce fragment marqué sera ensuite clivé par des réactions chimiques spécifiques reconnaissant chaque base. Une fois coupés ces fragments de tailles différentes sont lus par électrophorèse sur gel de polyacrylamide. Ceci permettant de déterminer l'ordre et la nature des bases et donc le séquençage de l'ADN étudié (42).

(2) Séquençage Sanger

La seconde méthode, bien plus connue, a été développée par le britannique Frederick Sanger (*The Sanger method*). Cette méthode avant-gardiste de séquençage est parue en deux volets. Tout d'abord en 1975, Sanger a mis au point une technique permettant de déterminer la séquence d'un ADN par synthèse complémentaire avec l'aide de l'ADN polymérase (43). Une première extension est faite en présence des quatre deoxyribotriphosphates marqués au P^{32} . Un double système est ensuite employé. Tout d'abord le « système du moins » (*the « minus » system*) qui consiste à enlever un des quatre deoxyribotriphosphates lors de la réaction d'extension avec l'ADN polymérase. Puis les mix sont dénaturés et séparés par électrophorèse sur gel d'acrylamide pour visualiser les différences de taille et donc estimer les séquences. Ce système ne peut pas à lui seul établir la séquence de la matrice, c'est pourquoi il est couplé à un système similaire, le « système du plus » (*the « plus » system*). Toujours en partant de la première extension, une incubation est réalisée avec un seul des quatre deoxyribotriphosphates puis, comme pour le système du moins, dénaturé et déposé sur gel acrylamide afin d'y apprécier les différences de taille. Donc par exemple, nous allons avoir dans un mix l'ADN polymérase et les dGTP, dTTP et dCTP pour le système « moins » et uniquement dATP pour le système « plus ». Sanger améliora sa méthode de séquençage, qui comportait quelques problèmes notamment pour déterminer la taille des homopolymères, en développant un système avec un « terminateur de chaîne » (44). Cette démarche utilise des analogues de terminaisons de chaînes nucléiques plutôt que les deoxy nucléoside-triphosphate (dNTP) originels. Il s'agit d'une « dideoxy » méthode. En utilisant un terminateur de chaîne comme celui-ci, l'ADN généré

par l'extension possède une base spécifique à l'extrémité 3'. Originellement deux terminateurs étaient utilisés : un arabinose triphosphate et un 2', 3'-dideoxy nucléoside triphosphate (ddNTP). Ces analogues ayant comme particularité qu'ils ne contiennent pas de groupe 3' hydroxyle, empêchent ainsi l'extension du brin précisément à l'endroit où le triphosphate devait s'incorporer. De même que pour sa méthode « plus ou moins », Sanger réalise la lecture de la séquence par migration sur gel d'acrylamide. Lors d'une réaction classique, des dNTPs marqués au P³² et un terminateur A/T/C ou G sont ajoutés au mix de réaction. Donc si par exemple un ADN matrice est mis en présence d'une ADN polymérase, de dNTPs et de ddTTPs, un mix de fragments avec un ddTTP en 3' est obtenu. En faisant de même avec chacun des nucléosides et en combinaison avec le système « plus ou moins », il est possible de déterminer le profil de la séquence des brins d'ADN nouvellement synthétisés après l'analyse du gel d'acrylamide.

En 1977, Sanger et ces collaborateurs publier le séquençage du premier génome complet d'un organisme, le ϕ 174 (44). Ce fut le premier d'une longue série, en effet de nos jours il existe près de 173.353.076 séquences publiées sur Genbank (Figure 19) (<http://www.ncbi.nlm.nih.gov/gate1.inist.fr/genbank/statistics>). Les années 80 ont vu naître le séquençage d'un grand nombre de génomes comme par exemple le SV40 (45), le phage lambda (46) ou encore le virus Epstein Barr (47). Ces années ont vu apparaître également un centre, aujourd'hui bien connu, accessible et regroupant les séquences génétiques publiées : Genbank a été créé en 1982 par le NIH (National Institutes of Health) pour regrouper les séquences génétiques publiées (48). Avec l'avancée des technologies de séquençage et l'enrichissement des banques de données, les méthodes et techniques d'analyse ont dû elles aussi aller de l'avant. Ainsi des programmes comme BLAST (49), FASTA (50) sont apparus, permettant l'identification et la comparaison des séquences déjà présentes dans les bases de données. On peut donc dire que l'apparition et l'explosion du séquençage Sanger (dans un premier temps) a fait naître la bio-informatique.

Nous sommes encore loin du séquençage Sanger que l'on connaît aujourd'hui. Cependant bénéficiant de l'invention de la PCR (51) mais également des avancées des méthodes de détection de la fluorescence (52,53), des fluorochromes eux-mêmes (54,55) et de l'électrophorèse par capillaire (56), le séquençage Sanger a peu à peu évolué vers une automatisation. De plus de grandes sociétés comme Applied Biosystems (ABI) (52), Beckman Coulter ou encore Perkin Elmer se sont employées à développer et automatiser de plus en

plus cette technologie de séquençage émergente. En 1986, lors de son premier rapport ABI montra qu'il était possible de récolter les informations de séquençage directement sur un ordinateur, sans avoir à passer les méthodes « manuelles » jusque-là utilisées (52). Toujours en s'appuyant sur les mêmes principes, il suffisait d'utiliser des fluorochromes de couleurs différentes pour chacune des bases A, C, G et T. Puis, après une co-electrophorèse dans un tube de gel de polyacrylamide, les ADNs séparés étaient analysés à proximité du fond du tube par un laser transmettant l'information sur un ordinateur. Une année plus tard ABI publia le premier programme permettant l'interprétation automatisé du séquençage (57). Cette même année, en 1987, Venter et ces collaborateurs du NIH sont les premiers à utiliser un séquenceur automatisé et ils publient la première séquence de génome utilisant le séquençage d'ADN entièrement automatisée. En 1990, le projet de séquençage humain (Human genome project) fut présenté par le NIH et le département de l'énergie américain (DOE). Ce projet ne comprenait pas seulement le séquençage du génome humain mais aussi celui d'autres modèles tels qu'*Escherichia coli*, *Caenorhabditis elegans* (58) et bien d'autres encore. C'est à cette période que naquit l'institut de la recherche génomique (TIGR) (1992). Outre l'automatisation des techniques de séquençage, l'efficacité de celui-ci a très rapidement augmenté aussi. En effet la taille des fragments séquencés par Sanger a quadruplé en l'espace de quelques années ce qui a permis de séquencer des génomes toujours de plus en plus gros. De plus des systèmes ont été développés afin de mieux estimer les erreurs de séquençage dû à l'assignement des bases, comme le score *phred* (59,60). Malgré un énorme succès de la méthode de séquençage Sanger et de l'automatisation de ces machines, cette première génération de séquenceurs reste couteuse et peu adaptée à de gros volumes d'échantillons et/ou à du séquençage de très grands génomes.

(3) Le pyroséquençage

Le pyroséquençage est un séquençage en temps réel basé sur la détection des pyrophosphates produits et non sur la fluorescence comme pour la méthode Sanger. Cette technique de séquençage s'appuie sur les travaux de Nyren (61–63) ainsi que les propriétés de l'ADN polymérase et d'une suite de réactions enzymatiques en cascade. Le pyroséquençage est une technique qui va doser la production de pyrophosphate (PPI) grâce à son utilisation dans une réaction générant de la lumière.

Comme les autres réactions de séquençage le pyroséquençage commence par l'utilisation de l'ADN polymérase. Cette réaction enzymatique va libérer du PPI qui sera à son tour utilisé dans une réaction de sulphurylase qui produira de l'adénosine triphosphate (ATP). L'ATP produite va ensuite être consommée par une réaction de luciférase émettrice de photons et donc de lumière. Dans ces conditions, l'intensité lumineuse est proportionnelle à la quantité de PPI produit et par conséquent de deoxynucléotides (dATP, dCTP, dGTP ou dTTP) incorporés. Le signal lumineux généré est mesuré par un capteur (exemples : photodiode, photomultiplicateur ou encore un dispositif à transfert de charge (*charge-coupled device*, CCD)) qui va convertir l'intensité lumineuse en série de pics dont la hauteur est proportionnelle au nombre de nucléotides ajoutés. Cette série de pics, aussi appelée pyrogramme, permet de déterminer la séquence nucléique. Deux stratégies de pyroséquençage ont été développées et publiées par Ronaghi et ses collaborateurs à deux ans d'écart : en phase solide (63) et en phase liquide (64).

Le premier système est un procédé dont la matrice est fixée sur un support solide et est nommé système en phase solide. Il nécessite une étape de lavage entre chaque ajout de deoxynucléotides, étape indispensable pour éliminer les dNTPs non incorporés mais aussi les ATPs produits. Le second système est un procédé en phase liquide, la réaction se fait en solution sans support. Pour ce système, l'étape de lavage a été substituée par l'addition d'une enzyme de dégradation de nucléotides. Malgré une révolution technologique et de nombreux avantages tel que :

- un coût moins élevé que les technologies classiques de l'époque
- l'absence de gel d'électrophorèse, pas d'amorces marquées ni de dNTPs marqués
- une facilité d'adaptation avec du multiplexage ;

Le pyroséquençage garde encore de nombreux inconvénients :

- les étapes de lavage pour le système en phase solide entraînent une perte de signal
- l'activité enzymatique de l'apyrase diminue au fil des cycles
- des problèmes de bruit de fond peuvent être observés lors de la contamination de PPI

- l'extension incomplète de l'ADN polymérase joue un rôle sur la taille des fragments séquencés mais aussi sur le ratio signal/ bruit de fond
- des problèmes de « mismatch » des amorces. Une solution a été publiée en 1998 et suggère l'utilisation de système « tige boucle » (64).
- l'incapacité de la technologie de quantifier les homopolymères de plus de 5 répétitions à cause de la luminosité qui n'est plus linéaire au-delà de ces 5 répétitions.

Dans les années 2000 de nombreuses équipes ont utilisé le pyroséquençage pour différentes applications. Comme par exemple la détection de mutations sur un exon du gène p53 ou de SNPs (polymorphisme d'un seul nucléotide) chez *Mycobacterium tuberculosis* ou encore l'analyse de la structure secondaire de l'ADN (65–70).

Le pyroséquençage a ouvert de nombreuses nouvelles possibilités d'analyse de séquences d'ADN, notamment grâce aux prémices des séquenceurs hauts débits automatisés, tel le PSQ 96 system (www.pyrosequencing.com). La rapidité de séquençage de cette méthode (comparativement au séquençage Sanger) a non seulement permis le re-séquençage de gènes mais également l'identification d'éventuelles mutations par les multiples pyrogrammes générés permettant la mise en évidence des variations nucléotidiques. Cette stratégie a été utilisée avec succès pour le re-séquençage et la mise en évidence du gène P53 suppresseur de tumeur (68). Le re-séquençage n'est pas la seule opportunité qu'offre le pyroséquençage, les caractéristiques de la méthode permettent un typage extrêmement facile, précis et robuste. Ainsi l'identification de polymorphismes de simple nucléotide (SNP) est l'une des autres possibilités que propose cette technologie (67,69,70). De la même façon, si le pyroséquençage est capable de détecter des mutations uniques, cette application peut être utilisée afin de mettre en évidence l'apparition de résistance des pathogènes aux traitements (71). Outre les mutations de pathogènes déjà connues, cette technique peut aussi permettre l'identification de nouveaux pathogènes (72,73). La grande variabilité de séquences de certains pathogènes les rend très difficiles à identifier par des méthodes classiques comme la PCR, lors de diagnostic classique. Cette méthode est capable de mettre en évidence des pathogènes méconnus ou déjà identifiés et permet d'estimer la diversité des pathogènes dans des échantillons donnés (74,75).

2. Haut débit

Le séquençage a pris un tournant avec l'apparition de nouvelles méthodes mettant fin à la suprématie du séquençage par la méthode dideoxy (Sanger). En effet les nouvelles méthodes de séquençage massif ont considérablement augmenté le nombre de séquences générées. Cependant l'augmentation du nombre de séquences a été accompagnée de la diminution de la taille des amplicons produits. En 2005, Margulies et ces collaborateurs ont présenté une méthode de séquençage s'appuyant sur le pyroséquençage (76). C'est dans cette étude que fut présenté le premier séquenceur massif, le *454 life science* (454 Life Sciences/Roche, Branford, CT). Un an après, c'est Solexa (société rachetée par Illumina en 2007 (Illumina, Inc., San Diego, CA)) qui commercialise son premier séquenceur utilisant une autre technologie de séquençage, le *genome analyzer* (77). Puis c'est au tour de PacBio (Pacific Biosciences, Menlo Park, CA) (78) et enfin le dernier en liste est Ion Torrent (Ion Torrent/Life Technologies, South San Francisco, CA) (79). Ces quatre plateformes sont capables de produire des informations similaires dans l'ensemble (80,81), cependant quelques différences comme la taille, la qualité et la quantité des séquences générées sont à noter. Le nombre de séquences attendues en fonction de l'application que l'on souhaite est un facteur à prendre également en compte lorsque l'on choisit l'une de ces plateformes de séquençage. En outre toutes ces plateformes de séquençage sont en perpétuelle amélioration et de plus en plus utilisées en recherche (82–86).

(1) 454 Roche

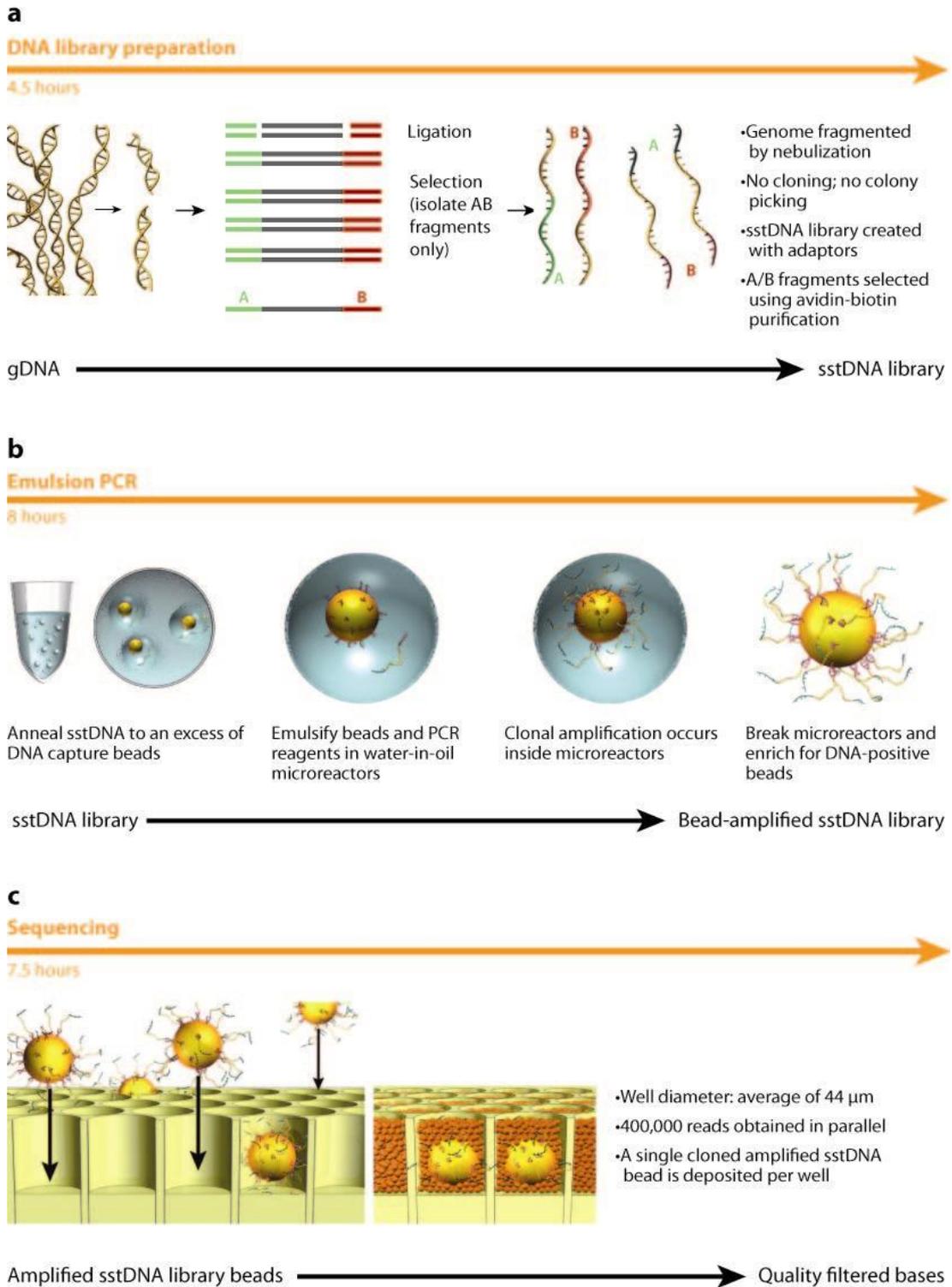
a. *Historique et principe*

La première machine de séquençage dit « massif » commercialisée fut le *454 life science* (76) en 2005. Ce séquenceur est le premier de la deuxième génération de séquenceur. Basé sur la méthode de pyroséquençage, le *454* utilise une technologie de système d'amplification en émulsion (Figure 21). Le principe de ce séquençage en émulsion peut se décomposer en quatre étapes basiques :

- 1. la création d'une « librairie » d'ADN simple brin matrice
- 2. une amplification clonale de la librairie en émulsion
- 3. le séquençage et la génération de métadonnées

- 4. l'analyse des données par des outils bio-informatiques

Les ADNs complémentaires (ADNc) sont tout d'abord fragmentés aléatoirement par nébulisation afin d'avoir une matrice de petite taille. Puis des adaptateurs sont fixés aux extrémités des ADNs matrices, ces adaptateurs sont utiles non seulement pour la purification mais aussi pour la quantification, l'amplification et l'étape de séquençage. Une pré-amplification de la matrice est alors faite avec des amorces spécifiques contenant la séquence des adaptateurs 454. Puis vient l'étape d'amplification clonale en émulsion. Les fragments sont alors attachés à des billes sous une condition ne permettant l'attachement que d'un seul fragment par bille. Chacune de ces billes est capturée dans une gouttelette de mélange réactionnel de PCR en émulsion d'huile. Des micros réactions d'amplification en émulsion se font alors à la surface des billes de capture dans ces gouttelettes. Cette PCR en émulsion va ainsi permettre la création de millions de copies clonales de chaque fragment attachés à la surface des billes (Figure 21). Après rupture de l'émulsion, les ADNs portés par les billes sont dénaturés. Les billes ne portant alors plus que des clones d'ADNs simple brin sont alors déposées dans des puits de capteur optique. Chaque puits ne peut accueillir qu'une seule bille, et donc qu'un seul clone d'ADN. De plus petites billes, porteuses d'enzymes nécessaires à la réaction de séquençage, sont déposées dans chacun des puits. La technologie 454 s'appuyant sur le pyroséquençage, chaque ajout de nucléotide entraîne une réaction en chaîne aboutissant à un signal lumineux enregistré grâce à une caméra et à des capteurs optiques. Ces millions de copies de séquences donnent des pyrogrammes qui seront analysés par bioinformatique (Figure 22). La qualité du signal est appréciée afin de ne garder pour l'analyse que les séquences de bonne qualité. De plus, afin de réduire la quantité de mauvaise assignation du signal lumineux, des systèmes (analogues au système de Phred score (Ewing et al. 1998), couramment utilisé pour le séquençage Sanger) d'estimation du taux d'erreur ont été développés (Marguiles et al. 2005).



AR Mardis ER. 2008.
 Annu. Rev. Genomics Hum. Genet. 9:387–402

Figure 21 : Schématisation du principe de la technologie de pyroséquençage 454.
 A. préparation de la librairie.
 B. principe de la PCR en émulsion.
 C. étape de séquençage.

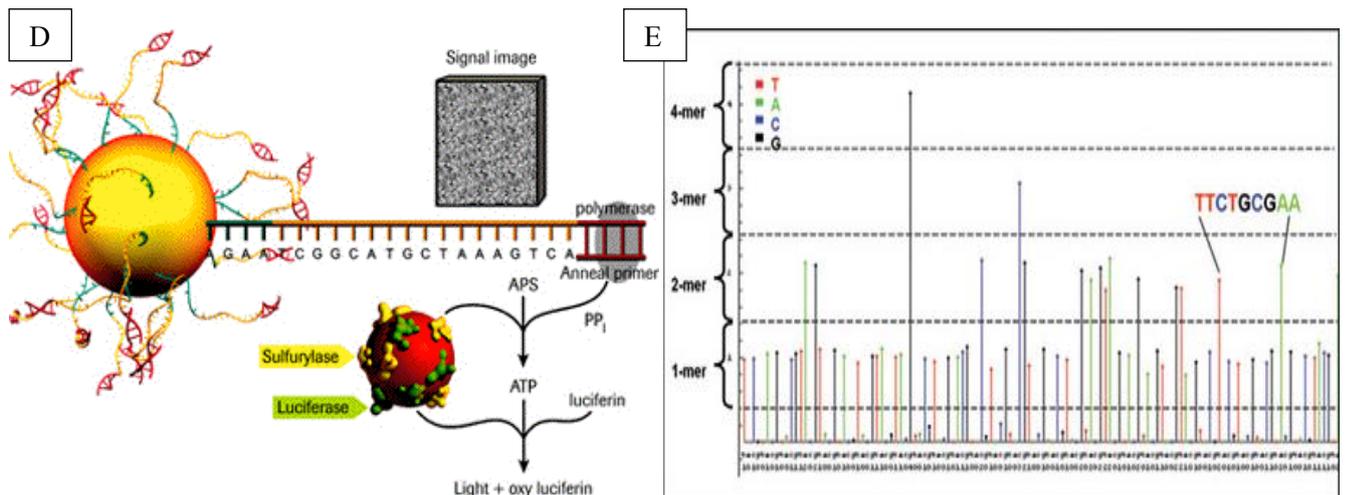


Figure 22 : Schématisation du principe de la technologie de pyroséquençage 454.

D. Représentation de la chimie du séquençage.

E. Analyse des pyrogrammes générés.

Au fil des années, des avancées technologiques ont permis de faire évoluer la méthode 454 diminuant son taux d'erreur et permettant aussi l'allongement de la taille des brins séquencés. Taille qui a permis à ce séquenceur de se placer en tête de la catégorie des séquenceurs à haut débit. De nos jours le 454 comprend 2 types d'appareils (Figure 23):

- GS FLX titanium
 - o GS FLX titanium XL + (environ 1 million de lectures, taille des fragments séquencés pouvant aller jusqu'à 1.000 pb)
 - o GS FLX titanium XL R70 (environ 1 million de lectures, taille des fragments séquencés pouvant aller jusqu'à 600 pb)
- GS junior
 - o GS junior + (environ 100.000 lectures, taille des fragments séquencés d'environ 700 pb)
 - o GS junior (environ 100.000 lectures, taille des fragments séquencés d'environ 400 pb)



GS Junior+ System

The GS Junior+ System brings the power of 454 Sequencing Systems directly to the laboratory benchtop.



GS FLX System

The GS FLX+ System features the unique combination of long reads, exceptional accuracy and high-throughput, making the system well suited for larger genomic projects.

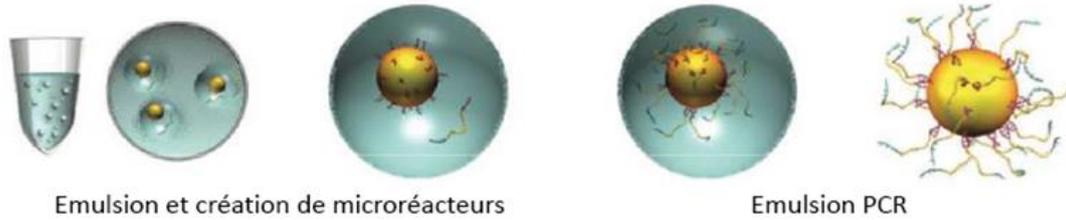
Figure 23 : Représentation des séquenceurs Roche utilisant la technologie 454.
<http://454.com/products/index.asp>

Ces machines sont capables de générer des lectures de taille allant d'environ 400 paires de bases (pb) pour les plus petites à 1.000 pb pour les plus performantes. Il est important de noter que les tailles des séquences générées ne sont pas fixes. Elles dépendent de la fragmentation mais également des amplicons eux-mêmes. Cette deuxième génération de séquenceur a permis de combler les lacunes de la première, représenté par la technologie Sanger, en diminuant le coût de la séquence tout en augmentant le volume de données générées. En effet, la première génération de 454 générait environ 20Mb (Méga-base)/run d'une taille d'environ 100 pb/lecture en 2006, et aujourd'hui ce séquenceur est capable de générer près de 700Mb/run d'une taille moyenne de 700pb/lecture.

(2) Illumina

La technologie illumina a été initié par la société Solexa avec la sortie d'un nouvel appareil de séquençage en 2004 (77). Cette méthodologie ne s'appuie pas sur le même principe que le 454. Pour cette technologie illumina la matrice est immobilisée sur un support solide et l'amplification se fait par création de ponts (Figure 24).

– Emulsion huile/eau puis ePCR (emulsion PCR)



– Immobilisation sur phase solide puis bridge amplification

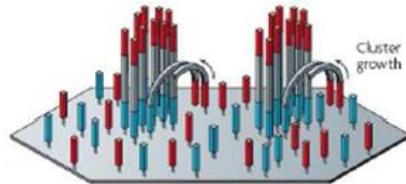


Figure 24 : Représentation des différences entre le séquençage en émulsion (Roche 454) et le séquençage par amplification en pont (illumina).

a. Historique et principe

L'histoire du séquençage illumina commence en Angleterre dans le milieu des années 90, à Cambridge dans un laboratoire qui a contribué à la première ébauche de génome humain. Cette équipe travaillait notamment sur des nouveaux fluorophores et a eu l'idée d'un nouveau concept de séquençage massif. Solexa fut fondée en 1998. Elle a d'abord acquis technologies comme « Manteia's molecular clustering technology » en 2004 permettant l'amplification d'une seule molécule d'ADN sur un cluster afin d'augmenter la fidélité du signal. En 2005 Solexa acquiert la compagnie « Lynx Therapeutic ». Un an après, en 2006, Solexa lance son premier appareil, le « Genome analyser », qui peut produire près d'1 gigabase de données en un seul run. En 2007 Illumina fait l'acquisition de Solexa qui a permis cette année le séquençage de près d'une centaine de génomes d'origine différente (plantes, microorganismes, génomes d'animaux). La technologie Illumina n'a pas cessé de s'améliorer elle aussi, avec de nouvelles machines pouvant générer plus d'un terabase par run.

La technologie illumina s'appuie sur un séquençage massif en parallèle utilisant les propriétés du séquençage par synthèse : chaque base est détectée après incorporation lors de la synthèse du brin complémentaire. Bien que la technologie Illumina soit complètement différente de celle précédemment décrite, elle suit un schéma similaire (Figure 25 (87)).

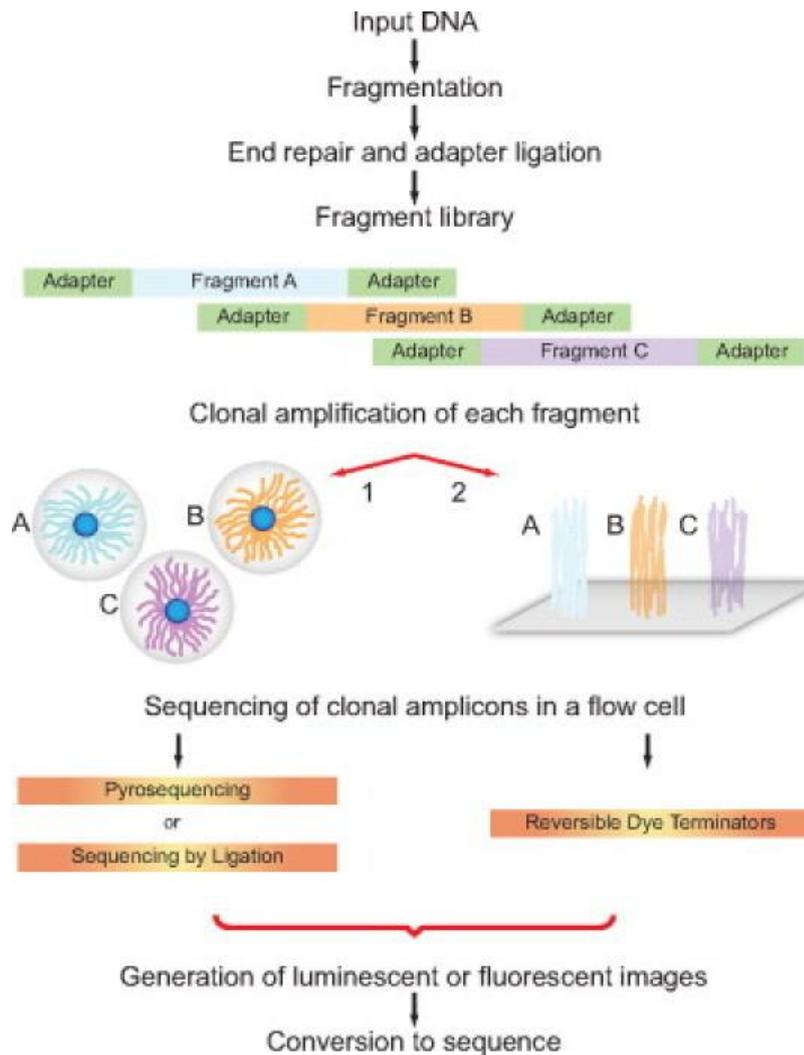


Figure 25 : Processus de séquençage pour les technologies Roche et Illumina. La technologie Roche (amplification en émulsion) est représentée sur la gauche, la technologie Illumina est représentée sur la droite de cette figure.

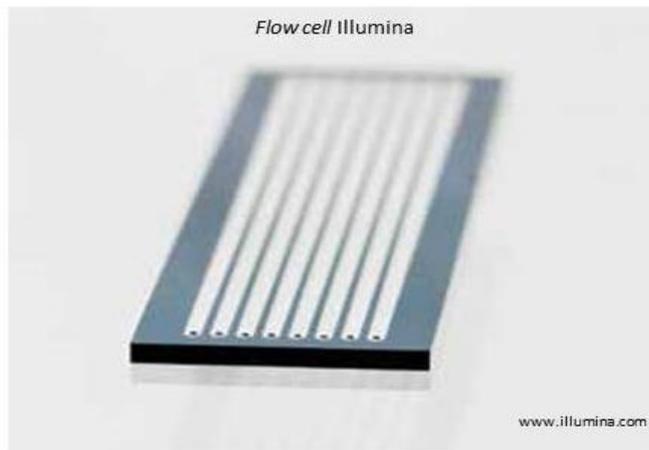
Voelkerding et al. 2010

Le principe de ce séquençage peut être simplifié en 6 étapes :

- Préparation de la librairie (des banques d'ADN)
- Fixation de la matrice sur la « flow cell » (cellule)
- Amplification en pont
- Séquençage
- Acquisition et analyse de l'image (base calling)

- Analyse des données générées

Tout commence par la fragmentation aléatoire des ADNs génomiques, puis des adaptateurs sont liés à chaque extrémité des fragments. Ces fragments sont ensuite attachés aléatoirement à la surface d'un support solide appelé « flow cell ». C'est sur cette cellule que les matrices seront multipliées par un système d'amplification en pont (Figure 26). Pour cela des ADNs simple brin vont dans un premier temps être attachés à la surface de la flow cell grâce aux adaptateurs situés à leurs extrémités. La première étape d'amplification va permettre la création de ponts doubles brins puis, après dénaturation, une nouvelle étape d'amplification pourra être initiée pour obtenir à nouveau des ADNs simple brin. Cette étape d'amplification aboutit à l'obtention d'un cluster, ou groupe d'ADNs clonaux autour de l'ADN initialement fixé à la cellule. Une fois l'étape d'amplification achevée vient la phase de séquençage. Pour cela les ADNs sont dénaturés et l'extrémité 3' est bloquée afin d'empêcher toute nouvelle formation de pont. Des amorces sont hybridées sur la matrice et des nucléotides marqués sont ajoutés avec la polymérase. Les nucléotides possèdent également un terminateur réversible bloquant la synthèse après leur incorporation. Ce nucléotide incorporé est excité par un laser permettant l'émission de la fluorescence de celui-ci. L'acquisition de la couleur est faite en temps réel à chaque cycle pour chaque assimilation de nucléotide. Chacun des nucléotides (A/T/G/C) est en effet porteur d'une couleur différente permettant l'analyse de l'image acquise. La séquence est déduite par différenciation de la couleur des fluorochromes, ce qui est appelé « base calling ». Une fois l'acquisition finie, le terminateur est clivé ce qui permet l'ajout d'un nouveau nucléotide qui bloque à son tour la synthèse. Ces cycles de séquençage sont répétés et permettent de déterminer la séquence des fragments. Une fois le séquençage des groupes de séquences (cluster) terminé, les données sont analysées en passant dans un premier temps par une série de filtres servant à vérifier la qualité des séquences et éliminer celles qui ne correspondent pas à l'attente minimum fixée par le constructeur. Les séquences seront alors alignées sur des génomes de référence avec une stringence plus ou moins importante, suivant les attentes des utilisateurs (Figure 27).



Flow cell : cellule d'analyse pour le système illumina, il est possible d'ajouter plusieurs échantillons sur une même cellule. Elle comprend 8 lignes et permet l'analyse simultanée des échantillons sur la flow cell.

Figure 26 : Représentation d'une cellule de séquençage Illumina dite « flow cell ».

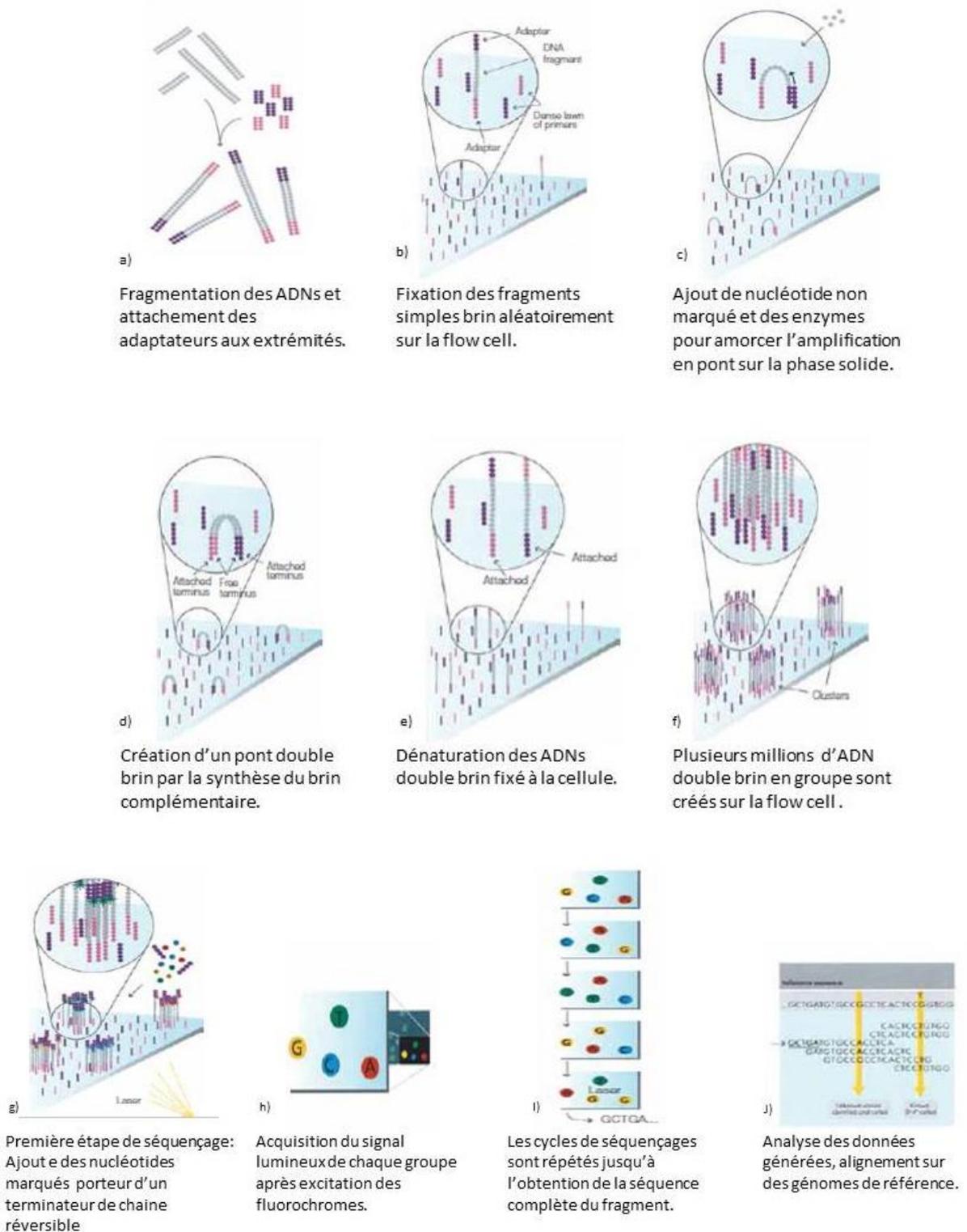


Figure 27 : Schématisation du principe de la technologie de séquençage Illumina

D'après <http://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.ilmn>

A ses débuts cette technologie n'est possible que sur une seule machine, le Genome analyser (GA). Le HiSeq a ensuite vu le jour permettant de générer une plus grande quantité de séquences, suivi par une version plus petite, le MiSeq, qui génère moins de séquences, mais de plus grande taille et beaucoup plus rapidement. Récemment une nouvelle machine a vu le jour : le NextSeq (Figure 28).



Machine	GA	MiSeq	HiSeq	NextSeq
Taille	2 x 100 pb	2 x 300pb	2 x 125 pb	2 x 150 pb
Quantités	95 Gb	15 Gb	1000 Gb	120 Gb
Nombre de lectures	640 M	25 M	4000 M	400 M
Temps	14 jours	8h30 heures	2.5jours	30 heures

Figure 28 : Représentation des séquenceurs Illumina avec leurs performances.

De gauche à droite, Genome Analyser, MiSeq, HiSeq, NexSeq.

D'après <http://systems.illumina.com/systems.ilmn> daté du 10/10/2014

(3) Life technologie

Issue de la fusion de deux géants de l'industrie (Applied Biosystems et Invitrogen) en 2008, Life technologie a fait un grand pas dans le séquençage de nouvelle génération en 2010 avec l'acquisition d'Ion torrent. C'est en Février 2010 qu'Ion torrent lance sa première machine de séquençage (88). La technologie Ion torrent s'appuie elle aussi sur la méthode de séquençage par synthèse et sur l'amplification en émulsion (PCR en émulsion) comme d'autres plateformes. Sa méthode de séquençage est quant à elle est très différente. Tout comme le pyroséquençage qui utilise les pyrophosphates produits lors des réactions d'amplification, Ion torrent s'appuie sur les réactions d'incorporation des bases et les substrats générés, plus particulièrement sur les ions hydrogènes (H^+) relâchés pendant la réaction (figure 29). Ce relargage d'ions change le pH de la solution indiquant la nature de la base incorporée. Cette particularité permet à ce type de séquençage de se passer de système de lecture optique, réduisant ainsi drastiquement le temps d'un « run ». Ce système de séquençage est aussi appelé séquençage semi-conducteur (semiconductor sequencing) et il est disponible sur deux types d'instruments : l'Ion Personal Machine (PGM) et l'Ion Proton. De plus, cette technique utilise des puces traduisant les informations biologiques en information numérique (A/T/G/C deviendront alors 0/1).

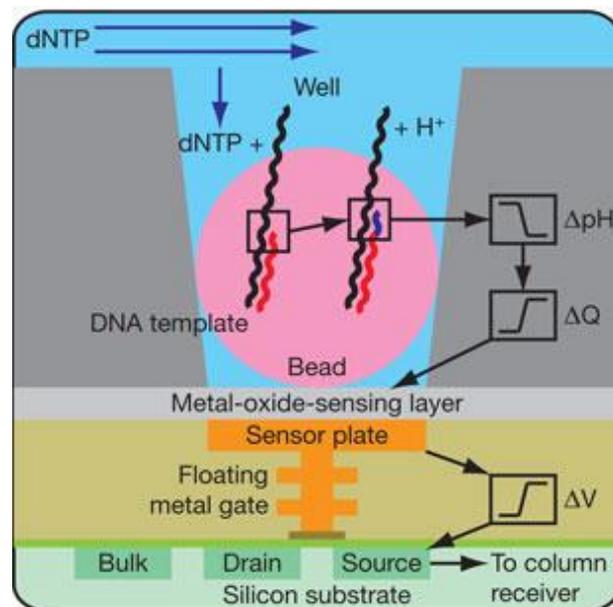


Figure 29 : Schématisation du relargage de l'ion H^+ durant la réaction de séquençage par la technologie semi-conductrice d'ion torrent.

<http://biochimej.univ-angers.fr/Page2/COURS/9ModulGenFoncVeg/5MethEtudGenFonc/1MethodeSEQUENGAGE/1SEQUENGAGE.htm>

a. *Principe et méthode*

Il est possible de décomposer le principe de ce séquençage en 4 grandes étapes :

- Préparation des librairies
- Préparation et amplification de la matrice de séquençage
- Séquençage
- Analyse des données

La préparation de la librairie a pour but de lier les adaptateurs aux extrémités de l'ADN que l'on souhaite séquencer. Ce processus commence par la récupération des ADNs ou ADNc et la fragmentation de ces derniers afin d'uniformiser la matrice (généralement entre 200 et 400 pbs suivant, suivant la chimie utilisée). Ensuite les adaptateurs sont liés à la matrice et il est possible de multiplexer les échantillons grâce à un système de code barre permettant la différenciation des fragments après le séquençage. Il est important de vérifier la concentration d'ADN afin d'avoir un ratio permettant la fixation d'un fragment par bille lors de l'amplification en émulsion.

L'étape de préparation de la matrice de séquençage se divise en deux sous étapes. Tout d'abord, une amplification utilisant le principe de l'amplification clonale en émulsion et la fixation de la matrice à la surface des sphères magnétiques. Les fragments d'ADN générés par l'étape de création de librairie vont être séparés afin que chaque fragment soit fixé à une bille avec un ratio d'un fragment par bille. Ces sphères portent à leur surface des amorces complémentaires aux adaptateurs permettant la fixation par l'extrémité. Cette solution est ensuite mélangée à de l'huile afin de former une émulsion de microgouttelettes. Ceci permet l'amplification en émulsion, comme pour la technologie 454, avec une bille par microgouttelette. Une fois l'amplification monoclonale réalisée, l'émulsion est cassée et un processus d'enrichissement est effectué (Figure 30).

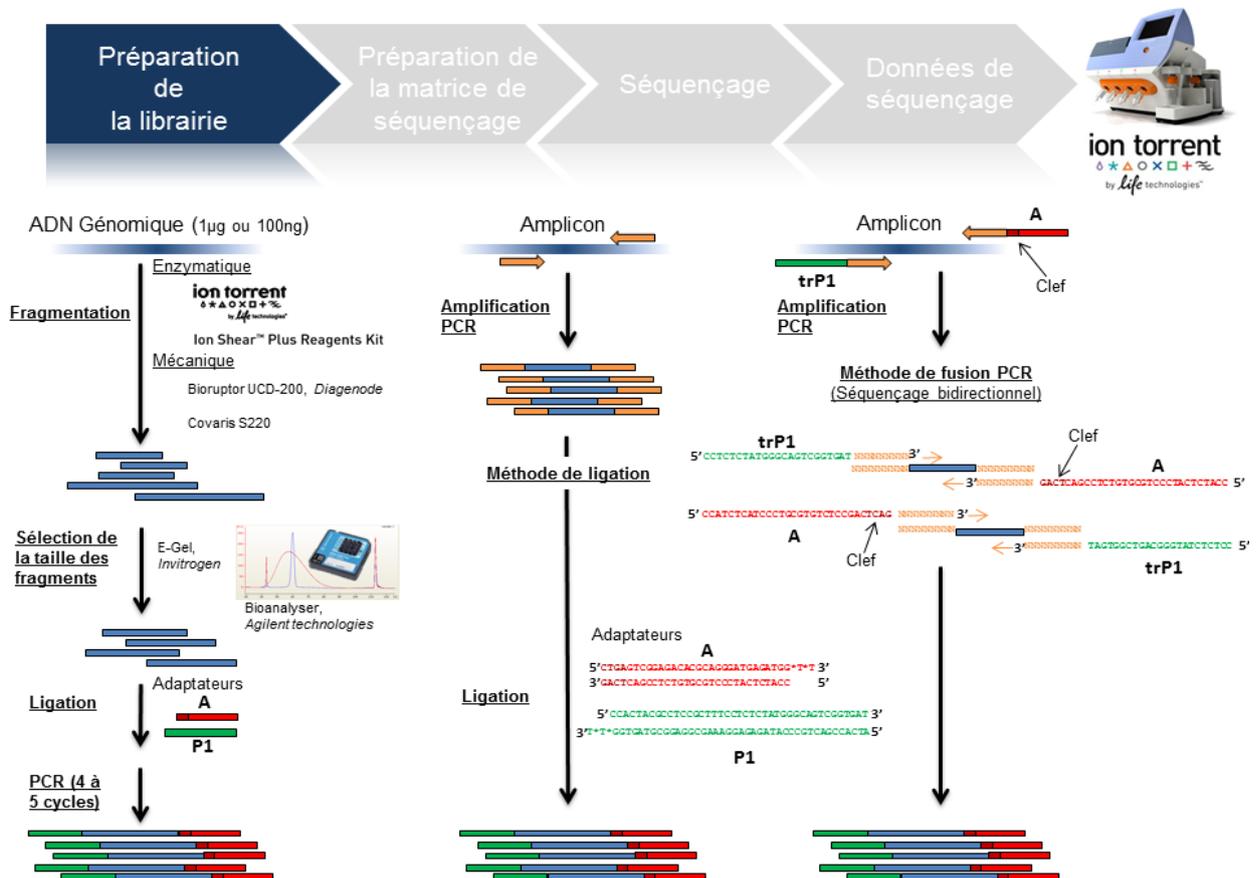


Figure 30 : représentation de l'étape de préparation des librairies de la technologie Ion torrent (semiconducteur).

http://www.bioriqami.com/wp-content/uploads/2013/06/Workflow-sequençage-PGM_Sequen%C3%A7age2.png

Chaque bille sera alors déposée dans un puits sur la puce de séquençage. Utilisant les bases de la chimie de pyroséquençage, les nucléotides seront ajoutés un par un et incorporés grâce à une polymérase. Cependant la technologie Ion torrent ne fonctionne pas comme les autres plateformes : elle mesure directement les changements de pH lors de la réaction de polymérase. Pour cela chaque puits est « noyé » dans une solution de PCR composé d'un seul nucléotide. A chaque incorporation de nucléotide lors de la formation du brin complémentaire, des ions hydrogènes sont libérés. La variation de pH induite par la libération d'ions hydrogène (majoré par la présence de millions de copies clonales par puits) est mesurée et convertie en voltage après chaque ajout de nouveau nucléotide. Le voltage est ensuite converti en nombre de bases ajoutées. Par exemple, si le nucléotide ajouté n'est pas incorporé alors il n'y aura pas de libération d'hydrogène et il n'y aura pas de modification du pH (Figure 31).

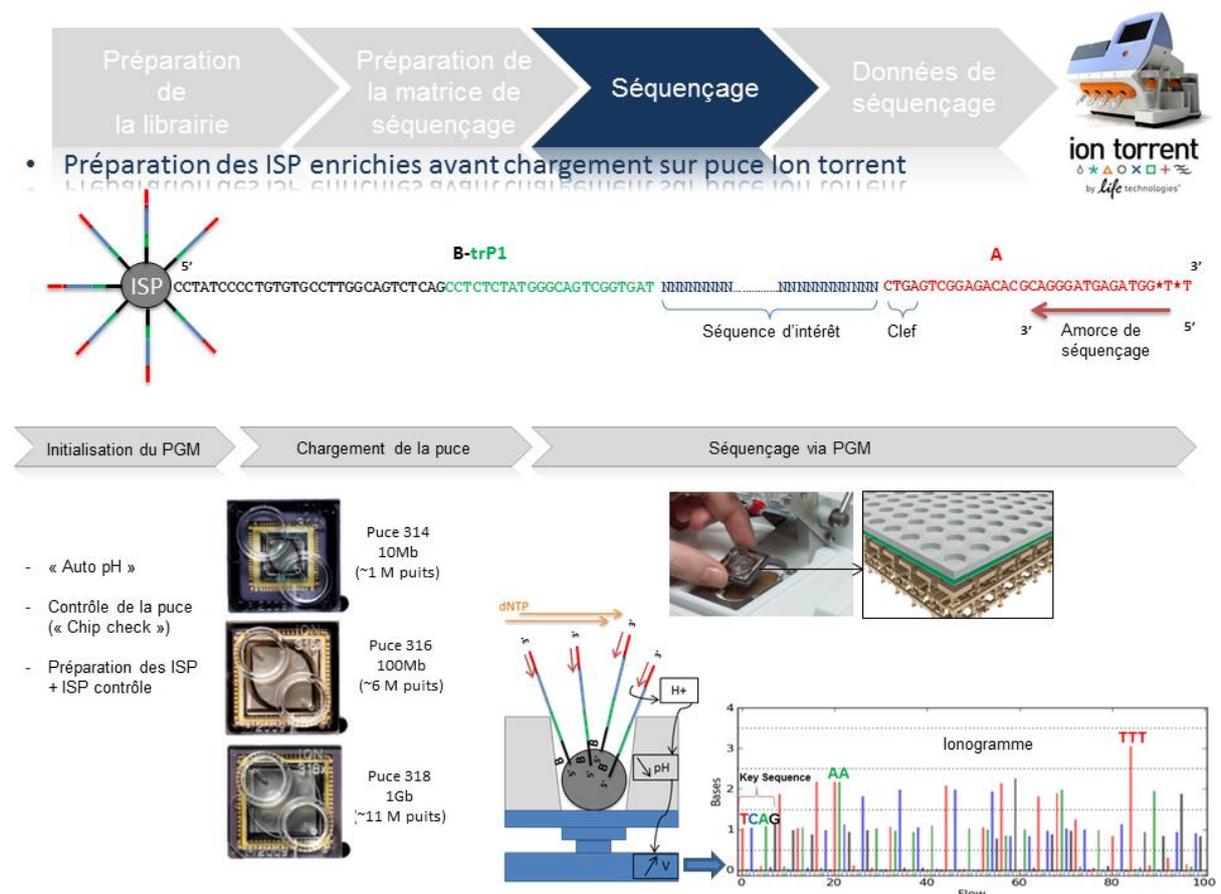


Figure 31 : Représentation de l'étape de séquençage de la technologie Ion torrent (semiconducteur). http://www.biorigami.com/wp-content/uploads/2013/06/Workflow-sequencage-PGM_Sequen%C3%A7age2.png

Comme le pyroséquençage, cette méthode présente des difficultés pour identifier les longs homopolymères.

L'étape d'analyse des données démarre par l'application de différents filtres, ainsi que par la détermination de la qualité des données générées (Phred score). Dans le but de faciliter l'analyse des métadonnées, Ion torrent a développé et continue de développer des logiciels aidant les utilisateurs. De façon générale, les séquences générées sont alignées contre des séquences de référence afin d'assigner ces dernières (Figure 32).

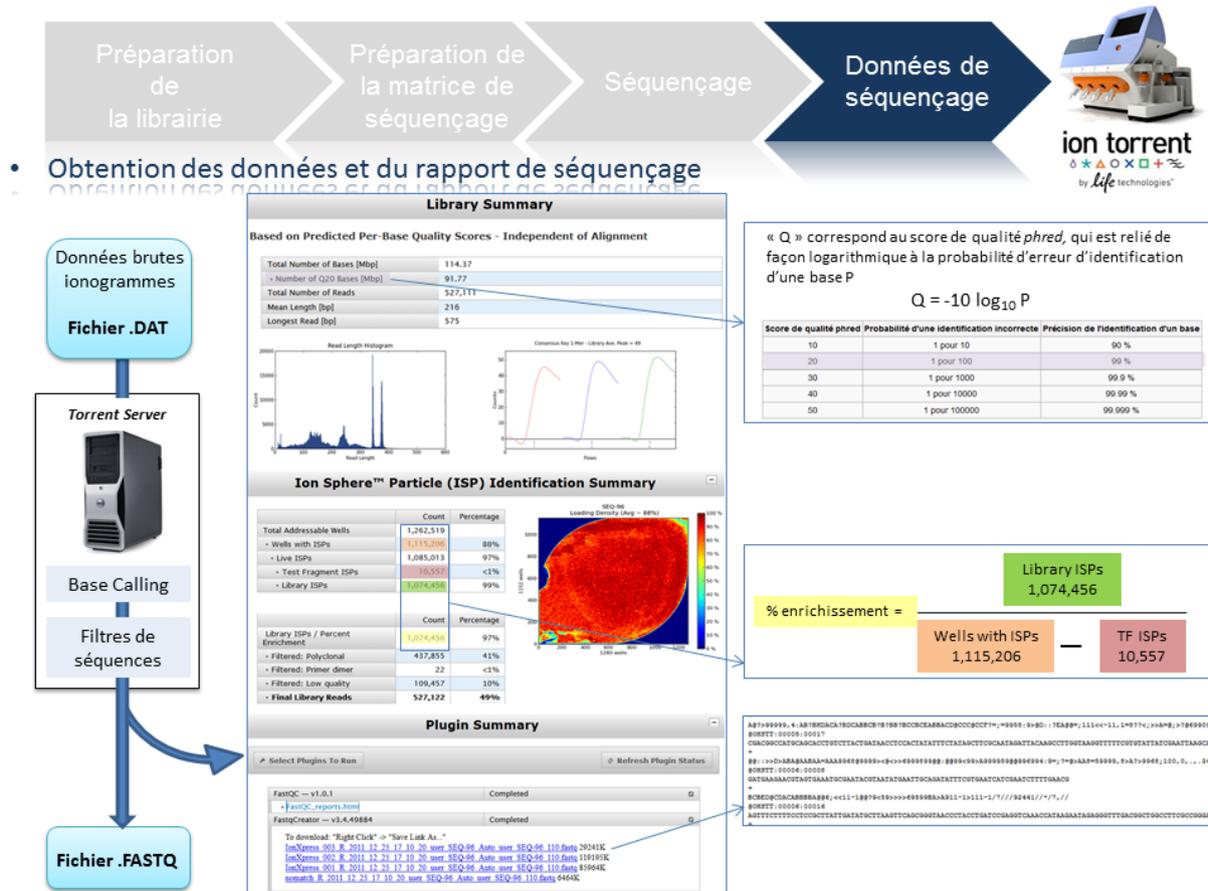


Figure 32 : Analyse des données, technologie Ion torrent

http://www.biorigami.com/wp-content/uploads/2013/06/Workflow-sequencage-PGM_Donnees-de-Sequen%C3%A7age.png

Au fil des ans, Ion torrent a développé et amélioré ces séquenceurs comme le montre cette figure sur l'évolution des machines (Figure 33):

	PGM 314	PGM 316	PGM 318	PI	PII (est. mid-2014)
# of sensors	1.2M	6.1M	11M	165M	660M
Total output	up to 100Mb	up to 1Gb	up to 2Gb	~10Gb	~32Gb (at launch)
Run time	2-4 hrs	3-5 hrs	4-7 hrs	2-4 hrs	2-4 hrs
Output/day*	up to 200Mb	up to 2Gb	up to 4Gb	~20Gb	~64Gb
Avg read length	up to 400b	up to 400b	up to 400b	up to 200b	100b
# of single reads	up to 0.6M	up to 3M	up to 5.5M	up to 82M	up to 330M
Chip price	\$99	\$299	\$499	\$699	~\$699
Reagent price**	\$250	\$250	\$250	\$300	~\$300
Instrument price	\$50k	\$50k	\$50k	\$149k	\$149k

* assumes two runs/day
 ** template prep plus sequencing per chip

Figure 33 : Evolution des séquenceurs utilisant la technologie du séquençage semi-conducteur.
<http://allseq.com/knowledgebank/sequencing-platforms/life-technologies-ion-torrent>

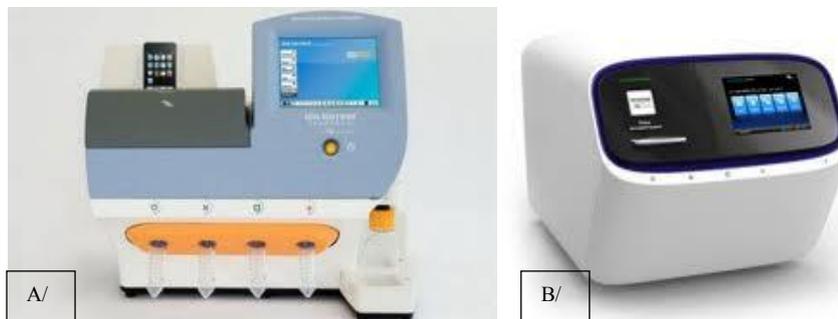


Figure 34 : Illustration des séquenceurs utilisant la technologie Ion Torrent.
 A/ Personal Genome Machine.
 B/ Ion Proton

Ces deux types de séquenceurs partagent la même technologie pourtant ils ne sont pas utilisés pour les mêmes applications de par la différence de quantité de données qu'ils génèrent (Figure 34).

Malgré une arrivée sur le marché un peu plus tardive que les autres plateformes, la technologie de séquençage semi-conductrice Ion torrent a déjà conquis de nombreuses équipes autant pour des approches de recherche de mutations (89–92), que pour le diagnostic d'une pathologie (71), pour le séquençage de génomes complets (93,94), pour l'identification de pathogènes (95,96), et même dans des analyses de changement de populations bactériennes (97).

3. Outils d'analyses

Les grandes avancées des technologies de séquençage ont permis d'accroître nos connaissances et nos banques de données. Cependant, la quantité infiniment plus importante de données générées par un *run* s'accompagne d'une difficulté d'analyse. C'est pourquoi de nombreuses plateformes se sont développées afin de faciliter les études utilisant ces nouvelles technologies révolutionnaires. Les étapes d'analyse après un séquençage à haut débit peuvent se diviser en trois grandes parties :

- Les contrôles post séquençage.

Cette étape correspond à l'estimation des erreurs dues au séquençage. Elle comprend le contrôle de la qualité des séquences, la recherche des duplicats et l'application d'un filtre pour éliminer des séquences dites contaminantes (séquences ADN génomique de l'hôte, bactériennes..).

- L'annotation des séquences

Cette étape est la plus importante dans l'analyse des données générées par un *run* de séquençage haut débit. En effet, l'importante divergence des séquences virales et les lacunes des banques de données, font de l'annotation un réel challenge. Un très grand nombre d'études ont été menées dans ce but. Les logiciels disponibles en ligne METAVIR (98) et VIROME (99) ont par exemple été développés pour l'annotation des lectures virales. SPHINX permet quant à lui l'annotation de tous les taxons par une méthode hybride de reconnaissance et d'alignement combinée (100); MetalID est un outil permettant l'annotation et l'estimation des taxons dans tous les environnements (101).

- L'assemblage des lectures

L'assemblage des données générées permet d'obtenir des séquences plus longues indiquant la couverture des gènes, et menant au « contigage » de génomes. La présence de régions conservées rend l'assemblage plus aisé, cependant certaines régions de deux génomes distants génétiquement peuvent être identiques. Des algorithmes ont été développés pour permettre cet assemblage des séquences tout en diminuant les potentielles erreurs de liaisons de séquences. On retrouve parmi ces algorithmes GENOVO réalisant des

assemblages de novo (102), ou encore MetaORFA permettant un assemblage de novo mais possédant un algorithme (ORFome) qui augmente la sensibilité de recherche d'homologie pouvant également améliorer les analyses de diversités des métadonnées (103).

Il existe aussi des logiciels permettant d'estimer la quantité de séquences assignées à une famille virale ou à une autre. GAAS (104) ou encore GRAMMY (105) font partie de ces outils d'estimation des compositions de communautés microbiennes de données de séquençage. Une revue présente l'ensemble des différents outils utilisés pour interpréter des données de séquençage haut débit (106) (Tableau 2).

	GENOMICS	(NON-VIRAL) METAGENOMICS	VIRAL METAGENOMICS
Quality control	<i>Lucy (Li and Chou 2004); Phred (Ewing et al. 1998); Mothur (Schloss et al. 2009); TagDust (Lassmann et al. 2009); PIQA (Martínez-Alcántara et al. 2009); SolexaQA (Cox et al. 2010); PRINSEQ (Schmieder & Edwards 2011)</i>		
Pre-processing	Duplicates removal (artifacts from 454/Roche sequencing)	-	<ul style="list-style-type: none"> * Artificial duplicates may bias species relative abundances and the community structure <i>CD-HIT-454 (Li and Godzik 2006)</i>
	Filtering	-	<ul style="list-style-type: none"> * Contamination from human genomic material is a major issue in human viral metagenomes. * Removal of bacterial-like reads must be carefully evaluated (part of these might come from genes of bacterial origin transferred to their phages or from excised prophages mistakenly annotated as bacteria). <i>Deconseq (Schmieder and Edwards 2011)</i>
Annotation	-	<i>BLAST (Altschul et al. 1990); MEGAN (Huson et al. 2007); MG-RAST (Meyer et al. 2008);</i> <i>TETRA (Teeling et al. 2004b);</i> <i>SPHINX (Mohammed et al. 2011);</i> <i>Phymm/PhymmBL (Brady and Salzberg 2011);</i> <i>PhyloPythia (McHardy et al. 2006)</i>	<ul style="list-style-type: none"> * Annotation is challenging in viral metagenomics because of the high divergence of viral sequences and the limited number of reference genomes for similarity searches. * Annotation of bacterial-annotated reads must be carefully evaluated (part of these might come from genes of bacterial origin transferred to their phages or from excised prophages mistakenly annotated as bacteria).
			<i>ProVide (Ghosh et al. 2011); MGTAXA (http://mgtaxa.jcvi.org);</i> <i>MetaVir (Roux et al. 2011); VIROME (Bhavsar et al. in preparation);</i> <i>VMGAP (Lorenzi et al. 2011)</i>
Assembly	<i>Euler (Pevzner et al. 2001)</i> <i>SOAPdenovo; (Li et al. 2009);</i> <i>AbySS (Simpson et al. 2009);</i> <i>ALLPATHS (Butler et al. 2008);</i> <i>Newbler (Roche);</i> <i>CLC (CLC bio, Aarhus, Denmark)</i> <i>Velvet (Zerbino et al. 2008);</i> <i>CAP3 (Huang et al. 1999)</i>	<ul style="list-style-type: none"> * Only partial assembly can be made, due to the complexity of a metagenomic sample and the limited depth of sequencing. We may reconstruct partial or complete genomes of dominant species in case of low community diversity. * The presence of conserved genomic regions between distantly related species may lead to the generation of chimeric contigs. * Assembly is complicated by the different coverage across species due to uneven species frequency in the sample. <i>Genovo (Laserson et al. 2011); MetaORFA (Ye and Tang 2009); MetaIDB (Peng et al. 2011); MetaVelvet (Namiki et al. 2012); Bambus 2 (Koren et al. 2011); MAP (Lai et al. 2012)</i>	
Estimation of genotype abundances, community structure and diversity	-	<i>GAAS (Angly et al. 2009); GRAMMY (Xia et al. 2011); GASiC (Lindner and Renard, 2012); Circonspect (Angly et al. 2006); PHACCS (Angly et al. 2005)</i>	

Tableau 2 : revue des principaux outils pour l'analyse génomique, métagénomique et de métagénomique virale. Chaque étape de l'analyse des métadonnées est passée en revue.

Fancello, L. et al., Computational tools for viral metagenomics and their application in clinical research. Virology (2012), <http://dx.doi.org/10.1016/j.virol.2012.09.025>

4. Applications du séquençage à haut débit

Les applications du séquençage à très haut débit sont diverses et variées. Dans la mesure où l'on peut obtenir du matériel génétique (ADN, ARN), il est possible d'employer le séquençage à haut débit. Quelle que soit la chimie utilisée, ce type de séquençage a pour objectif de produire une quantité très importante de données pour un voire plusieurs échantillons en parallèle (multiplexage possible des échantillons). Les opportunités d'application qu'offre le séquençage à haut débit peuvent se diviser en grandes catégories (Figure 35):

- Le séquençage *de novo*

On appelle séquençage *de novo* l'acquisition de séquences génétiques d'un nouvel organisme pour lequel il n'existe pas encore de référence. Le séquençage *de novo* peut aussi servir dans des analyses de métagénomique. La métagénomique est une méthode qui consiste à étudier l'ensemble des acteurs d'un environnement complexe par séquençage des acides nucléiques. Cette méthode permet d'établir la liste des espèces présentes (qu'elles soient bactériennes, virales, ou même d'origine eucaryote). Elle peut être réalisée sur des échantillons divers comme les différentes flores, tissus, matière organique ou non (fèces, terre, végétaux..). Il est possible de cibler une étude métagénomique, ainsi on parlera d'analyse d'un virome, qui est une méthode permettant d'apprécier l'ensemble des virus d'un échantillon donné.

- Le re-séquençage

Comme son nom l'indique le re-séquençage permet de vérifier les séquences ayant été obtenues par le passé. Cette méthode va permettre de mettre en évidence l'apparition de mutations, de dérives génétiques par rapport aux séquences obtenues auparavant. Ainsi, il est par exemple possible de faire ressortir les liens entre l'apparition d'une résistance et d'une mutation.

- L'analyse fonctionnelle

L'analyse fonctionnelle demande une analyse plus fine et une quantité de séquences plus importante. En effet, elle permet l'étude de l'interactome (étude des interactions entre deux acteurs, exemple : ADN/protéine), l'analyse de méthylations, l'étude de la réponse immunitaire d'un organisme vis-à-vis d'une agression virale, bactérienne, etc., mais aussi l'étude de facteurs cellulaires ou de maladies génétiques.

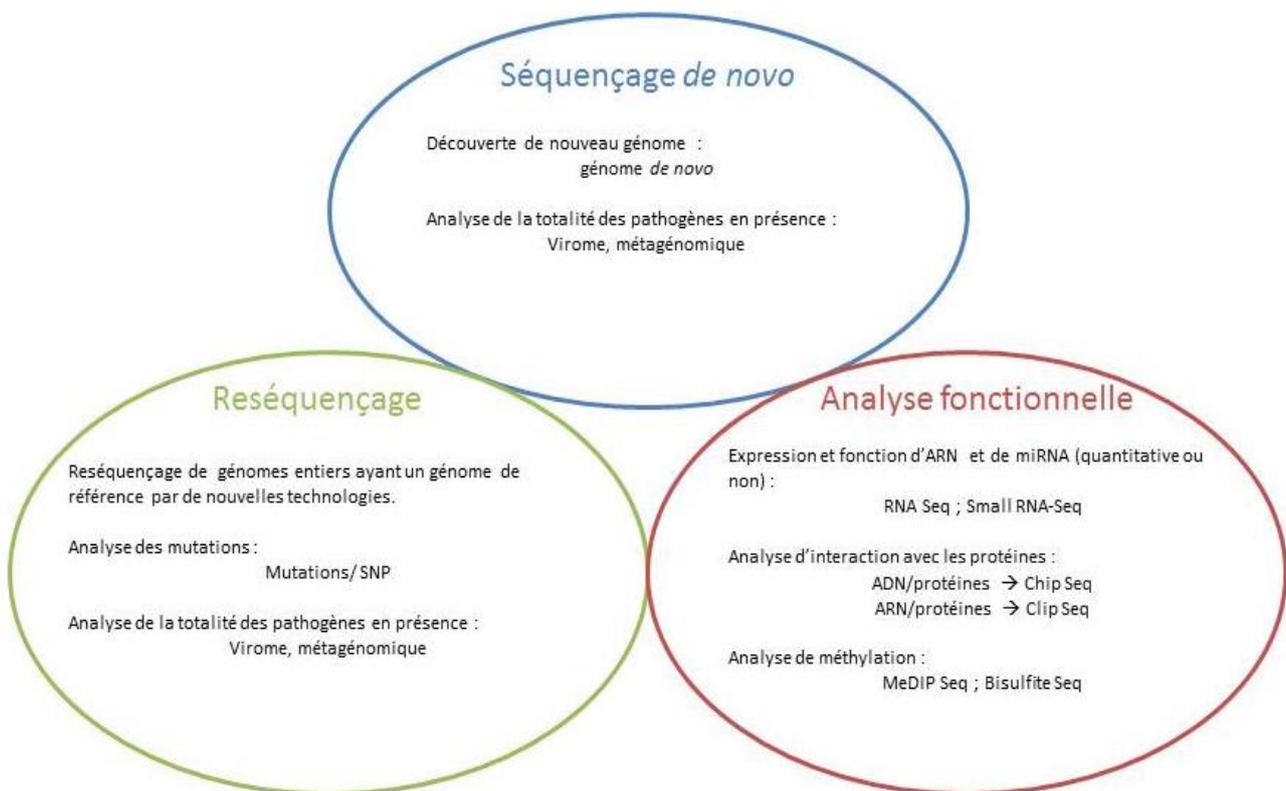


Figure 35 : représentation du regroupement possible des applications du séquençage à haut débit.

Le séquençage à haut débit est utilisé dans de nombreuses applications. Le choix de la plateforme dépend de la question à laquelle l'utilisateur souhaite répondre. De plus, il est important de noter que ces différentes technologies ne sont pas apparues la même année, et que chaque machine à un coût élevé. Il est donc logique que les technologies les plus « anciennes » regroupent le plus grand nombre d'utilisations publiées. De plus en plus de chercheurs s'appuient sur ces outils de nouvelle génération ouvrant de nouvelles opportunités. La très grande quantité de séquences générées rend ces techniques très fiables et extrêmement précises. Le séquençage à haut débit a significativement contribué à l'avancée dans la compréhension mais également à la découverte de pathogènes. C'est aussi un outil très important et très puissant en immunologie venant compléter les outils déjà existants comme les puces à ADN (107).

De nombreux outils et plateformes d'analyse ont suivi l'explosion du séquençage dernière génération et se sont développés autour de celui-ci. Parmi la grande quantité d'applications pour lesquels le séquençage à très haut débit a pu servir, voici quelques exemples.

(1) Séquençage de génomes complets

La puissance des séquenceurs haut débit a permis l'enrichissement des banques de données. En effet, la taille et la quantité importante des séquences générées par ce nouvel outil rendent la couverture d'un génome complet plus aisée. Le séquençage haut débit n'est pas seulement utilisé pour la caractérisation de pathogènes. Ainsi des génomes complexes ont été séquencés comme celui de la pomme (*Malus domestica*) (108) ou encore du cacaoyer (*Theobroma cacao*) (109) publiés en 2010 et 2011 respectivement. Ce genre de données permet d'accélérer l'étude des gènes impliqués dans des résistances aux maladies, mais aussi dans le développement, l'arôme et même pour les croisements génétiques. De même le génome de populations indigènes (d'Afrique du sud) (110) a été mis en ligne permettant la comparaison et la mise en évidence de variation génomique en fonction de l'habitat et l'hygiène de vie. Parmi les nombreux génomes microbiens qui ont été ajoutés aux banques de données, on peut citer *Bacillus anthracis* (111), *Chlamydomophila psittaci* (112) ou encore *Salmonella enterica* (113). De même, de nombreux génomes complets de virus ont été publiés comme par exemple un virus influenza A (H7N9) retrouvé chez un patient en Chine. Le séquençage haut débit a permis le séquençage de ce génome sans passer par une étape de culture du virus éliminant des étapes pouvant introduire des biais (114). Parmi les génomes viraux issus de cette nouvelle technologie, on peut également citer le premier génome de paramyxovirus aviaire de type 4 du continent africain publié en 2012 (115). Le séquençage de génome complet permet non seulement d'enrichir nos connaissances et nos banques de données mais aussi de mettre en évidence les mutations et l'apparition de résistance aux traitements comme pour le virus de l'hépatite B (116).

(2) Métagénomique

Une autre application très importante du séquençage à haut débit est l'utilisation sans *a priori* afin de récolter un maximum d'information dans les échantillons. La puissance des séquenceurs de dernière génération a permis d'estimer les populations virales (virome) et bactériennes (bactériome) d'un grand nombre d'organismes, flores voire d'environnements. Tant dans le monde vétérinaire qu'en médecine humaine, sur des patients sains ou des patients présentant des signes cliniques, les analyses métagénomiques permettent de répondre à des questions très variées.

Targeted disease	Nucleic acid	Samples	New virus discovered	Sequencing method	Viral particles isolation	Assembly	Annotation	Reference
Lower respiratory tract infection	DNA RNA	Nasopharyngeal aspirates	Parvovirus, coronavirus	Sanger	Ultracentrifugation; 0.22 µm filtering	Not performed	BLAST	(Allander et al., 2005)
Human merkel cell carcinoma	RNA	Cell carcinoma tissues (biopsies)	Polyomavirus	454/Roche	(Direct nucleic acids extraction)	Not performed	BLAST	(Feng et al., 2008)
Diarrhea	RNA	Stool	Astrovirus, torque teno virus, norovirus, picobirnavirus, enterovirus, nodavirus	Sanger	Centrifugation; 0.45 µm filtering	Not performed	BLAST	(Finkbeiner et al., 2008)
Acute respiratory infections and diarrhea	RNA	Nasopharyngeal aspirates, stool	-	454/Roche	Centrifugation	Not performed	BLAST, SSEARCH	(Nakamura et al., 2009)
Fatal transplant-associated disease	RNA	Brain, cerebrospinal fluid, serum, kidney, liver	Arenavirus	454/Roche	(Direct nucleic acids extraction)	CAP3 (Huang and Madan, 1999)	BLAST	(Palacios et al., 2008)
Hemorrhagic fever	RNA	Liver biopsies, serum	Arenavirus	454/Roche	(direct Nucleic acids extraction)	GCG Package (Accelrys, San Diego, CA, USA)	CLC RNA Workbench (CLC bio, Århus, Denmark)	(Briese et al., 2009)
Acute flaccid paralysis	DNA	Stool	Bocavirus, picornaviruses, circovirus, nodavirus, dicistroviruses	454/Roche, Sanger	Centrifugation; 0.45µm filtering	Sequencher (Gene Codes Corporation, Ann Arbor, MI USA)	BLAST	(Victoria et al., 2009)
Cystic fibrosis	DNA	Sputum	-	454/Roche	0.45 µm filtering; CsCl gradient	PHRAP (www.phrap.org)	BLAST, MG-RAST	(Willner et al., 2009)
Upper respiratory tract infection	RNA	Nasopharyngeal aspirates	-	Illumina	(Direct nucleic acids extraction)	Geneious (http://www.geneious.com)	BLAST	(Greinger et al., 2010)
Encephalitis	RNA	Frontal cortex (biopsy)	Astrovirus	454/Roche	(Direct nucleic acids extraction)	GreenPortal website (http://tako.cpmc.columbia.edu/Tools)	BLAST	(Quan, 2010)
Chronic fatigue syndrome	DNA RNA	Serum	-	454/Roche, Sanger	0.22 µm/0.45 µm filtering; ultracentrifugation	miraEST (Chevreux et al., 2004)	BLAST	(Sullivan et al., 2011)
Acute exacerbation of idiopathic pulmonary fibrosis	RNA	Bronchoalveolar lavage and serum	-	Illumina	(Direct nucleic acids extraction)	Not performed	MegaBLAST, BLAST	(Wootton et al., 2011)
Lower respiratory tract infections	DNA & RNA	Nasopharyngeal aspirates	Rhinovirus C	454/Roche	0.22 µm/0.45 µm filtering; ultracentrifugation	miraEST (Chevreux et al., 2004)	MegaBLAST, BLAST	(Lysholm et al., 2012)
Hemorrhagic fever	RNA	Serum	-	454/Roche	(Direct nucleic acids extraction)	Newbler (Roche); CLC (CLC bio, Aarhus, Denmark)	BLAST, MEGAN	(McMullan et al., 2012)
Cystic fibrosis	DNA	Lung tissue (biopsies)	-	454/Roche	0.45 µm filtering; CsCl gradient	CAP3 (Huang and Madan, 1999)	BLAST	(Willner et al., 2011)
Tropical febrile illness	DNA RNA	Serum	Circovirus	Illumina	(Direct nucleic acids extraction)	Not performed	BLAST	(Yozwiak et al., 2012)

Figure 36 : Etude de métagénomique virale sur des échantillons humains pour une application d'étude de cas clinique. De gauche à droite : la maladie ciblée, type d'acide nucléique (ARN/ADN), type d'échantillon, éventuel virus découvert, technique utilisée, isolement des particules virales, outils bioinformatique utilisés (assemblage, annotation), référence de l'étude.

Fancello, L. et al., *Computational tools for viral metagenomics and their application in clinical research. Virology* (2012), <http://dx.doi.org/10.1016/j.virol.2012.09.025>

Des études de métagénomiques ont été réalisées sur des environnements variés comme les lacs français du Bourget et Pavin dont les diversités virales ont été illustrées en 2012 (parmi les 156.772 séquences reconnues pour le lac du Bourget et les 92.834 séquences pour le lac Pavin, 95.905 séquences et 47.345 séquences ont été associées à des séquences virales pour les lacs du Bourget et Pavin respectivement) (117) ; des grands fleuves comme le Mississippi dans lequel des séquences de virus de la peste porcine africaine ont été retrouvées (environ 6,9% des séquences générées ont été associées à des séquences virales) (118) ; ou encore des océans tel le Pacifique illustrant les variations des populations virales en fonction de l'ensoleillement (existence d'un « core » constant et d'une partie plus « flexible » de la population virale de cet océan) (119,120). Autant d'études qui permettent de connaître et mieux comprendre l'environnement dans lequel l'homme et tout être-vivant évoluent. Ces études pourraient permettre également de prédire et mieux se prémunir contre d'éventuels sauts d'espèce ou l'apparition de mutations à l'origine de problèmes de santé publique importants.

Un grand nombre d'études cliniques se sont appuyées sur la métagénomique en médecine humaine comme le montre la figure 35 (121). La première de ces analyses métagénomiques fut réalisée en 2008 et a permis la découverte d'un nouvel arenavirus lors de la recherche de la cause de décès de patients transplantés (122). Puis de nombreuses études ont suivi, utilisant elles aussi le séquençage haut débit pour caractériser l'agent responsable d'une pathologie comme chez des enfants touchés par une paralysie flasque (123), ou touchés par des diarrhées aiguës (124). Des études ont aussi visé à caractériser l'ensemble des pathogènes viraux présents lors de troubles respiratoires. Ainsi une équipe a mis en évidence la présence d'un virus H1N1 non seulement par séquençage à haut débit mais également grâce à une puce ADN, la Virochip chez 17 patients aux États-Unis (125). Une équipe suédoise a réalisé une étude plus descriptive, mettant en évidence la présence entre autre de *Paramyxoviridae*, *Orthomyxoviridae* et de *Picornaviridae* dans un pool d'échantillon clinique (126). Récemment une étude a été menée sur le microbiome humain démontrant non seulement leur importante présence à travers notre organisme mais aussi leur rôle (127).

Les recherches métagénomiques ne se cantonnent pas à la médecine humaine : la flore virale de nombreux animaux a déjà été caractérisée. Les viromes de chats (128,129) et chiens atteints de diarrhées ont notamment été caractérisés (130). Des espèces plus exotiques ont

également été étudiées. Les chauves-souris font l'objet de nombreuses études tout particulièrement depuis que les chiroptères ont été identifiés comme hôte réservoir des coronavirus responsables des syndromes respiratoires aigus (MERS-CoV, SARS-CoV (131,132)). Des analyses métagénomiques des chiroptères ont permis de montrer le pouvoir de dissémination et le potentiel zoonotique à l'échelle planétaire de ces mammifères (133–135). D'autres mammifères proches de l'homme tant par leurs proximités géographiques que génétiques ont été sujets d'études métagénomiques. En effet, les porcs sont suspectés d'être des hôtes intermédiaires permettant l'adaptation de certains virus (tels que l'influenza A) à l'homme (136–138). Tout comme pour les chiroptères, il est important de mieux comprendre l'environnement et la flore de ces animaux potentiellement porteurs de pathogènes ayant un fort pouvoir zoonotique, tant d'un point de vue santé humaine qu'animale (139–141). Ainsi des études ont cherché à connaître le virome de ces animaux urbains susceptibles d'être porteurs de pathogènes viraux comme celui des rats et des pigeons qui sont les espèces animales les plus représentées dans les milieux urbains (142,143).

(3) Identification et découverte de pathogènes

Cette technologie a permis l'identification de nombreuses bactéries comme par exemple la caractérisation de 35 génomes de salmonelles issues d'un épisode majeur de salmonellose aux Etats-Unis en 2009-2010 (113). De plus ces nouvelles techniques de séquençage à haut débit permettent de dépasser les limites de techniques d'identification classique telle que l'isolement, ainsi une équipe danoise a réussi à séquencer des bactéries rares et impossibles à cultiver (144). L'utilisation du séquençage à haut débit a également permis d'identifier un grand nombre de pathogènes viraux comme par exemple deux nouveaux arenavirus en 2008 et 2009, l'un responsable du décès de patients transplantés et l'autre étant associé à une fièvre hémorragique en Afrique du sud (122,145). Le NGS a permis d'identifier de nouveaux arbovirus de la famille des Rhabdoviridae (ayant un spectre d'infection très large et un pouvoir pathogène important, exemple : virus de la rage) chez des moustiques en Afrique de l'ouest (146) ; un nouveau picornavirus, potentiel agent étiologique d'une hépatite virale chez la dinde (147) ; l'identification et la caractérisation d'un nouveau virus de l'encéphalite murine de Theiler (148) ; ou le virus de l'hépatite C du chien, qui est le virus le plus proche phylogénétiquement du virus de l'hépatite C humaine (149). Il permet aussi l'identification de

virus émergent comme le virus Ebola associé à l'apparition des fièvres hémorragiques de 2008 en Ouganda (150). Mais aussi plus récemment un nouvel Orthobunyavirus, le virus Schmallerberg (portant le nom de la ville dans laquelle il a été découvert) qui a été identifié et associé à la maladie des ruminants émergente de 2011 au nord de l'Allemagne (151). Le séquençage à haut débit a eu également une importance dans la découverte de nouveaux virus chez des chauves-souris, reconnus comme hôtes-réservoirs de nombreux virus capables de franchir la barrière d'espèces (comme le SARS-CoV, le virus Ebola, le virus Nipah) (134,152–154). Ces nouvelles technologies permettent aussi de lever le voile sur des pathologies dont l'agent était jusque-là inconnu, comme pour la maladie foudroyante de la pintade qui est une entérite fulminante (155), le Flavivirus responsable de la maladie de Theiler chez les équins (156), ou un nouveau virus influenza porcine qui est le premier virus influenza de groupe D (95).

IV. Maladies virales aviaires

Les maladies aviaires, qu'elles soient respiratoires ou digestives, sont souvent à l'origine de pertes économiques importantes. Ce déficit peut être lié à la mort des animaux, à une surconsommation d'aliment, ou un défaut de maîtrise de l'environnement du bâtiment (157). Nous nous contenterons ici de présenter quelques notions introduisant nos travaux.

A. Principales Maladies respiratoires aviaires

Les maladies respiratoires ont un impact économique très important dans les élevages avicoles. Ces maladies qui sont souvent enzootiques, et dues à la concentration d'animaux dans ces élevages, touchent inévitablement l'ensemble du troupeau.

L'apparition de syndrome respiratoire est induite par l'association d'un ou plusieurs facteurs. On parle alors de « maladie respiratoire complexe » (MRC) qui est définie par l'association d'un ou plusieurs facteurs d'induction, d'un ou plusieurs facteurs environnementaux, et d'un ou plusieurs agents de surinfection, aboutissant à des lésions chez l'animal.

Les agents pathogènes primaires peuvent être des mycoplasmes, des virus ou l'association des deux. Les agents de surinfection (une ou plusieurs bactéries issues de la flore de l'oiseau ou de son environnement) entraînent une aggravation de ces lésions.

Les volailles peuvent être soumises à différents facteurs de « stress » ayant des origines diverses. Parmi ces stress, on citera les vaccinations, la température, l'hygrométrie, la ventilation, les erreurs d'élevage comme une densité d'oiseaux trop élevée, des défauts d'aliments ou de litière. La litière joue un rôle important dans un élevage de volaille car elle permet de réduire le dégagement d'ammoniac. Ce gaz est produit par la fermentation des fientes et induit une sensibilité accrue des animaux aux pathogènes respiratoires.

Chez les volailles pondeuses, les troubles respiratoires se manifestent surtout par une atteinte de l'appareil génital, du fait des spécificités anatomiques des oiseaux : en effet, les sacs aériens sont en contact étroit avec la grappe ovarienne et l'oviducte, ce qui peut expliquer que les virus à tropisme respiratoire ont également un tropisme génital (157).

Il existe un grand nombre d'agents pathogènes respiratoires dont les plus communs sont :

- L'herpesvirus responsables de la Laryngotrachéite infectieuse,
- Les virus influenza A faiblement pathogènes (les virus hautement pathogènes sont à l'origine d'une maladie systémique, très rapidement mortelle).
- Les paramyxovirus de type 1 (PMV-1), qui sont associés à la maladie de Newcastle
- Les métapneumovirus aviaires (aMPV), à l'origine d'une atteinte des voies respiratoires supérieures, qui peut être d'évolution bénigne ou sévère selon les complications bactériennes.
- Les coronavirus responsables de la bronchite infectieuse (virus de la bronchite infectieuse, IBV)
- Les poxvirus responsables de la variole aviaire ; dont une forme respiratoire est décrite
- Les mycoplasmes de la poule : *Mycoplasma gallisepticum* et *Mycoplasma synoviae*.
- *Escherichia coli* pathogènes (Avian Pathogenic E. coli), responsables de colibacillose respiratoire, qui se manifestent essentiellement par des aérosacculites fibrineuses.
- *Aspergillus (fulmigatus notamment)* responsables des aspergilloses chez les jeunes volailles
- *Ornithobacterium rhinotracheale* associés à des infections respiratoires très sévères chez la dinde.

Une infection peut se traduire ou non par des signes cliniques. On parle d'infection inapparente quand elle est en phase d'incubation ou lors d'un portage sain. Un oiseau porteur sain est un animal qui est infecté par un micro-organisme mais qui ne développe pas de maladie ou de signe clinique. Les signes respiratoires associés à ces agents sont divers : toux ; éternuements ; jetage (écoulement par les orifices nasaux) ; larmolements ; râles (bruit anormal observé à l'examen). Ces signes peuvent entraîner une diminution des performances autant dans le domaine de la croissance, que la ponte ou le gavage.

Une chute de ponte est un terme employé pour évoquer la diminution de la production d'un lot de volailles pondeuses. Elle est traduite par une variation de la courbe de ponte. Cette courbe correspond au nombre d'œufs pondus par jour et pour 100 oiseaux en fonction

du temps. Cette variation peut prendre plusieurs aspects suivant l'origine du facteur influent sur la productivité. Il existe un grand nombre de facteurs différents pouvant influencer sur la ponte, touchant ou non l'appareil reproducteur de l'animal. L'allure de la courbe de ponte sera fortement impactée par les maladies infectieuses ou les accidents techniques, alors que lors de problèmes d'origine toxique ou nutritionnelle la courbe oscillera de façon chronique.

Les principales causes d'une chute de ponte sont regroupées dans quatre catégories :

- les causes liées à l'alimentation ou aux troubles métaboliques

Une sous-alimentation en eau ou en aliment peut induire une réduction de la production d'œufs. La sur-alimentation (ou alimentation trop riche) peut elle aussi entraîner une modification de la ponte, par exemple par le manque d'exercice. La chute de ponte due à un problème métabolique peut être également induit par une mycotoxine (exemple : zéaralénone).

- les causes liées à la zootechnie

Les stress induits par les changements de luminosité, de température, d'air ambiant influent directement sur la ponte. La couvaie peut elle aussi entraîner un arrêt plus ou moins long de la ponte.

- les causes infectieuses

Les infections virales pouvant entraîner une chute de ponte ou des malformations de l'œuf, sont nombreuses. Les affections à adénovirus provoquent un syndrome appelé *Egg Drop Syndrom* (EDS 76), elles induisent des lésions au niveau de l'appareil génital et se traduisent par des œufs décolorés ou mous. La bronchite infectieuse, causée par une infection à coronavirus (IBV), se manifeste par des signes respiratoires et une déformation des œufs. D'autres infections par des virus respiratoires peuvent induire des chutes de pontes. Ainsi la maladie de Newcastle, l'influenza aviaire ou les métapneumoviroses provoquent une chute de ponte chez les oiseaux reproducteurs. On retrouve également des infections bactériennes comme les colibacilloses et mycoplasmoses.

- et les causes liées aux infections parasitaires

Le stress induit par une infection par les poux rouges peut induire une chute de ponte.

B. Principaux virus entériques aviaires

Tout comme les maladies respiratoires, les maladies entériques ont un impact économique important sur les élevages aviaires. Les agents étiologiques potentiellement responsables des entérites aviaires sont multiples. En effet, les causes peuvent être bactériennes, parasitaires, virales ou encore d'une toute autre origine, comme les intoxications ou le stress. La multitude de facteurs rendent le diagnostic plus complexe, d'autant plus qu'une combinaison de plusieurs facteurs peut être observée. Seules les principales causes d'entérites virales seront présentées dans cette thèse. Cependant, on peut souligner que l'importance accrue des entérites aviaires provient autant des agents pathogènes ayant co-évolué avec leur hôte que des pratiques alimentaires. Ces habitudes ont probablement déséquilibré la flore digestive des volailles.

Les virus retrouvés majoritairement lors de maladie entérique aviaire sont les adénovirus, les astrovirus, les coronavirus, les entérovirus-like, les réovirus et les rotavirus. Certains virus, comme le virus de la maladie de Newcastle peuvent également induire des signes entériques, notamment dans leurs formes aiguës. Toutefois, il est essentiel de souligner que certains virus (ainsi que des bactéries entéropathogènes telles que les enterobactéries) peuvent être portés de façon asymptomatique. Les méthodes de diagnostic sont principalement basées sur la PCR et RT-PCR ; l'histopathologie peut également être réalisée en amont de ces tests de biologie moléculaire afin d'orienter le diagnostic. Pour cela, il est nécessaire de connaître le site de réplication de la cible. Ces virus ont une transmission horizontale dans la plupart des cas. Cependant certains virus entériques comme les rotavirus ou les astrovirus sont susceptibles d'être transmis de façon verticale également (157).

1. Les entérites à virus ADN

Les adénoviroses aviaires sont de plus en plus présentes dans les élevages. Les adénovirus font partie de la famille des *Adenoviridae*, ont une taille de 74 à 80 nm et peuvent induire des maladies diverses. Cette famille est subdivisée en cinq genres : *Mastadenovirus*, *Aviadenovirus*, *Atadenovirus*, *Siadenovirus* et *Ichtadenovirus* (158) (<http://www.vmri.hu/~harrach/AdVtaxlong.htm>). Les genres *Aviadenovirus*, *Siadenovirus* et

Atadenovirus peuvent tous trois induire une pathologie chez les espèces aviaires. Les *Siadenovirus*, anciennement appelés adénovirus du groupe 2, ont été identifiés comme responsables de l'entérite hémorragique du dindon, la maladie de la rate marbrée du faisan ou encore de la splénomégalie du poulet.

Cette entérite hémorragique de la dinde peut se manifester de plusieurs façons, de manière sub-clinique, traduite par une séroconversion ou de faibles symptômes. Elle peut toutefois se manifester de façon plus grave pouvant alors induire la mort de l'oiseau. La nature de la souche virale en cause, ainsi qu'une surinfection peuvent expliquer la différence de symptômes. Les signes cliniques sont traduits par un arrêt de l'alimentation et la présence de sang dans les fientes. L'examen nécroscopique révèle une congestion de l'intestin grêle, une muqueuse intestinale hémorragique, une rate et un foie hypertrophiée et une rate d'apparence marbrée. Ces virus peuvent également toucher les pintades et induisent un syndrome hémorragique.

Le diagnostic de cette maladie est basé sur les symptômes et les lésions observées lors de l'autopsie. Il existe un vaccin et, lors d'une affection, un traitement est préconisé pour contrôler les surinfections bactériennes (159).

Les entérovirus-like sont des virus de la famille des *Picornaviridae*. Cette famille est composée de 26 genres dont les *Entérovirus* (<http://www.picornastudygroup.com/taxa/genera.htm>). Il s'agit de virus ADN d'une taille allant de 22 à 30 nm. Ces entérovirus induisent une diarrhée et un retard de la croissance. Ils touchent de nombreuses espèces aviaires tel que les poulets, les pintades, les dindes, les perdrix ou encore les autruches ou les faisans. Les *Picornaviridae* sont également responsables d'une maladie infectieuse à tropisme nerveux, l'encéphalomyélite aviaire. Cette maladie est traduite par des troubles nerveux et un taux de mortalité variant entre 10 et 80%.

2. Les entérites à virus ARN

Les virus ARN sont connus pour avoir la faculté de muter plus facilement que les virus ADN. Les mutations apparaissent lors de la réplication du virus, en effet les virus ARN possèdent une polymérase n'ayant pas de système de correction de copie comme les virus ADN. Ce phénomène leur garantit un potentiel évolutif plus rapide faisant apparaître de nouveaux variants (mutants). On retrouve ainsi des virus aviaires ARN tel que le virus Influenza ou encore les coronavirus aviaires tous deux en constante évolution. Concernant

les virus entériques, on retrouve des astrovirus aviaires, des rotavirus, des reovirus aviaires mais également des coronavirus (le coronavirus de la dinde ou « Turkey coronavirus ») dont une introduction plus approfondie sera faite dans la suite de ce document.

Les astrovirus aviaires (« Turkey Astrovirus » de type 1 (TAsTV1) et, principalement, les TAsTV2) sont de petit virus, d'une taille allant de 25 à 35 nm, non enveloppés, de la famille des *Astroviridae*. Ils se répliquent dans l'épithélium des villosités intestinales et la bourse de Fabricius. Les astrovirus ont été identifiés dans un grand nombre d'espèces différentes, incluant l'Homme, les moutons, les chats, les chiens, les cerfs, les poulets, les dindes et les canards. Ces virus ont été retrouvés dans des élevages avicoles présentant des mortalités liées à des entérites (« poult enteritis mortality syndrome » PEMS) (160). Lors d'une infection à astrovirus, le système immunitaire de l'oiseau est très fortement impacté, permettant alors une sensibilité plus importante aux différents agents infectieux (157).

Les rotavirus sont des virus de plus grande taille 60-80nm, ils appartiennent à la famille *Reoviridae*. Les rotavirus ont un tropisme marqué pour le système digestif et sont retrouvés de façon ubiquitaire à travers la planète (161).

Les reovirus aviaires font eux aussi partie de la famille des *Reoviridae* et sont membres du genre *Orthoreovirus*. Ce virus est omniprésent parmi les élevages de volailles, il a été associé à différentes, maladies, notamment entériques (162).

La maladie de Newcastle est une des maladies virales les plus sévères affectant les espèces aviaires. Elle apparaît dans des zones endémiques de manière sporadique depuis de nombreuses années. Le virus de la maladie de Newcastle est un virus de la famille des *Paramyxoviridae*. Ce virus enveloppé, d'environ 200 à 300 nm de diamètre, possède un tropisme variant suivant la nature de la souche (163). En effet, il existe des souches vélogènes viscérotropes provoquant des lésions hémorragiques du tube digestif ; vélogènes neurotropes entraînant des symptômes respiratoire et nerveux ; des souches mésogènes induisant des symptômes respiratoires ; des souches lentogènes ou encore avirulentes n'induisant ni symptômes ni lésions (pouvant être très faible dans le cas des souches lentogènes). Les signes cliniques varient suivant les espèces touchées. Ainsi les dindes et les poulets vont présenter des signes respiratoires sévères, des diarrhées ainsi que des signes nerveux alors que les palmipèdes sont quant à eux considérés comme résistants aux souches virulentes pour les galliformes mais peuvent héberger de manière asymptomatique le virus (157).

Résultats

I. L'agent de la maladie foudroyante de la pintade : un nouveau coronavirus

A. Introduction

La maladie foudroyante de la pintade, aussi appelée « maladie X » par les professionnels, est une entérite aiguë, entraînant un très fort taux de létalité. Cette maladie se caractérise par une intense prostration du lot d'oiseaux, des lésions entériques, et une dégénérescence du pancréas prenant alors une couleur blanchâtre (pancréas dit « neigeux »).

Cette maladie a été observée pour la première fois dans les années 80 dans les élevages du sud-ouest et sud-est de la France, la première description clinique publiée datant de 1986. Sur la base des premières études, l'origine virale fut très rapidement favorisée (165). Plusieurs travaux ont été conduits dans le but d'identifier l'agent étiologique de cette maladie. Cependant, ces investigations étant basées sur les méthodes de virologie classique, seules disponibles à cette époque, et notamment sur la microscopie électronique, n'ont pu conduire à l'identification du virus. De nombreux candidats ont été proposés : toga-like virus (166), reovirus, herpesvirus (167). Plus récemment, un astrovirus a été associé à l'entérite catarrhale chez la pintade (168). Toutefois ce virus n'a pas été détecté dans les lots d'oiseaux affectés par la maladie foudroyante. Suite à l'échec de l'identification par des méthodes classiques de virologie cellulaire et moléculaire, une approche sans *a priori* a été privilégiée. Nous avons employé le séquençage à très haut débit pour cette étude. Les données obtenues ont permis de mettre en évidence un candidat (coronavirus), appartenant à une famille virale déjà connue pour induire des entérites. Les coronavirus font partie de l'ordre des Nidovirales et de la famille des *Coronaviridae*, subdivisée en deux sous-familles : les torovirus et les coronavirus (http://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/vetebrate-official/1230.aspx Consulté le 11/08/2014).

Les coronavirus possèdent un génome à ARN positif non segmenté et enveloppé présentant un spectre d'hôte très large. Ils sont classés en quatre groupes (1, 2, 3 et 4), classification basée à l'origine sur leurs réactivités antigéniques (169), mais qui fut rapidement changée au profit d'une comparaison de l'organisation du génome complet des coronavirus (170). Récemment, une nouvelle nomenclature a été adoptée, répartissant la sous-famille des

Coronavirinae en quatre genres : *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* et *Deltacoronavirus* (171). Il est donc possible de différencier les différents membres de la sous-famille des *Coronavirinae* sur la base de leur organisation génomique (Figure 37) (172,173). Les *Alphacoronavirus* et *Betacoronavirus* regroupent des virus ayant pour hôtes des mammifères ; les virus ayant essentiellement des espèces hôtes aviaires sont, quant à eux, regroupés dans le genre *Gammacoronavirus* mais le coronavirus du béluga est également un *Gammacoronavirus* et les *Deltacoronavirus* comprennent aussi des coronavirus aviaires (Figure 38).

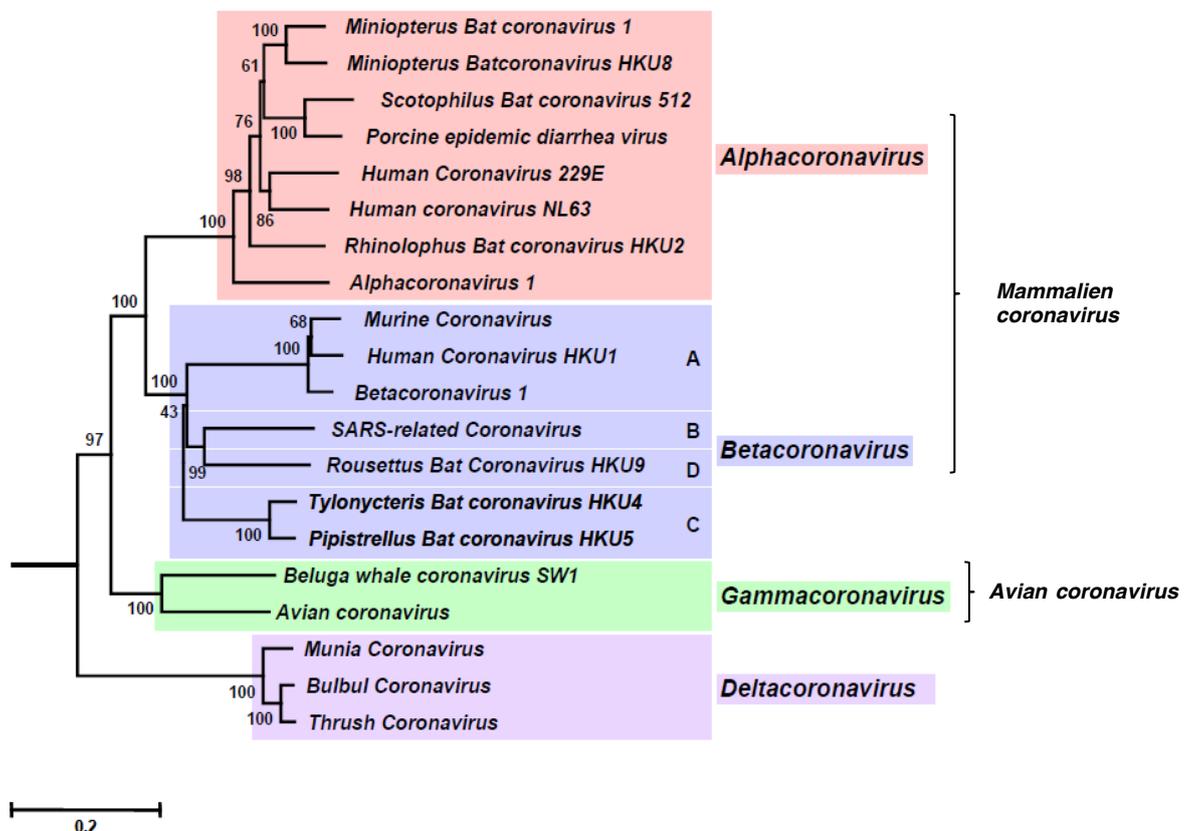
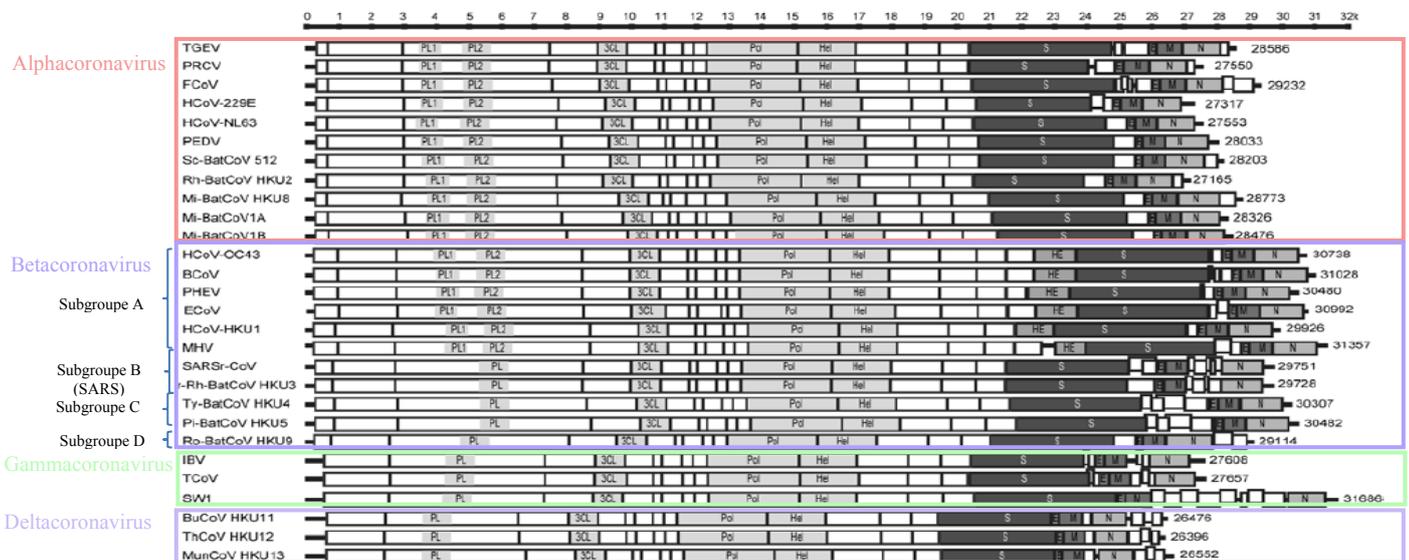


Figure 37 : Arbre phylogénique d'un alignement multiple du gène RdRp et du gène codant pour l'hélicase, avec un torovirus équin (equine torovirus Berne) comme *outgroup*.

http://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/vertebrate-official/1230.aspx Consulté le 11/08/2014

Affectant les mammifères comme les espèces aviaires, ces virus sont principalement associés à des troubles respiratoires ou digestifs, cependant ils sont parfois également liés à des atteintes du système nerveux, urogénital ou hépatique (174). Les recherches consacrées aux coronavirus ont considérablement augmenté depuis 2003, suite à la l'identification du coronavirus responsable du SRAS (syndrome respiratoire aigu sévère) (175,176). Ils infectent classiquement leur hôte de façon *espèce-spécifique*, induisant une infection aiguë ou persistante. Les voies de transmission de ces virus sont habituellement respiratoires ou fécales-orales.



Graphique 38 : Organisation génomique des différents membres des genres de la famille des Coronaviridae. **PL1**, papaine-like protéase 1; **PL2**, papaine-like protéase 2; **PL**, papaine-like protéase; **3CL**, chymotrypsine-like protéase; **Pol**, RNA-dépendent RNA polymérase; **Hel**, hélicase; **HE**, hémagglutinine estérase; **S**, spike; **E**, enveloppe; **M**, membrane; **N**, nucléocapside.

D'après Woo et al. 2010

Tous les coronavirus possèdent une organisation génomique similaire, avec un gène codant pour la réplicase (correspondant à près des 2/3 du génome complet du virus) porté en 5' et des gènes codants pour les protéines structurales sur la partie 3' du génome. Ces gènes sont ordonnés de la façon suivante : le gène codant pour l'hémagglutinine estérase (présent seulement chez quelques virus du genre *Betacoronavirus*), le gène codant pour la protéine de spicule (S), le gène codant pour la protéine de membrane (M), et enfin le gène codant pour la nucléocapside (N) et pour une protéine interne (I) (Figure 38) (177). Le génome des coronavirus est le plus gros des virus à ARN. En effet, les membres de cette famille

possèdent un génome allant de 27,3 Kb pour le plus petit (souche 229E du coronavirus humain HCoV-229E) à 31,3 Kb (virus de l'hépatite murine MHV) (174).

Récemment, l'émergence d'un nouveau coronavirus responsable d'un syndrome respiratoire aigu chez l'homme, nommé *Middle East Respiratory Syndrome-CoV* (MERS-CoV), a été observée. Ce virus du genre *Betacoronavirus* a été isolé pour la première fois en 2012 et possède un caractère zoonotique. En effet, les analyses phylogénétiques ont montré une étroite relation entre le MERS-CoV et deux virus de chauve-souris, *Tylonycteris* bat coronavirus HKU4 (BtCoV-HKU4) et *Pipistrellus* bat coronavirus HKU5 (BatCoV-HKU5) (178). Des tests sérologiques ont démontré une séroprévalence élevée chez les dromadaires dans les zones touchées (179,180), suggérant un réservoir du genre *Camelus* mais n'écartant pas la théorie du réservoir chiroptères. Tout comme pour le SRAS ayant pour origine un *Betacoronavirus* de chiroptère qui s'est propagé chez la civette (181), le MERS-CoV serait donc un virus présentant un large spectre d'hôte (182).

Afin d'identifier l'agent étiologique de la « maladie X », une analyse sans *a priori* a été conduite suite à l'échec des méthodes de diagnostic classique. Les résultats du séquençage à haut débit ont permis de mettre en évidence un gammacoronavirus proche du Turkey coronavirus (TCoV) : le coronavirus de la pintade (gfCoV). La séquence complète du gène de spicule a été obtenue par séquençage Sanger en complément des données de séquençage haut débit, permettant une comparaison phylogénique. Afin de confirmer l'association du virus à cette entérite, des lots d'oiseaux ont été testés par PCR spécifique permettant de confirmer les liens de causalité. De plus, des cas spontanés (déclaré avant et après le début de cette étude) ont été criblés et ont confirmé la présence de gfCoV renforçant notre hypothèse. Nous avons également vérifié la présence de ce virus dans les différents organes de pintades infectées expérimentalement dans le but d'apprécier sa répartition tissulaire. Un marquage immunohistochimique à l'aide d'un anticorps anti-TCoV a permis de valider l'affinité entérique et la proximité génétique de ce coronavirus avec le turkey coronavirus. Le caractère fulminant de cette entérite reste toutefois atypique par rapport aux autres entérites aviaires à coronavirus puisque le TCoV (entérique également) cause très peu de clinique chez la dinde.

Les données de séquençage à haut débit nous ont permis d'obtenir en première intention une couverture d'environ 70% du génome complet. Par la suite et en s'appuyant sur un

alignement avec un génome TCoV de référence, la séquence complète du génome du gfCoV a été obtenue en complétant avec du séquençage Sanger.

L'analyse phylogénétique basée sur le génome complet de ce coronavirus a clairement démontré que ce gfCoV est très proche des TCoV nord-américains. De plus, l'organisation du génome du gfCoV est identique à celle des souches de TCoV. Une analyse génétique du génome complet a révélé une recombinaison au niveau du gène S, qui semble provenir d'un échange entre un virus de type « IBV » ayant donné la plupart de son génome et un virus - encore inconnu - qui aurait « donné » son gène de spicule. L'hypothèse de recombinaison a déjà été émise lors de l'étude de l'origine du TCoV (183). Les recombinaisons sont des phénomènes d'échange d'une partie de génome ; ce genre d'échange n'est pas rare au sein de cette famille virale. Un grand nombre de recombinaisons ont déjà été identifiées chez les coronavirus (184), comme par exemple pour le SARS-CoV dont l'origine proviendrait de plusieurs recombinaisons (185).

Aucun lien avec un type de production de pintades n'a pu être mis en évidence. Malgré une sur-représentation de cette maladie dans les élevages de plein air, des lots d'oiseaux n'étant jamais sortis des bâtiments d'élevages ont été infectés par ce coronavirus entérique. Les informations épidémiologiques disponibles à ce jour ne nous ont pas permis d'identifier d'espèce aviaire sauvage en contact avec les élevages de pintades : une recherche de virus ciblée sur les principaux oiseaux commensaux (étourneaux, moineaux, etc...) pourrait être entreprise, mais elle serait très aléatoire. L'origine de la dissémination du donneur du gène S reste donc actuellement floue. Outre l'hypothèse d'une origine aviaire sauvage, plusieurs autres pistes peuvent être explorées. La proximité de bâtiments d'élevages aviaires d'autres espèces (poulets, canards, dindes etc..) ou encore l'épandage de matière fécale dans les champs avoisinants les bâtiments d'oiseaux infectés sont des causes potentielles d'introduction de matériel biologique contaminé dans les élevages. La transmission peut provenir soit d'aérosols (circuit de ventilation, fenêtre, brèche dans les murs...) soit directement des fientes (introduction par l'homme lors d'un passage entre différents bâtiments, entrée d'oiseaux commensaux potentiellement vecteurs dans le bâtiment).

D'une manière plus globale, on peut espérer que dans un futur proche, la surveillance des coronavirus et les séquences de coronavirus aviaires disponibles dans les banques de séquences publiques vont considérablement augmenter, ce qui pourrait permettre d'identifier l'espèce hôte correspondant au virus le plus proche du gfCoV.

Dans le prolongement de ce travail d'identification, nous avons initié des essais visant à isoler le virus dans différents systèmes. L'isolement du pathogène est en effet idéal pour obtenir un inoculum de référence stable pour les travaux suivants.

Deux systèmes ont été utilisés pour les travaux d'isolement du gfCoV. Plusieurs tests ont été réalisés sur système cellulaire d'une part et sur œufs embryonnés d'autre part. Nous avons essayé plusieurs lignées cellulaires : les cellules HRT18 (lignée épithéliale d'adénocarcinome de rectum humain permissive aux coronavirus bovin BCoV), les cellules Véro (lignée épithéliale de rein de singe vert permissive aux SARS-CoV et BCoV), les cellules A549 (lignée épithéliale de tumeur pulmonaire humaine permissive aux SARS-CoV), qui ont été utilisées pour l'isolement de coronavirus dans de précédentes études (186–192). Les lignées cellulaires aviaires DF1 (lignée cellulaire de fibroblastes de poulet), QT6 (lignée cellulaire de fibroblastes de caille), CEF (fibroblastes issus d'œufs de poule embryonnés), et DEF (fibroblastes issus d'œufs de canne embryonnés) ont également été testées, mais ces essais ont tous été infructueux. Aucun effet cytopathogénique n'a en effet pu être observé et après deux passages, les cellules et leur surnageant étaient tous négatifs pour le gfCoV par PCR. Sur œufs embryonnés, deux voies d'inoculation ont été testées : la voie allantoïdienne et la voie amniotique. Des œufs de poule ont été utilisés dans un premier temps, puis nous avons tenté l'isolement sur œufs de pintade pour disposer d'un système de culture pleinement homologue. Seule la voie allantoïdienne a pu être testée sur œufs de pintade car leur opacité gêne le repérage de la cavité amniotique. Aucun de ces essais d'ovoculture n'a permis d'isoler le gfCoV non plus. Les coronavirus de dinde, les plus proches génétiquement du gfCoV, sont également difficiles à isoler. Ils se multiplient exclusivement dans le liquide amniotique d'œufs embryonnés de dinde (dont le volume et la facilité d'accès sont limités) (193). Les IBV en revanche se multiplient plus facilement sur œufs quelle que soit la voie d'inoculation (la voie allantoïque est privilégiée car elle permet d'obtenir un titre viral plus élevé en plus d'être facile d'accès).

A ce jour, les tentatives d'isolement n'ont donc pas été concluantes, ce qui n'est d'ailleurs pas très étonnant si l'on se réfère aux difficultés rencontrées avec les autres coronavirus entériques aviaires. Nous avons donc mis au point une méthode de titrage moléculaire permettant de mieux contrôler la dose inoculée lors d'infections expérimentales notamment. Une gamme étalon ciblant le gène de la nucléocapside N a donc été réalisée. Pour cela, le gène N de la souche vaccinale IBV H120 (souche Massachussetts) a été cloné

dans un plasmide accepteur (pSCB). Le plasmide obtenu a été quantifié et cette gamme nous a permis d'estimer le nombre de copies d'ARN présentes dans les échantillons (cas spontanés). Elle a également permis de mener une étude préliminaire de cinétique d'excrétion virale. Nous avons observé une excrétion précoce du virus (dès 1 jour post infection) expliquant la rapidité de dissémination de celui-ci. Des reproductions expérimentales de la maladie sont prévues afin de poursuivre la caractérisation des processus pathologiques. Nous avons déjà confirmé la possibilité d'une transmission par contact direct. Des travaux sur l'évolution virale sont également envisagés avec un suivi génétique du virus durant un passage en série. Pour cela des oiseaux seront infectés et mis en contact avec un lot sain, puis le lot d'oiseaux nouvellement infectés servira à son tour de « donneur » à un nouveau lot d'oiseaux sains. Chaque passage fera l'objet d'une autopsie et de prélèvements qui seront séquencés pour suivre l'apparition d'éventuelles mutations.

Dans le cadre de cette étude, le séquençage à haut débit sans *a priori* a permis de s'affranchir de difficultés rencontrées lors des tentatives d'identification par des méthodes classiques. Si on considère le nombre important d'essais qu'il aurait fallu réaliser avant de trouver l'agent étiologique de cette maladie (dans l'hypothèse où une méthode ait porté ses fruits), l'utilisation de cette nouvelle technique nous a permis de gagner beaucoup de temps et d'argent. De plus, la quantité très importante de séquences générées par cet outil a permis de couvrir plus de la moitié du génome complet, facilitant ainsi la suite de sa caractérisation génétique. Nous avons donc pu répondre à une question très précise par une approche de séquençage massif sans *a priori*. Les méthodes de virologie classique sont maintenant nécessaires pour approfondir l'étude du nouveau pathogène.

B. Articles

Novel Avian Coronavirus and Fulminating Disease in Guinea Fowl, France

Etienne Liais,¹ Guillaume Croville,¹
Jérôme Mariette, Maxence Delverdier,
Marie-Noëlle Lucas, Christophe Klopp,
Jérôme Lluch, Cécile Donnadieu, James S. Guy,
Léni Corrand, Mariette F. Ducatez,
and Jean-Luc Guérin

For decades, French guinea fowl have been affected by fulminating enteritis of unclear origin. By using metagenomics, we identified a novel avian gammacoronavirus associated with this disease that is distantly related to turkey coronaviruses. Fatal respiratory diseases in humans have recently been caused by coronaviruses of animal origin.

Fulminating disease (also referred to as X disease) of guinea fowl (*Numida meleagris*) is an acute enteritis characterized by intense prostration and a very high death rate, leading to the almost complete destruction of affected flocks. Lesions are generally limited to severe enteritis and, in some birds, pancreatic degeneration. This disease is uncommon, but its fulminating evolution raises concerns of differential diagnoses with highly pathogenic avian influenza.

Fulminating disease has been reported for decades in the French guinea fowl industry, and although its viral origin was previously suspected, the virus remained unknown. During the 1980s, French groups investigated the etiology of the disease. Because propagation on cells or embryos remained unsuccessful and molecular tools were unavailable, etiologic hypotheses were based mostly on electron microscopy findings. The groups reached different conclusions, proposing candidates such as toga-like virus (1), reovirus,

or herpesvirus (2). More recently, astroviruses have been associated with catarrhal enteritis in guinea poult (i.e., young guinea fowl) (3), but these viruses were not detected in birds affected by fulminating disease (data not shown).

We investigated field cases and performed an experimental reproduction of fulminating disease and identified its agent by using unbiased high-throughput sequencing. We propose a gammacoronavirus of a novel genotype as the most likely causal agent of fulminating disease. Coronaviruses (CoVs) are zoonotic. The novel human Middle East respiratory syndrome CoV, a betacoronavirus that was first isolated in 2012 in Saudi Arabia, is most closely related to *Tytonycteris* bat CoV HKU4 and *Pipistrellus* bat CoV HKU5 (4); severe acute respiratory syndrome-CoV originated from a betacoronavirus that spread from bats to civets and humans (5).

The Study

We investigated field cases of fulminating disease among 5 flocks of guinea fowl in France during 2010 and 2011; in all cases, the birds' clinical signs were limited to severe prostration, a dramatic decrease in water and feed consumption, and a daily death rate of up to 20%. From each affected flock, 5–10 sick birds were euthanized, and necropsy was performed. Most birds showed marked congestive enteritis and a whitish, enlarged pancreas. Livers, spleens, pancreas, kidneys, and intestinal tracts from the birds were placed into 10% buffered formalin for histopathologic examination or stored at -80°C for virologic analyses.

Fifteen 6-week-old guinea fowl poult were housed in 2 poultry isolators. One group of 5 birds that had been orally inoculated with the clarified and filtered (0.45- μ m mesh) intestinal contents of diseased birds from 1 field case flock shared an isolator with another group of 5 uninoculated guinea poults, placed as contact birds. A third group of 5 uninoculated birds were placed in the other isolator. Inoculated and contact birds showed severe prostration as early as 2 days after infection and died or had to be euthanized within 6 days. Uninoculated birds showed no clinical signs of illness throughout the experiment. Necropsy was performed on birds that died or were euthanized, and comprehensive samples of tissues and fluids were obtained for subsequent investigations. At necropsy, the most notable lesion was an acute and marked enteritis; in some birds, a mildly enlarged and whitish pancreas was observed.

The intestinal contents of experimentally infected poults were pooled, clarified by centrifugation, and pelleted by ultracentrifugation (100,000 \times g, 2 h). The pellets were treated with RNase (Ambion, 20 μ g/mL) and DNase (Invitrogen, 10 U/ μ L) (both from Life Technologies, Grand

Author affiliations: French National Institute for Agricultural Research (INRA), Toulouse, France (E. Liais, G. Croville, M. Delverdier, M.-N. Lucas, L. Corrand, M.F. Ducatez, J.-L. Guérin); Ecole Nationale Vétérinaire, Toulouse (E. Liais, G. Croville, M. Delverdier, M.-N. Lucas, L. Corrand, M.F. Ducatez, J.-L. Guérin); INRA 31326, Castanet-Tolosan, France (J. Mariette, C. Klopp; J. Lluch, C. Donnadieu); and North Carolina State University, Raleigh, North Carolina, USA (J. Guy)

DOI: <http://dx.doi.org/10.3201/eid2001.130774>

¹These authors contributed equally to this article.

Table 1. Distribution of reads generated by sequencing of pooled intestinal contents from guinea fowl poults with fulminating disease France, 2010–2011*

Read type	No. (%) reads
Total reads generated	476,430
Cellular reads	142,739 (30)
Bacterial reads	246,787 (51.8)
Archaea reads	35,271 (7.4)
Phage, plant, and insect virus reads	32,477 (6.8)
Eukaryote virus reads	19,155 (4.0)
Coronavirus reads	1,441 (7.5)†

*Sequencing was performed by using MiSeq (Illumina, San Diego, CA, USA) and analyzed by using the GAAS (<http://gaas.sourceforge.net/>) algorithm with an expected value of 10^{-3} (7).

†Percent within eukaryote virus reads.

Island, NY, USA) to concentrate the viral material. RNA and DNA were extracted separately, and a random reverse transcription PCR was performed, as described (6), to generate unbiased PCR products of ≈ 300 bp. High-throughput sequencing was performed by using a MiSeq platform (Illumina, San Diego, CA, USA). A total of 476,430 sequences were generated (Table 1), 10.8% of which matched with known viral sequences, as determined by using GAAS software (<http://gaas.sourceforge.net/>) (7) with an expected value of 10^{-3} ; 7.5% of the eukaryotic viral reads were similar to avian CoVs, such as turkey CoV (TCoV) and infectious bronchitis virus (IBV). A total of 1,444 reads related to CoVs were aligned against the most similar TCoV genome available in GenBank (TCoV/CA/MG10; accession no. EU095850). The reads fairly aligned to almost the whole TCoV genome; the overall coverage was 78.86% (data not shown). A CoV-specific reverse transcription PCR was performed and the result was positive, specifically in intestinal tissues of experimentally infected birds (Table 2) (8). Furthermore, immunochemistry was performed by using a monoclonal antibody specific for TCoVs (9). We observed an intense, cytoplasmic, and granular labeling in enterocyte villi of inoculated birds only (Figure 1), suggesting a substantial intestinal replication of a TCoV-related virus. A 3,680-nt consensus full sequence of the spike (S) gene was completed by classical PCR and Sanger sequencing (GenBank/EMBL accession no. HF544506). A BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) search followed by a phylogenetic analysis performed on the complete S gene showed that guinea fowl fulminating enteritis virus corresponds to a distinct genotype of CoV, clustering within *Gammacoronavirus* genus, which also includes TCoV and

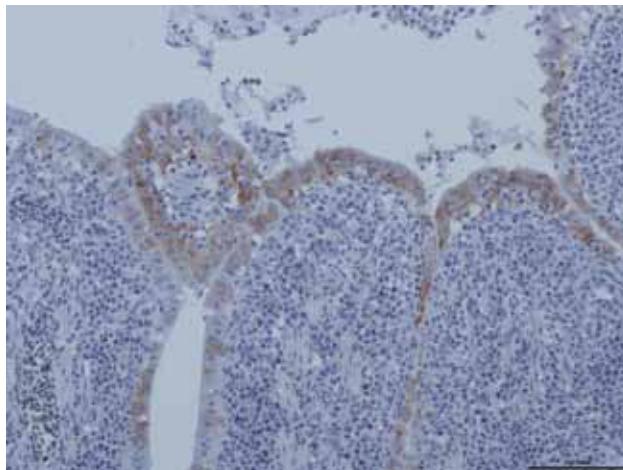


Figure 1. Replication of guinea fowl coronavirus in intestinal epithelium cells of experimentally infected birds as evidenced by immunohistochemical testing with a turkey coronavirus-specific monoclonal antibody, France, 2010–2011 (9). Scale bar indicates 0.1 mm.

IBV (10). The complete S gene sequence of guinea fowl CoV (GFCoV) was most similar to that of the TCoV S gene (minimum Kimura distance 22.6% at the nucleotide level between GFCoV/FR/2011 and TCoV/VA/74/2004, TCoV/CA/MG10, TCoV/IN/517/1994, or TCoV/ATCC). The GFCoV S gene was more similar to those of North American TCoV than to those of European (French) TCoV strains (Figure 2) (11).

CoVs are positive-sense RNA viruses that are subject to frequent mutations and recombination events, resulting in the emergence of novel viruses, such as severe acute respiratory syndrome (in 2003) and Middle East respiratory syndrome CoV in humans (4,5). Avian CoVs associated with nonclinical carriage or with respiratory, genital, renal, or enteric diseases have been identified in many avian species (12). CoV infection causes mild enteritis in different avian species, mainly turkeys, partridges, and quails. Avian CoVs are usually classified as gammacoronaviruses, although a few bird CoVs have also recently been described as deltacoronaviruses (13).

In the past, a recombination event led to the emergence of TCoV: the S gene of IBV recombined with an unknown virus (likely of avian origin), which resulted in a host change (chicken to turkey) and a tropism switch (respiratory to enteric). IBV and TCoV share <36% similarity for the S gene,

Table 2. Tissue tropism of coronavirus in experimentally and naturally infected guinea fowl poults, as detected by reverse transcription PCR, France, 2010–2011*

Case	Duodenum	Ileum/colon	Pancreas	Spleen	Bursa of Fabricius
Inoculated	5/5	5/5	0/5	0/5	1/5
Contact	3/5	5/5	0/5	0/5	1/5
Uninfected control	0/5	0/5	0/5	0/5	0/5
Field†	5/5	5/5	NT	NT	NT

*Data are no. of coronavirus PCR-positive birds/total no. of birds. NT, not tested.

†Intestines (pooled duodenum, ileum, colon samples) from 5 field cases were tested (5–10 birds tested per clinical case). The 5 cases were selected on the basis of clinical signs (severe prostration and death rate >50%).



Figure 2. Phylogenetic analysis of the complete spike (S) gene of the fulminating disease of guinea fowl coronavirus (gfCoV, in **boldface**) in relation to turkey coronaviruses (TCoVs) at the nucleotide level. The tree was generated by using MEGA 5.05 (<http://megasoftware.net/mega.php>) by the neighbor-joining method, Kimura 2-parameter model, and pairwise deletion. Bootstrap values (1,000 replicates) >75 are indicated on the nodes. Only a partial S gene sequence (1,657 nt) was available for quail CoV/Italy/Elvia/2005. Scale bar indicates Kimura distance.

but their full genomes are >86% similar (14). Although the origin of GFCoV is still unknown, the distance of its S gene to TCoV and IBV S genes suggests not only a common ancestor but also a current separate evolutionary path. A quail CoV similar to TCoV has been described (15); its available S gene sequence also clusters with North American TCoV (Figure 2). However, the comparable partial S gene sequences of quail CoV/Italy/Elvia/2005 and GFCoV differ greatly (genetic distance 30%). A few cases of guinea fowl fulminating disease are diagnosed each year in France but have no apparent epidemiologic link to each other. The severity of the disease in the field may suggest a poor adaptation of the pathogen to guinea fowl. This pathologic pattern differs greatly from TCoV enteritis in turkeys and makes GFCoV of potential interest for comparative studies of CoV pathobiology. Virus reemergence may indicate that ≥ 1 other species may also be asymptomatic carriers.

Conclusion

We identified an avian gammacoronavirus related to TCoVs in fulminating disease of guinea fowl. The epidemiologic reservoir of this virus still needs to be clarified, and the sequencing of the full genome of the pathogen is warranted to fully assess its phylogenetic relationship with other gammacoronaviruses, its epidemiology, and its origin.

Acknowledgments

We thank Simon Roux for help with bioinformatics.

E.L. is supported by a grant from the French Ministry of Agriculture and Region Midi-Pyrénées. This project was supported by the French Comité Interprofessionnel de la Pintade and by the French Ministry of Agriculture.

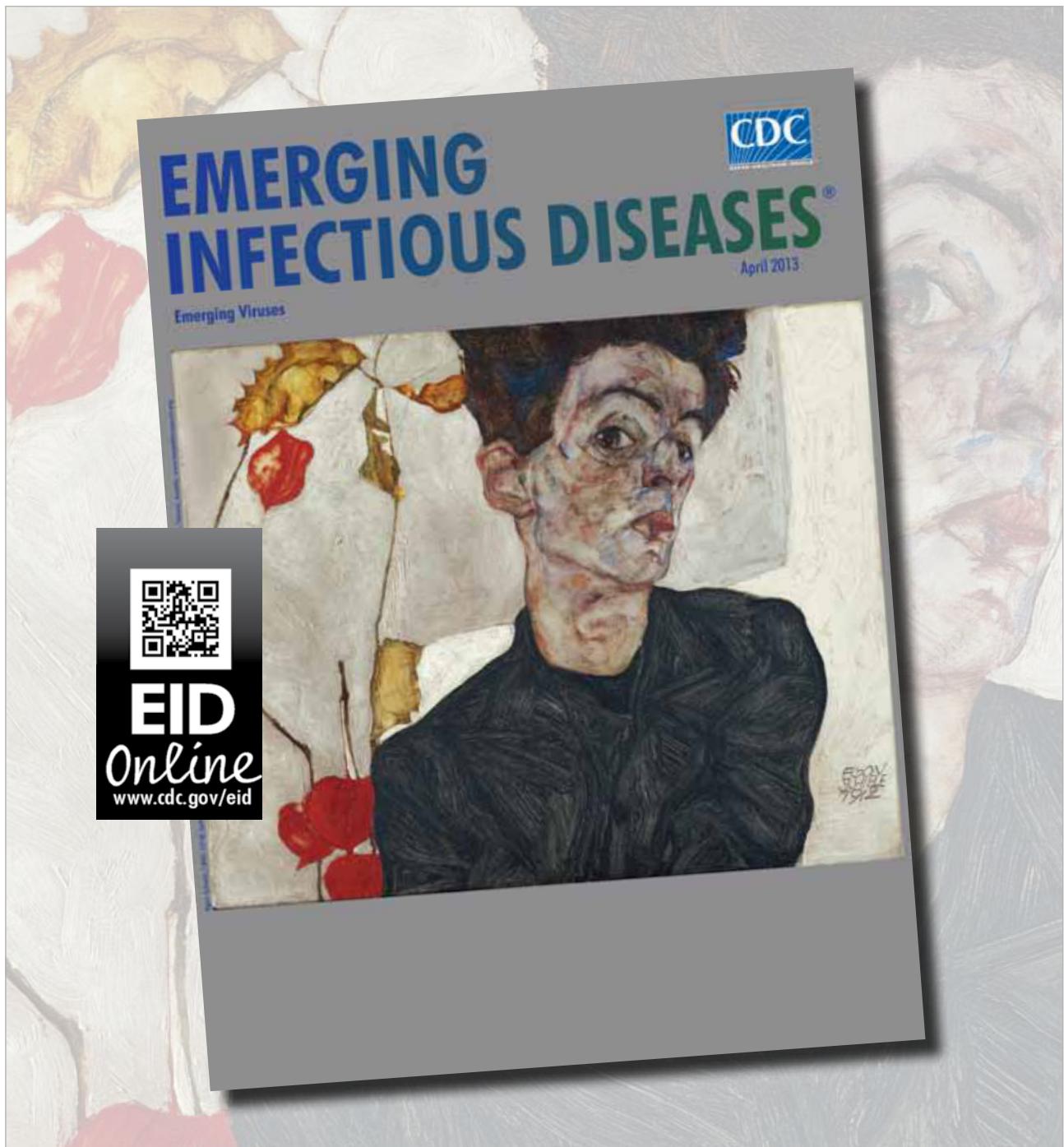
Mr Liais is a PhD candidate in virology at the Host-Pathogen Agents Interactions Joint Research Unit, National Veterinary School, and INRA, Toulouse, France. His research interests are the host interactions of avian influenza viruses and metagenomic approaches in virology.

References

1. Brahem A, Demarquez N, Beyrie M, Vuillaume A, Fleury HJ. A highly virulent togavirus-like agent associated with the fulminating disease of guinea fowl. *Avian Dis.* 1992;36:143–8. <http://dx.doi.org/10.2307/1591729>
2. Andral B, Lagadic M, Louzis C, Guillou JP, Gourreau JM. Fulminating disease in guinea fowl: etiological investigation. *Point Veterinaire.* 1987;19:515–20.
3. Toffan A, Catania S, Salvato A, De Battisti C, Vascellari M, Toson M, et al. Experimental infection of poult and guinea fowl with genetically distinct avian astroviruses. *Avian Pathol.* 2012;41:429–35. <http://dx.doi.org/10.1080/03079457.2012.704980>
4. van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* 2012;3:pil00473-12.
5. Gouilh MA, Puechmaile SJ, Gonzalez J-P, Teeling E, Kittayapong P, Manuguerra J-C. SARS-coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. *Infect Genet Evol.* 2011;11:1690–702. <http://dx.doi.org/10.1016/j.meegid.2011.06.021>
6. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol.* 2009;83:4642–51. <http://dx.doi.org/10.1128/JVI.02301-08>
7. Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, et al. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLOS Comput Biol.* 2009;5:e1000593. <http://dx.doi.org/10.1371/journal.pcbi.1000593>
8. Stephensen CB, Casebolt DB, Gangopadhyay NN. Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay. *Virus Res.* 1999;60:181–9. [http://dx.doi.org/10.1016/S0168-1702\(99\)00017-9](http://dx.doi.org/10.1016/S0168-1702(99)00017-9)
9. Breslin JJ, Smith LG, Barnes HJ, Guy JS. Comparison of virus isolation, immunohistochemistry, and reverse transcriptase–polymerase chain reaction procedures for detection of turkey coronavirus. *Avian Dis.* 2000;44:624–31. <http://dx.doi.org/10.2307/1593102>
10. Jackwood MW, Hall D, Handel A. Molecular evolution and emergence of avian gammacoronaviruses. *Infect Genet Evol.* 2012;12:1305–11. <http://dx.doi.org/10.1016/j.meegid.2012.05.003>
11. Maurel S, Toquin D, Briand FX, Queguiner M, Allee C, Bertin J, et al. First full-length sequences of the S gene of European isolates reveal further diversity among turkey coronaviruses. *Avian Pathol.* 2011;40:179–89. <http://dx.doi.org/10.1080/03079457.2011.551936>
12. Cavanagh D. Coronaviruses in poultry and other birds. *Avian Pathol.* 2005;34:439–48. <http://dx.doi.org/10.1080/03079450500367682>

13. Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, Lau JHN, et al. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *J Virol*. 2012;86:3995–4008. <http://dx.doi.org/10.1128/JVI.06540-11>
14. Jackwood MW, Boynton TO, Hilt DA, McKinley ET, Kissinger JC, Paterson AH, et al. Emergence of a group 3 coronavirus through recombination. *Virology*. 2010;398:98–108. <http://dx.doi.org/10.1016/j.virol.2009.11.044>
15. Circella E, Camarda A, Martella V, Bruni G, Lavazza A, Buonavoglia C. Coronavirus associated with an enteric syndrome on a quail farm. *Avian Pathol*. 2007;36:251–8. <http://dx.doi.org/10.1080/03079450701344738>

Address for correspondence: Jean-Luc Guérin, INRA, UMR 1225, École Nationale Vétérinaire de Toulouse, 23, Chemin des Capelles 31076 Toulouse Cedex 3, France; email: jl.guerin@envt.fr



AN AVIAN CORONAVIRUS ASSOCIATED WITH FULMINATING DISEASE OF GUINEA FOWL

E. Liais^{1,2*}, G. Croville^{1,2}, L. Corrand^{1,2}, M. Delverdier^{1,2}, J. Mariette³, M.-N. Lucas^{1,2}, C. Klopp³, J. Lluch⁴, C. Donnadieu⁴, J.S. Guy⁵, M. Ducatez^{1,2} and J.-L. Guérin^{1,2}

¹ INRA, UMR 1225 and Université de Toulouse

² Ecole Nationale Vétérinaire, 23, chemin des Capelles 31076 Toulouse, France

³ Plateforme bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA, 31326, Castanet-Tolosan, France

⁴ GeT-PlaGe, INRA ENVT UMR444, laboratoire de Génétique cellulaire 31326, Castanet-Tolosan, France

⁵ College of Veterinary Medicine, NCSU, Raleigh, NC, USA

Summary

Fulminating disease of Guinea fowl is an acute enteritis characterized by a very high mortality. This disease has been reported for decades in French guinea fowl industry and although its viral origin was previously suspected, the virus remained unknown. Here, we reproduced experimentally the disease by inoculating guinea poults with intestinal content of diseased birds and identified the causative agent using unbiased high throughput sequencing (1).

Introduction

X disease is an acute enteritis has been reported for decades in French guinea fowl industry. This disease also called fulminating disease is characterized by a very high mortality and generally pancreatic degeneration. First investigations have been carried out in the 1980s' to identify the etiological cause of the disease. While a few candidates have been proposed, none had been clearly detected in affected birds (2,3,4). In this study, we reproduced experimentally the disease by inoculating guinea poults with intestinal content of diseased birds and identified a gammacoronavirus as most likely the causative agent (1).

Material and Methods

We investigated field cases of guinea fowl fulminating disease in 2010 and 2011 and collected intestinal content from diseased birds. We inoculated 6 weeks old guinea poults with clarified and filtered (0.45 µm mesh) intestinal content. Three groups of five birds were constituted: (i) the first group was orally inoculated with the field filtered sample, (ii) the second group was placed in the same isolator as inoculated birds, (iii) the third group was housed in a distinct isolator and

* Corresponding author

was not infected: it was our negative control group. Viral particles of intestinal contents were concentrated by ultracentrifugation and nuclease treatment (RNase (Ambion, 20µg/mL) and DNase (Invitrogen, 10U/µL)). RNA and DNA were extracted and a (RT-)PCR was optimized to generate PCR products of 300bp thanks to tagged random primers previously described (5). An unbiased metagenomic analysis was performed by using MiSeq platform (Illumina, San Diego, CA, USA). Data reads were analysed using GAAS software (<http://gaas.sourceforge.net/>) with an E-value of 10^{-3} . Phylogenetic analysis was performed using MEGA4.

Results

In our *in vivo* experiment, all inoculated and contact birds (groups 1 and 2) showed clinical signs (severe prostration, decrease in water and feed consumptions) and died or had to be sacrificed within 6 days (Table 1). The unbiased high-throughput sequencing analysis identified an avian coronavirus which was confirmed by coronavirus-specific RT-PCR on samples from both experimentally infected birds and spontaneous field cases. 476,430 sequences were generated, of which 10.8 % matched known viral sequences; and 7.5 % of the eukaryotic viral reads were similar to avian coronavirus sequences. Blast and phylogenetic analyses on the virus genome showed that this fulminating enteritis virus is indeed a coronavirus (named guinea fowl coronavirus, GfCoV) that clusters with gammacoronaviruses and especially Turkey coronavirus (TCoV). GfCoV spike (S) gene was interestingly more similar to North American TCoV than to European (French) TCoV strains S genes.

Discussion

We have identified a gammacoronavirus in French guinea fowls that we named gfCoV. The very high mortality associated with the fulminating disease in the field may suggest a poor adaptation of the pathogen to its host, which greatly contrasts with previously characterized enteritic avian gammacoronaviruses such as TCoV so far associated with limited morbidity. The epidemiological reservoir of gfCoV still needs to be clarified and further analyses are warranted to fully assess its phylogenetic relationship with other gammacoronaviruses.

In conclusion we showed that an avian coronavirus is the agent of “fulminating disease” of guinea fowl. Coronaviruses cause enteritis in different avian species. Both pathological and epidemiological patterns of this peracute infection are still to be clarified.

Acknowledgement

The authors would like to acknowledge the support of the COST Action FA1207.

References

- ¹ Liais E, Croville G, Mariette J, Delverdier M, Lucas MN, Klopp C, Lluch J, Donnadiou C, Guy JS, Corrand L, Ducatez MF, Guérin JL. Novel avian coronavirus associated with fulminating disease in guineafowl, France, 2010-2011. *Emerg Infect Dis.* 2014;20(1):105-8.
- ² Brahem A, Demarquez N, Beyrie M, Vuillaume A, Fleury HJ. A highly virulent togavirus-like agent associated with the fulminating disease of guineafowl. *Avian Dis.* 1992;36:143-8 .
- ³ Andral B, Lagadic M, Louzis C, Guillou JP, Gourreau JM. Fulminating disease in guinea fowl: etiological investigation. *Point Veterinaire.* 1987;19:515-20.
- ⁴ Toffan A, Catania S, Salviato A, De Battisti C, Vascellari M, Toson M, et al. Experimental infection of poults and guinea fowl with genetically distinct avian astroviruses. *Avian Pathol.* 2012;41:429-35.
- ⁵ Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, et al. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. *J Virol.* 2009;83(9):4642-51.

Table 1: Clinical signs after experimental infection of guinea fowls with gfCoV. Green: absence of clinical signs; yellow: severe prostration, decrease in water and feed consumptions; red: death. BF: bursa of Fabricius.

Batch	Experimental follow-up								Clinical and necropsic recording	
	D1	D2	D3	D4	D5	D6	D7	D8		
Inoculated birds (group 1)	A1	Green	Green	Yellow	Yellow	Red				Degenerative pancreas +++ / atrophy Thymus & BF
	A2	Green	Green	Yellow	Yellow	Red				-
	A3	Green	Red							-
	A4	Green	Green	Yellow	Yellow	Red				Degenerative pancreas +++ / atrophy Thymus & BF
	A5	Green	Green	Yellow	Yellow	Yellow	Red			Degenerative pancreas +++ / atrophy Thymus & BF
Contact birds (group 2)	C1	Green	Green	Yellow	Red					-
	C2	Green	Green	Red						Muroid enteritis
	C3	Green	Green	Yellow	Yellow	Yellow	Red			Degenerative pancreas +++ / atrophy Thymus & BF
	C4	Green	Green	Yellow	Red					Degenerative pancreas +
	C5	Green	Green	Yellow	Red					-
Control birds (group 3)	T1	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Lameness
	T2	Green	Green	Green	Green	Green	Green	Green	Green	-
	T3	Green	Green	Green	Green	Green	Green	Green	Green	-
	T4	Green	Green	Green	Green	Green	Green	Green	Green	-
	T5	Green	Green	Green	Green	Green	Green	Green	Green	-

Full genome sequence of French guinea fowl coronavirus associated with fulminating disease

Mariette F. Ducatez*, Etienne Liais, and Jean-Luc Guérin

Affiliation:

INRA, UMR 1225 and Université de Toulouse, INP, ENVT, 31076 Toulouse, France

* Corresponding author: Mariette F. Ducatez, m.ducatez@envt.fr, Tel: +33561193249, Fax: +33561193975

Keywords: coronavirus, Nidovirus, guinea fowl, genome organization

Running title: Full genome sequence of French Guinea fowl coronavirus

Abstract

Guinea fowl coronavirus (gfCoV), a recently characterized avian coronavirus, was identified from outbreaks of fulminating disease (peracute enteritis) in guinea fowls in France. The full-length genomic sequence was determined to better understand its genetic relationship with avian coronaviruses. The full-length coding genome sequence was 26,985 nucleotides long with 11 open reading frames and no hemagglutinin-esterase gene: a genome organization identical to that of turkey Coronavirus [5' UTR – replicase (ORFs 1a, 1ab) – spike (S) protein – ORF3 (ORFs 3a, 3b) – small envelop (E or 3c) protein – membrane (M) protein – ORF5 (ORFs 4b, 4c, 5a, 5b) – nucleocapsid (N) protein (ORFs N and 6b) – 3' UTR]. This is the first complete genome sequence of a gfCoV and confirms that the new virus belongs to group III coronaviruses.

Coronaviruses (CoVs) are enveloped viruses with positive-sense, non-segmented RNA genomes of 27 to 32 kb. CoVs infect a wide range of hosts causing various degrees of morbidity and mortality. Group I Coronaviruses (Alphacoronaviruses) contain viruses that infect for example humans (HCoV-229E and HCoV-NL63) but also cats and dogs (with feline CoV and canine CoV, respectively), or pigs (with the porcine transmissible gastroenteritis virus, TGEV for example). Similarly, group II CoVs (Betacoronaviruses) may infect humans (examples: HCoV-OC43, HCoV-HKU1, Severe Acute Respiratory Syndrome (SARS), or Middle East Respiratory Syndrome (MERS) virus), horses (with ECoV), or cattle (with BCoV). In contrast, group III CoVs (Gammacoronaviruses) primarily infect birds: chickens, peafowl, and partridges harbour infectious bronchitis virus (IBV) while turkeys have turkey CoV (TCoV) and guinea fowl may be infected with guinea fowl CoV (gfCoV).

Gammacoronavirus strains have however been isolated from a whale and a wild felid (Gavier-

widen D, 2012). Group IV CoVs (Deltacoronaviruses) have been detected in birds (with BuCoV, MuCoV, SpCoV, etc), or pigs (with PorCoV) (Chan et al., 2012). Interestingly CoVs of the Groups I, II, and IV have been detected in Chiroptera (bats), thought to be the reservoir of CoVs (Balboni et al., 2012; Chen et al., 2011).

In the present study we focused on a new member of the group III CoVs, gfCoV, and aimed at sequencing its full genome to better understand its molecular relationship with gammacoronaviruses.

To determine the full genome of GfCoV/FR/2011, we first analysed the data generated on a MiSeq Illumina platform and previously described (Liais et al., 2014). Briefly, pooled intestinal contents of experimentally infected guinea fowl poult were clarified, ultracentrifuged, and treated with nucleases to concentrate viral material. RNA was extracted, and a random RT-PCR was performed to generate unbiased PCR products of about 300 bp (Liais et al., 2014; Victoria et al., 2009). The sequences generated that matched with avian coronaviruses sequences, as determined using GAAS software (Angly et al., 2009), were extracted for further analysis and visualized using Integrative Genomics Viewer (IGV) with the closest Blast hit as reference genome: TCoV MG10 (accession number: EU095850) (Thorvaldsson et al., 2013). Primers were designed based on the known sequence data to amplify missing genome fragments by PCR. Sanger sequencing was then performed with PCR primers. The full genome sequence was submitted to EMBL and was attributed the following accession number: (in process). Sequence analysis was carried out using BioEdit version 7.0.8.0 (Hall, 1999), Muscle for the alignment (Edgar, 2004), and Mega version 5.05 for the phylogeny (Tamura et al., 2011).

The sequences were assembled into one contiguous coding sequence of 26,985 nucleotides was obtained. The entire genome had a GC content of 38.3%, identical to the turkey coronavirus (TCoV) MG10 genome (Gomaa et al., 2008).

GfCoV and TCoV genomes have the same organization: (i) a 5' untranslated region (UTR), (ii) two large slightly overlapping ORFs coding for the replicase: 1a and 1ab, (iii) gene coding for the spike (S) protein, (iv) ORF3 (ORFs 3a, 3b), (v) gene coding for the small envelop (E or 3c) protein, (vi) gene coding for the membrane (M) protein, (vii) ORF5 (ORFs X (4b and 4c), 5a, 5b), (viii) genes coding for the nucleocapsid (N) protein (ORFs N and 6b), and (ix) 3' UTR (Table 1). While some viral proteins were of the exact same size such as 3a, 3b, 4b, 5a, and N with 57, 64, 94, 65, and 409 aa, respectively; variability was observed for the other proteins but not more between gfCoV and the TCoV available genomes than among TCoV genomes. Interestingly, gfCoV/FR/2011 harboured a shorter small envelop protein than its TCoV counterparts (Table 1). Further studies are warranted to understand the impact of avian coronaviruses protein sizes in the biology of the viruses.

Phylogenetic analysis on the full genome of GfCoV/FR/2011 showed it clearly clustered with North American TCoV strains (Figure1, supported by a high bootstrap value of 500), as was observed when a phylogeny of the S gene only had been carried out (Liais et al., 2014). The genetic distance between gfCoV/FR/2011 and TCoV ranged between 10.7% and 11.4%, while these distances were larger between gfCoV/FR/2011 and the representative IBV strains tested as expected: 13.5% to 15.0% (Table 2). A simplot analysis comparing the GfCoV/FR/2011 full genome to its closest TCoV and infectious bronchitis virus (IBV) Blast hits showed that the 3 genomes are highly similar throughout the genome (74 to 100% similarity, with no significantly higher identity of GfCoV/FR/2011 with TCoV or IBV

genomes), except for the S gene (Figure 2). GfCoV S gene was indeed more closely related to its closest TCoV than IBV strain but also more similar to both viruses on the S gene than on the rest of its genome (Figure 2), suggesting a recombination event as was hypothesized for the origin of TCoV (Jackwood et al., 2010).

The present study showed that gfCoV/FR/2011 harbours a genome organization very similar to that of TCoV strains. In addition, and again like TCoV, gfCoV/FR/2011 likely originated from a recombination event between an IBV-like (or TCoV-like) virus that would have given most of its genome and a so far unknown CoV that would have contributed by its spike gene. Despite the similarity of their genomes and their enteric tropism, TCOVs cause little clinical signs while gfCOVs are usually associated with extremely high mortalities. Further studies are ongoing to understand the host range of gfCoV/FR/2011 and its determinants of pathogenicity.

Acknowledgements

This work was supported by the 'EPICOREM' grant of the Agence Nationale de la Recherche (ANR), by the French Comité Interprofessionnel de la Pintade (CIP) and by the French Ministry of Agriculture. E.L. is supported by a scholarship from French Ministry of Agriculture and Region Midi-Pyrénées. We would like to thank the Plateforme bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA, 31326, Castanet Tolosan, France; the GeT-PlaGe, INRA ENVT UMR444, laboratoire de Génétique cellulaire 31326, Castanet-Tolosan, France; and the Plateau de Génomique GeT-Purpan, UDEAR UMR 5165 CNRS/UPS, CHU PURPAN, Toulouse, France, for technical assistance with sequencing and sequence analysis.

References

- Angly, F.E., Willner, D., Prieto-Davo, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., Antonopoulos, D.A., Barott, K., Cottrell, M.T., Desnues, C., Dinsdale, E.A., Furlan, M., Haynes, M., Henn, M.R., Hu, Y., Kirchman, D.L., McDole, T., McPherson, J.D., Meyer, F., Miller, R.M., Mundt, E., Naviaux, R.K., Rodriguez-Mueller, B., Stevens, R., Wegley, L., Zhang, L., Zhu, B., Rohwer, F., 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS computational biology* 5, e1000593.
- Balboni, A., Battilani, M., Prosperi, S., 2012. The SARS-like coronaviruses: the role of bats and evolutionary relationships with SARS coronavirus. *The new microbiologica* 35, 1-16.
- Cao, J., Wu, C.C., Lin, T.L., 2008. Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. *Virus research* 136, 43-49.
- Chan, J.F., Li, K.S., To, K.K., Cheng, V.C., Chen, H., Yuen, K.Y., 2012. Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic? *The Journal of infection* 65, 477-489.
- Chen, H., Guan, Y., Fan, X., 2011. Animal reservoirs for SARS-like coronavirus in southern China. *Hong Kong medical journal = Xianggang yi xue za zhi / Hong Kong Academy of Medicine* 17 Suppl 6, 36-40.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32, 1792-1797.
- Gavier-widen D, R.-D.M., Decaro N, Buonavoglia C. 2012. Coronavirus infections, In: Gavier-Widen D, D.J., Meredith A (Ed.) *Infectious diseases of wild mammals and birds in Europe*. Blackwell, 234-240.

Gomaa, M.H., Barta, J.R., Ojkic, D., Yoo, D., 2008. Complete genomic sequence of turkey coronavirus. *Virus research* 135, 237-246.

Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis for Windows 95/98/NT. In: *Nucleic Acids Symp*, 95-98.

Jackwood, M.W., Boynton, T.O., Hilt, D.A., McKinley, E.T., Kissinger, J.C., Paterson, A.H., Robertson, J., Lemke, C., McCall, A.W., Williams, S.M., Jackwood, J.W., Byrd, L.A., 2010. Emergence of a group 3 coronavirus through recombination. *Virology* 398, 98-108.

Liais, E., Croville, G., Mariette, J., Delverdier, M., Lucas, M.N., Klopp, C., Lluch, J., Donnadieu, C., Guy, J.S., Corrand, L., Ducatez, M.F., Guerin, J.L., 2014. Novel avian coronavirus and fulminating disease in guinea fowl, France. *Emerging infectious diseases* 20, 105-108.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* 28, 2731-2739.

Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192.

Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., Delwart, E., 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *Journal of virology* 83, 4642-4651.

Table 1. Genes and coding regions for gfCoV/FR/2011

ORF	Location	Size in nt	Size in aa	Size in aa	Size in aa	Size in aa	Size in aa	Size in aa	Size in aa	Size in aa
	(gfCoV)	(gfCoV)	(gfCoV)	(TCoV/VA	(TCoV/TX	(TCoV/TX	(TCoV/IN	(TCoV/	(TCoV/	(TCoV/
				-74/03) ¹	-GL/01) ¹	-1038/98) ¹	-517/94) ¹	ATCC) ²	540) ²	MG10) ³
5' UTR*	1-463	>463	-	-	-	-	-	-	-	-
1a	464-12,307	11,844	3,948	3,947	3,949	3,950	3,952	3,957	3,945	3,951
1ab (<i>1b</i>)	464-12,280	19,884	6,628	6,596	6,602	6,602	6,605	2,654	2,652	6,601
	and 12,280-	8,067	2,689							
	20,346									
S	20,294-23,914	3,621	1,207	1,226	1,225	1,224	1,226	1,203	1,203	1,226
3a	23,917-24,087	171	57	57	57	57	57	57	57	57
3b	24,090-24,281	192	64	64	64	64	64	64	64	64
E (3c)	24,265-24,540	276	92	99	109	99	99	103	99	99
M	24,543-25,214	672	224	223	225	223	223	223	222	223
4b	25,218-25,499	282	94	94	94	94	94	94	94	94
4c	25,423-25,533	111	37					52	56	

5a	25,578-25,772	195	65	65	65	65	65	65	65	65
5b	25,772-26,011	240	80	82	82	92	82	82	80	82
N	25,957-27,183	1,227	409	409	409	409	409	409	409	409
6b	27,191-27,445	255	85					74	73	
3' UTR*	27,447	27,471	-	-				-	-	-

* Incomplete sequences, nt: nucleotides, aa: amino acids

¹ as described in (Jackwood et al., 2010)

² as described in (Cao et al., 2008) with 1b described rather than 1ab (size in aa in italic font)

³ as described in (Gomaa et al., 2008)

Table 2. Estimates of evolutionary distances (in number of substitutions per site) over sequence pairs between IBV, TCoV, and GfCoV/FR/2011.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1-GF-CoV/FR/2011		<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>
2-EU022526_TCoV-ATCC	0.111		<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>						
3-GQ427176_TCoV/IX-1038/98	0.108	0.073		<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>						
4-GQ427174_TCoV/IX-GL/01	0.114	0.079	0.039		<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>
5-GQ427173_TCoV/VA-74/03	0.112	0.076	0.046	0.046		<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>						
6-EU095850_MG10	0.111	0.075	0.050	0.045	0.031		<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>							
7-GQ427175_TCoV/IN-517/94	0.107	0.069	0.056	0.057	0.050	0.043		<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>
8-EU022525_TCoV-540	0.114	0.076	0.069	0.070	0.066	0.062	0.039		<i>0.002</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>												
9-FN430415_NGA/A116E7/2006	0.135	0.138	0.128	0.134	0.133	0.134	0.131	0.138		<i>0.002</i>														
10-FN430414_ITA/90254/2005	0.136	0.146	0.138	0.145	0.143	0.145	0.140	0.147	0.100		<i>0.002</i>	<i>0.003</i>	<i>0.002</i>											
11-KF377577_strain_4/91_vaccine	0.147	0.150	0.140	0.145	0.146	0.149	0.144	0.150	0.107	0.113		<i>0.002</i>	<i>0.003</i>	<i>0.003</i>	<i>0.002</i>	<i>0.002</i>								
12-AJ311317_Beaudette_CK	0.149	0.134	0.134	0.132	0.137	0.137	0.133	0.139	0.109	0.113	0.111		<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.003</i>	<i>0.002</i>
13-AY851295_strain_Mass_41	0.150	0.136	0.131	0.128	0.138	0.136	0.133	0.140	0.108	0.114	0.113	0.067		<i>0.001</i>	<i>0.001</i>	<i>0.002</i>								
14-AY641576_isolate_Peafowl/GD/KQ6/2003	0.150	0.137	0.133	0.128	0.137	0.135	0.134	0.138	0.107	0.115	0.106	0.068	0.020		<i>0.002</i>									
15-GU393336_serotype_Holte	0.139	0.120	0.124	0.129	0.129	0.126	0.122	0.131	0.100	0.109	0.109	0.085	0.085	0.087		<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>
16-GU393337_serotype_lowa_97	0.139	0.119	0.121	0.125	0.128	0.125	0.121	0.129	0.100	0.108	0.105	0.086	0.086	0.088	0.006		<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>
17-GU393334_serotype_Gray	0.136	0.112	0.113	0.118	0.114	0.113	0.107	0.115	0.095	0.104	0.106	0.087	0.083	0.085	0.071	0.071		<i>0.000</i>	<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>
18-GU393338_serotype_JMK	0.136	0.112	0.113	0.119	0.114	0.113	0.107	0.116	0.095	0.104	0.105	0.087	0.082	0.085	0.071	0.071	0.003		<i>0.001</i>	<i>0.002</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.002</i>
19-KF696629_strain_Connecticut_vaccine	0.137	0.109	0.099	0.106	0.106	0.107	0.101	0.117	0.093	0.104	0.103	0.075	0.075	0.076	0.068	0.069	0.064	0.063		<i>0.001</i>	<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>
20-AY514485_serotype_California_99	0.148	0.129	0.101	0.108	0.112	0.119	0.115	0.129	0.102	0.112	0.105	0.102	0.100	0.100	0.089	0.084	0.081	0.079	0.055		<i>0.001</i>	<i>0.002</i>	<i>0.002</i>	<i>0.001</i>
21-GQ504721_strain_Arkansas_Vaccine	0.145	0.123	0.095	0.104	0.107	0.119	0.113	0.126	0.098	0.110	0.089	0.095	0.095	0.089	0.083	0.077	0.065	0.064	0.057	0.048		<i>0.001</i>	<i>0.001</i>	<i>0.001</i>
22-FJ888351_strain_H120	0.149	0.140	0.117	0.120	0.127	0.133	0.134	0.142	0.097	0.112	0.088	0.085	0.084	0.068	0.090	0.086	0.091	0.087	0.081	0.077	0.058		<i>0.001</i>	<i>0.002</i>
23-GU393332_serotype_Delaware_072	0.150	0.139	0.117	0.123	0.126	0.137	0.137	0.142	0.117	0.131	0.107	0.126	0.128	0.113	0.119	0.112	0.115	0.114	0.108	0.099	0.077	0.056		<i>0.001</i>
24-GQ504723_strain_Georgia_1998_Vaccine	0.149	0.130	0.101	0.104	0.112	0.125	0.120	0.131	0.117	0.129	0.119	0.120	0.118	0.117	0.111	0.105	0.101	0.101	0.081	0.077	0.058	0.088	0.044	

The number of base substitutions per site from averaging over all sequence pairs are shown. Standard error estimate(s) are shown above the diagonal, in italic font. Estimates of evolutionary distances over sequence pairs between gCoV/FR/2011 and its counterparts are indicated in

bold font. The Analyses were conducted using the Maximum Composite Likelihood model. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). All positions containing gaps and missing data were eliminated.

Figure legends

Figure 1. Phylogenetic analysis of the complete genomes of GFCoV/FR/2011 (in bold font) in relation to all available full genomes of turkey coronaviruses (TCoV) and full genomes of representative infectious bronchitis viruses (IBV) at the nucleotide level. The tree was generated using MEGA 5.05 and the Maximum Likelihood method. Bootstrap values (500 replicates) >75 are indicated on the nodes.

Figure 2. Simplot analysis of full genomic sequence for GFCoV/FR/2011 (query) and its closest TCoV (in blue) and IBV (in red) blast hits. The spike gene area is indicated on the plot.

Figure 1

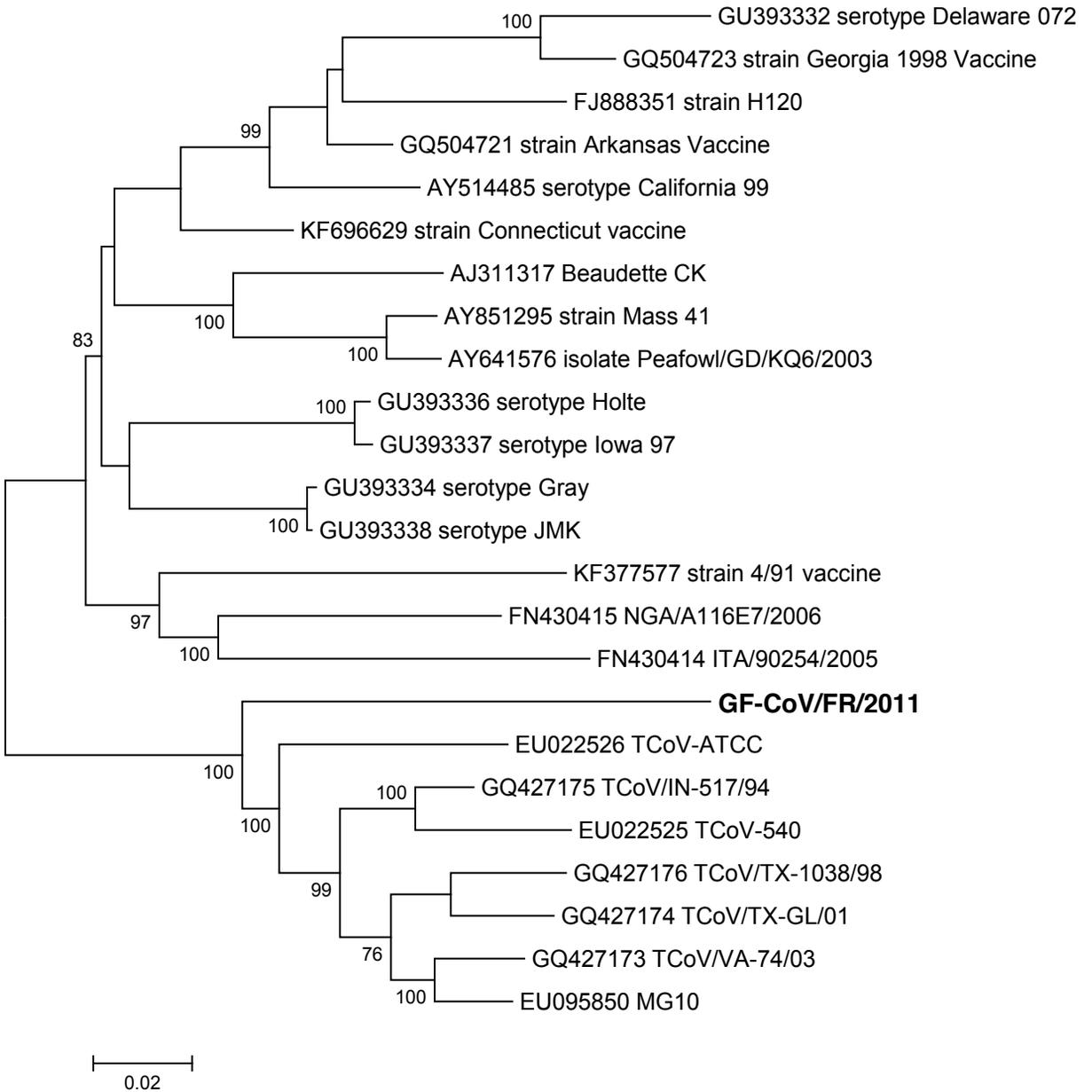
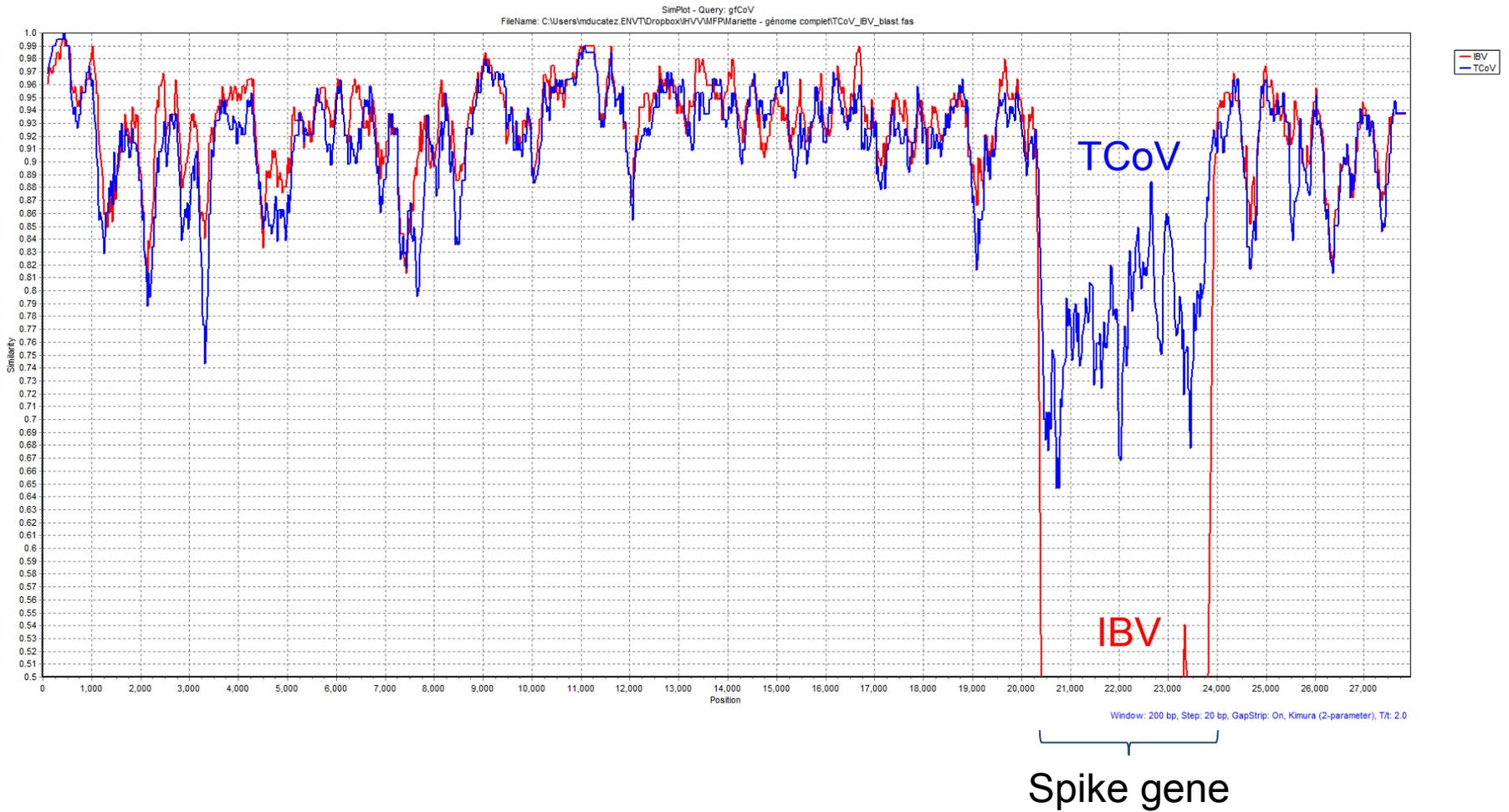


Figure 2



II. Application des approches de séquençage à haut débit sans *a priori* à la détection et caractérisation de virus respiratoires

A. Introduction

Les maladies respiratoires sont parmi les plus répandues dans les élevages avicoles, pour lesquels elles sont des facteurs de pertes économiques importantes. Ces syndromes entraînent également une surconsommation d'antibiotiques. Ils sont le plus souvent multifactoriels et font entrer en jeu des agents infectieux - qu'ils soient viraux, bactériens ou fongiques - et des facteurs environnementaux (température, hygrométrie, poussières). Parmi les espèces aviaires, les oiseaux aquatiques et notamment les canards sont connus pour être les hôtes naturels de virus comme le virus influenza (137,194), ou de bactéries comme *Chlamydia psittaci* (195). Ils sont également hôtes intermédiaires de plusieurs virus capables de franchir la barrière d'espèce tels certains coronavirus (196,197).

Pour appréhender la complexité des agents pathogènes présents chez un hôte, les méthodes de biologie moléculaire classique présentent un inconvénient qui est aussi leur plus gros atout : leur spécificité. En effet, la spécificité des techniques classiques ne leur permet pas de détecter de « nouveaux » agents pathogènes. Les récentes innovations dans les technologies de séquençage mais aussi dans les méthodes d'analyses informatiques ont permis de rendre plus accessible le séquençage à haut débit. Ainsi le séquençage de nouvelle génération (NGS) a été utilisé ces dernières années en virologie vétérinaire, permettant notamment la découverte de nouveaux virus (95,134,147,149,151,152,155,198,199) mais également d'établir le « virome » (représentation de l'ensemble des virus présents dans un échantillon) chez plusieurs espèces telles que la chauve-souris (133), des rongeurs (souris, campagnols, rats) (142) ou encore le porc (140). Ces techniques de séquençage sans *a priori* ont également été employées en médecine humaine, permettant de mettre en évidence de nombreuses espèces virales dans des échantillons de selles de patients atteints de diarrhée (200), lors d'études de métagénomique chez des enfants atteints de paralysie flasque aiguë non polio (201), ou dans l'arbre respiratoire de patients atteints de mucoviscidose (202) ou ayant une infection sévère de l'appareil respiratoire profond sévère (induite par des virus de la famille des *Picornaviridae*, *Orthomyxoviridae* et *Paramyxoviridae* (126)).

Dans notre travail, nous avons cherché à identifier, en tant que preuve de concept, le « virome » respiratoire de canards atteints de syndromes respiratoires et/ou de chutes de ponte. Nous avons plus particulièrement ciblé les virus respiratoires à ARN, compte tenu de leur importance toute particulière en virologie respiratoire.

Les échantillons biologiques soumis à l'analyse étaient des écouvillons trachéaux prélevés sur plusieurs lots de canards présentant des signes cliniques, en particulier des « chutes de ponte ». Ces syndromes affectant les oiseaux reproducteurs correspondent à une atteinte inflammatoire de l'oviducte. Les particularités anatomiques et physiologiques des oiseaux (en particulier la présence de sacs aériens abdominaux qui entourent les organes digestifs et génitaux) expliquent une relation directe entre pathologie respiratoire et génitale (Figure 39) : de fait, les virus associés aux chutes de ponte chez les espèces aviaires sont pour l'essentiel des virus respiratoires (par exemple : bronchite infectieuse, métapneumovirus ou paramyxovirus aviaire de type 1 chez la poule).

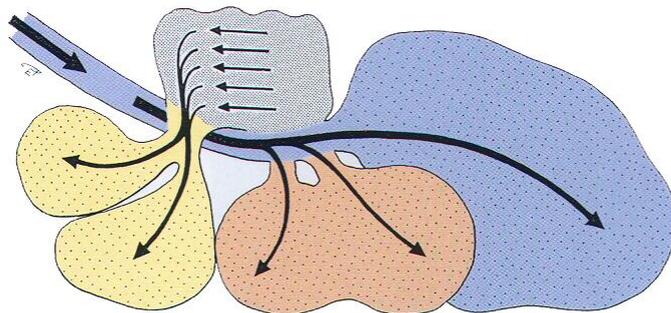


Figure 39 : Les sacs aériens sont une spécificité majeure de l'appareil respiratoire des oiseaux et expliquent les liens entre pathologie respiratoire et génitale.

Des lots de canes reproductrices atteintes de syndrome « chute de ponte », ou de canards mâles épidémiologiquement associés, ont été inclus dans cette étude : le travail a donc été basé sur un recrutement orienté sur critère syndromique.

Les échantillons ont ensuite été traités de manière très semblable à la description présentée précédemment pour l'identification du coronavirus de la pintade :

Les échantillons (écouvillons trachéaux) ont été repris dans une solution de tampon phosphate salin (PBS) puis les particules virales ont été concentrées par ultracentrifugation d'une part et traitement nucléase d'autre part (RNase et DNase). Les acides nucléiques ont été extraits et amplifiés par transcriptase inverse et PCR aléatoire (les amorces aléatoires portent une étiquette simplifiant l'identification de chaque échantillon après l'étape de séquençage). Les bibliothèques ont été créées en suivant les indications constructeur pour le

séquençage illumina sur la machine MiSeq. Nous avons choisi de ne cibler que les séquences de virus ARN suite à l'identification d'une optimisation nécessaire dans le protocole. En effet, une étape de Klenow est nécessaire pour les ADN viraux (permettant la création d'un double brin d'ADN à partir d'une molécule d'ADN simple brin). Cette étape a fait l'objet d'un ajout dans le protocole utilisé pour ce projet de thèse pour les expériences suivantes.

Un séquençage massif a ensuite été réalisé avec le séquenceur Illumina MiSeq et a permis de générer près de 2 millions de lectures pour un ensemble de 11 échantillons, soit en moyenne 200 000 lectures par point d'analyse. Ces séquences ont été soumises à une analyse bioinformatique à l'aide de l'algorithme GAAS, comme évoqué dans l'introduction de la thèse.

Une analyse bio-informatique délicate

Le pourcentage de séquences potentiellement virales s'est avéré faible (x 0,57 % des séquences total générées en moyenne), mais le répertoire de virus potentiellement identifiés est pourtant considérable.

Des séquences correspondant potentiellement à de nombreuses familles virales ont été identifiées lors de l'analyse bioinformatique initiale. Par exemple, un très grand nombre de séquences apparentées aux herpesvirus ont été mises en évidence. Pourtant, quand ces séquences ont été alignées sur un génome de référence (le plus proche après comparaison avec les bases de données), la visualisation de cet alignement grâce à un logiciel de visualisation Integrative Genome Viewer (IGV, (203)) a montré que la totalité de nos séquences virales s'alignaient sur un même gène, codant pour la télomérase. La similarité entre télomérases cellulaires et virales explique cette confusion : ces séquences de télomérase cellulaire étant très proches des séquences virales, l'algorithme n'a pu discriminer et a donc considéré à tort ces séquences comme virales. Cette difficulté est commune à tous les virus à ADN (herpesvirus et poxvirus notamment), pour lesquels la coévolution hôte-virus a entraîné l'assimilation de gènes d'origine cellulaire dans le génome viral : cela signifie que la « détection » de séquences attribuées à ces familles devra toujours être soumise à des vérifications approfondies.

Parmi les familles virales pertinentes en virologie respiratoire, nous avons focalisé notre analyse sur les *Paramyxoviridae* et *Coronaviridae*, car nous avons détecté des séquences apparentées à ces familles dans plusieurs échantillons :

Métapneumovirus aviaires

À partir de séquences de *Paramyxoviridae* générées par le séquençage haut débit, le gène de la glycoprotéine d'attachement (gène G) de métapneumovirus a été complété par séquençage Sanger. Cette séquence a rendu possible la caractérisation de ce virus par une étude phylogénétique. La présence de ce virus dans les élevages peut à lui seul expliquer les signes cliniques observés. Cependant, tous les échantillons étudiés n'ont pas révélé la présence de séquence de métapneumovirus. On peut donc dire que ces métapneumovirus n'expliquent pas l'essentiel des cas cliniques inclus dans cette étude.

Les métapneumovirus aviaires (aMPV) sont des virus responsables d'infections du système respiratoire supérieur. Ces virus font partie de la famille des *Paramyxoviridae* et du genre *Metapneumovirus*. Les *Paramyxoviridae* sont connus pour être responsables de maladies extrêmement variées tant en médecine humaine que vétérinaire. Les métapneumovirus aviaires comme les membres de la sous-famille des *Pneumovirinae* (dont ils font partie) sont des virus non segmentés, composés d'un génome ARN simple brin de polarité négative de 15Kb en moyenne (204). Les métapneumovirus aviaires sont classifiés en quatre sous-types A, B, C et D. Cette classification est basée sur leur différence nucléique du gène « G » codant pour la glycoprotéine. On retrouve les sous-types A et B principalement en Europe et en Asie. Le sous-type C est quant à lui retrouvé aux USA et en France. Le sous-type D a été rapporté uniquement en France (205). L'organisation génomique des métapneumovirus aviaire est la suivante :

3' – (région non codante) – N – P – M – F – M2 – SH – G – L – (région non codante) – 5' (206).
N : gène codant pour la nucléocapside d'enveloppe ; P : gène codant pour la phosphoprotéine (cofacteur du gène L intervenant dans la capsid virale) ; M : gène codant pour la protéine de matrice ; F : protéine de fusion ; M2 : gène codant pour la protéine M2 (jouant un rôle dans la transcription et le passage d'un cycle répliatif à l'assemblage du virus) ; SH : gène codant pour la petite protéine hydrophobe ; G : gène codant pour la glycoprotéine d'attachement ; L : gène codant pour la « grande » polymérase (*Large polymease protein*) (Figure 40).

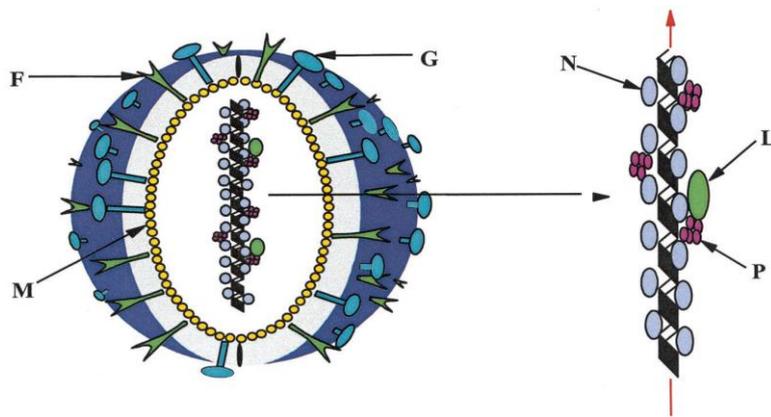


Figure 40 : Schématisation d'un pneumovirus. Le génome ARN est au centre de la capsidie formé de l'association des protéines virales.
Easton et al. 2004

Chez la dinde, l'infection à aMPVs prend la forme de la rhinotrachéite de la dinde. Parmi les signes cliniques observés chez la dinde on retrouve la toux, les écoulements nasaux et oculaires ou encore des conjonctivites. Les oiseaux de tous âges peuvent être touchés. Les aPMVs aviaires peuvent également infecter le poulet (207,208), le canard (209) mais également des espèces plus « exotiques » telles que l'Autruche (*struthio camelus*) (210) ou le faisan (211). Ce virus est un agent pathogène respiratoire très important des espèces aviaires et est retrouvé partout à travers la planète. Les métapneumovirus humains, responsables de syndromes respiratoires chez l'enfant, sont génétiquement plus proches des aMPV de sous-type C que des aMPV de sous-type A et B (207,209) (Figure 41).

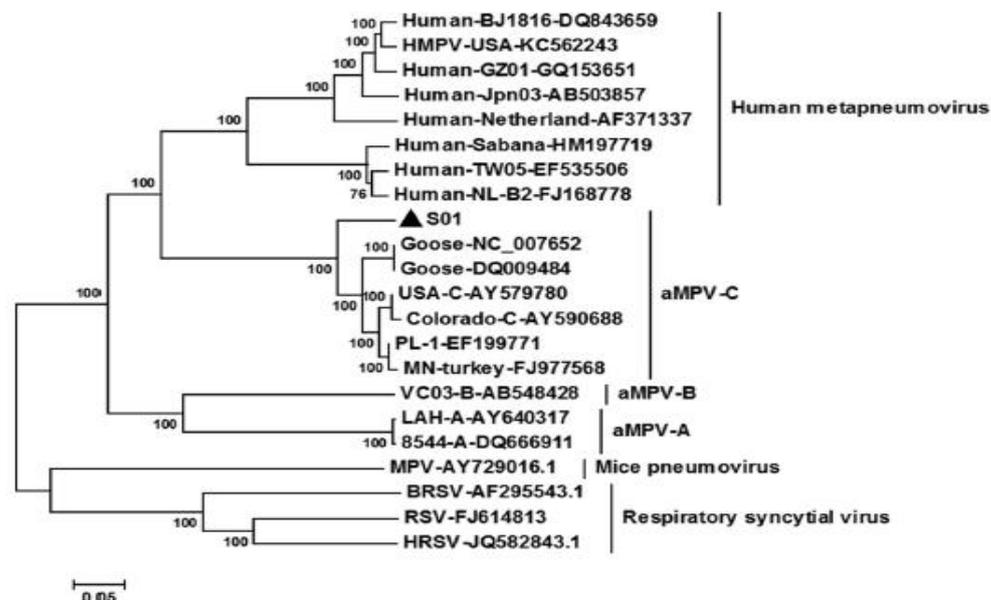


Figure 41 a: Relation phylogénétique des métapneumovirus basé sur le génome complet. Cet arbre a été généré par la méthode des voisins les plus proches (neighbor-joining) avec le logiciel MEGA 5.1, les valeurs de bootstraps sur les nœuds ont été estimées à partir de 1000 répétitions.
D'après *Sun et al. 2014*.

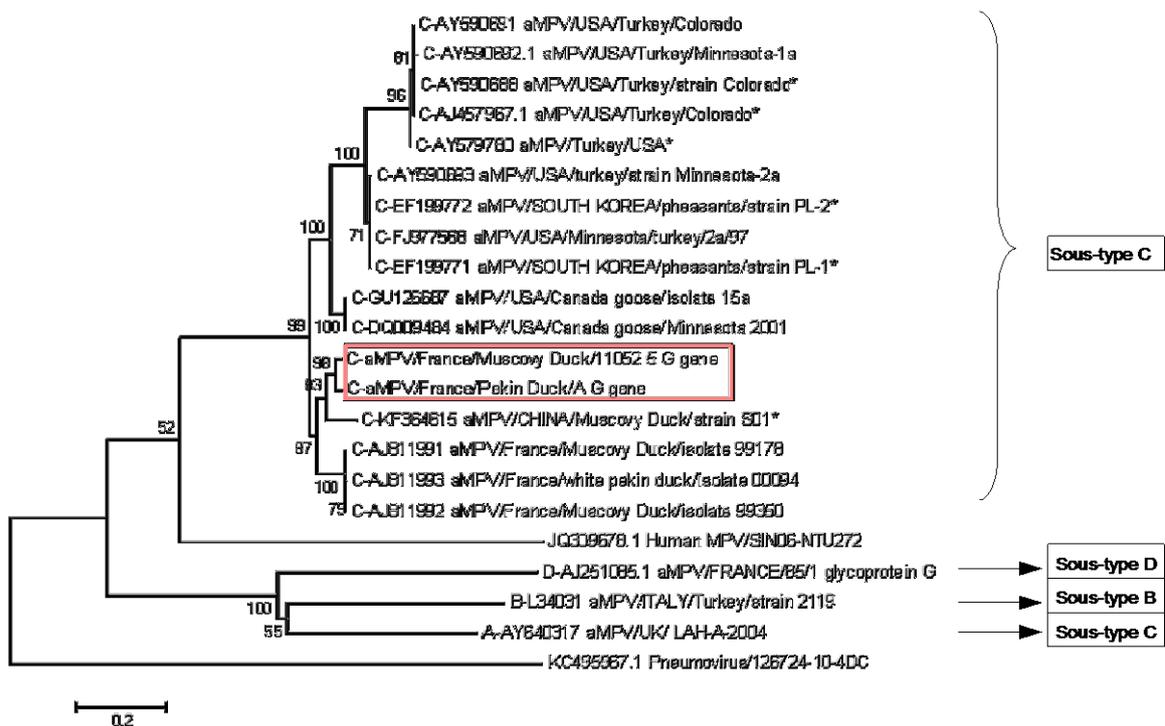


Figure 41 b: Analyse phylogénétique des métapneumovirus aviaires par la méthode du maximum de vraisemblance basée sur l'alignement du gène G des différents sous-types. L'arbre est représenté à l'échelle, avec la longueur des branches proportionnelle au nombre de substitutions par site. Les séquences encadrées proviennent de nos échantillons. L'arbre a été enraciné en utilisant la séquence Pneumovirus/126724-10-4DC comme groupe externe. Les valeurs de « bootstrap » indiquées à côté des nœuds ont été estimées à partir de 1000 répétitions.

La famille *Coronaviridae* renferme un grand nombre de virus pouvant expliquer les signes cliniques observés. Néanmoins tout comme pour les métapneumovirus, seuls quelques échantillons ont révélé la présence de coronavirus aviaire. Le nombre de séquences correspondant au genre *Coronavirus* n'a pas permis sa caractérisation. De plus l'amplification du gène de spicule, gène utilisé pour l'analyse phylogénétique, par méthode de biologie moléculaire classique (PCRs) est restée infructueuse. Ce résultat peut s'expliquer par une sensibilité plus grande du séquençage massif.

Nous avons observé que certains lots d'oiseaux présentaient une co-infection *Paramyxoviridae/Coronaviridae*. En effet, 3 pools présentaient des séquences de coronavirus et de métapneumovirus. Ce genre de co-infection a déjà été documenté notamment chez des enfants présentant des signes cliniques respiratoires (212,213). Il pourrait donc s'agir d'une co-infection respiratoire « classique ».

Une masse de données qui reste à exploiter

Parmi les séquences générées, une grande proportion est apparentée à des séquences bactériennes. Ces acides nucléiques d'origine bactérienne (essentiellement l'ARN 16S ribosomal) proviennent selon toute vraisemblance des étapes de création de la librairie pour le séquençage (extraction et purification d'ARN). Nous n'avons pas fait d'étude complémentaire sur ces séquences, cependant nous retrouvons des résultats similaires à ce qui est décrit dans des publications de métagénomique (133,134,142,143,214). La majeure partie des séquences ont été associées à la famille des *Enterobacteriaceae*, *Moraxellaceae* ou encore *Streptococcaceae*. Ces bactéries, bien que pathogènes opportunistes, colonisent classiquement la cavité oropharyngienne, ce qui peut expliquer leur détection.

L'application des approches NGS à la sphère respiratoire a permis en première intention de mettre en évidence a minima des coronavirus et métapneumovirus, dont il restera à évaluer la pathogénicité intrinsèque. Ces outils ne sont donc susceptibles de répondre qu'à une partie, préliminaire, des questions que se pose un infectiologue.

L'augmentation des quantités de données générées par les séquenceurs d'une part, et la puissance des algorithmes d'analyse d'autre part, améliorera sans doute très rapidement la pertinence et l'accessibilité de ces outils pour la détection et la caractérisation de virus, chez l'homme comme chez l'animal.

B. Article

Metagenomic analysis of respiratory viruses of duck using unbiased High-throughput sequencing

Etienne Iais^{1,2}, Guillaume Croville^{1,2}, Jerome Mariette³, Christophe Klopp³, Jerome Lluch⁴,
Cecile Donnadieu⁴, Fabien Lagoutte¹, Christelle Camus-Bouclainville^{1,2}, Mariette Ducatez^{1,2}
and Jean-Luc Guérin^{1,2*}

¹ Université de Toulouse, INP, ENVT, 31076 Toulouse, FRANCE

² INRA, UMR 1225, 31076 Toulouse, FRANCE

³ Plateforme bioinformatique Toulouse Midi-Pyrénées, UBIA, INRA, 31076 Castanet Tolosan, FRANCE

⁴ GeT-PlaGe, Genotoul, INRA, 31326, Castanet-Tolosan, FRANCE

Keywords: high-throughput sequencing, virus, metagenome, Duck, respiratory tract

Running title: respiratory virome and microbiome of duck

***Address for correspondence:** Jean-Luc Guérin, INRA, UMR 1225, École Nationale
Vétérinaire de Toulouse, 23, chemin des capelles 31076 Toulouse Cedex 3, France;

e-mail: jl.guerin@envt.fr

Tel : +33 561 193 258

Abstract – 245 words

Respiratory viruses of avian species are poorly investigated to date. In order to identify all the viral pathogens and bacterial co-infections in duck respiratory tract, tracheal swabs were collected on ducks showing respiratory clinical signs and/or egg drop syndromes. High-throughput sequencing was subsequently performed and allowed for the determination of putative viral co-infections, enabling characterization of duck respiratory viral flora.

A total of 11 flocks of Peking ducks or Muscovy ducks were sampled. After ultracentrifugation, clinical specimens were treated with DNase and RNase. RNA was then extracted and cDNA was synthesized using random primers. After the PCR step, 300 pb fragments were purified and loaded on a Miseq sequencing platform (Illumina).

More than 2 million reads (of about 150 nucleotides) were initially generated. The assembled contigs and singlet sequences were compared to Genbank virus database using GAAS tool. In average, 0.71% of our sequence reads matched a viral sequence referenced in the database. We identified a majority of sequences of bacteriophages, animal viruses (*Retroviridae*, *Reoviridae*, *Adenoviridae*, *Poxviridae*, *Picornaviridae* and *Paramyxoviridae*); as well as few sequences of plant and insect viruses. We focused on RNA viruses: interestingly, avian coronavirus (*Coronaviridae*) and avian Metapneumovirus (*Paramyxoviridae*) reads were detected. Altogether, these data resulted in the description of a duck respiratory RNA viral metagenome.

Unbiased high-throughput sequencing is a powerful approach to detect infectious agents in a clinical setting: it increases our understanding of the viral diversity in ducks and highlights the complexity of co-infections in poultry respiratory tract.

1. Introduction

Respiratory diseases are among the most prevalent pathological conditions in farm animals, resulting in poor animal welfare, economic losses, and increased antibiotics consumption. These syndromes are mostly multifactorial, due to complex interactions between pathogens (virus and/or bacteria and/or fungi), non-infectious environmental or farm management factors, and host factors. In poultry, several infectious agents have been isolated from birds showing respiratory signs, but a comprehensive knowledge of microbial flora is definitively lacking. Among poultry species, ducks are known to be natural reservoir of many viruses, such as influenza viruses, or zoonotic bacteria, such as *Chlamydia psittaci* (Laroucau *et al.*, 2009). Indeed, several studies have demonstrated that ducks can shed influenza viruses or paramyxovirus in water ponds or during migrations; this peculiar status of duck may be due to a long-term host-pathogens co-evolution in an aquatic environment (Cherbonnel *et al.*, 2007; Kim *et al.*, 2009; Stanislawek *et al.*, 2002).

To address the identification of duck respiratory viral flora, classical molecular methods have the drawback of their specificity and cannot detect emerging or yet unidentified viruses. Recent improvements in sequencing capacities and bioinformatics tools make now “Next Generation Sequencing” (NGS) accessible to veterinary clinical research, allowing prompt discovery of emerging pathogens (Donaldson *et al.*, 2010; Finkbeiner *et al.*, 2008; Hoffmann *et al.*, 2012; Phan *et al.*, 2011; Shan *et al.*, 2011; Victoria *et al.*, 2009; Willner *et al.*, 2009). NGS was used in human respiratory tract studies as well. Recent studies based on metagenomics, indeed evidenced viruses such as influenza virus (Yu *et al.*, 2014),

coronavirus (**Cotten et al., 2013**), or metapneumovirus in patients suffering from respiratory tract infections (**Lysholm et al., 2012**).

In animals, the implementation of NGS resulted in many new findings, for example in bats with a novel influenza virus (**Tong et al., 2012**), coronavirus (**Cotten et al., 2013**), but also in many other species (**Liais et al., 2014**). In this study, we applied a metagenomic approach to identify viral and bacterial co-infections in the trachea of ducks showing respiratory syndromes. We were able to detect avian viruses, as well as bacteriophages and bacterial 16S RNA sequences. These results illustrate the great diversity of infectious agents co-detected in respiratory syndromes, as detected by an unbiased metagenomic approach.

In this present study, we focused on RNA viruses of duck presenting clinical respiratory signs: swabs were collected in poultry and an unbiased method was performed for next generation sequencing (NGS) analysis. Numerous viral families were detected; after a first analysis of metagenomes, we focused on two viruses of interest in avian pathology: avian metapneumovirus and coronavirus. These two viruses were associated with different pathological pictures.

Avian metapneumovirus (aMPV) is a member of the genus *Metapneumovirus* in the family *Paramyxoviridae*, and is agent of upper respiratory tract infections, *e.* Members of *Paramyxoviridae* are known for causing respiratory diseases in human and animal. Metapneumovirus is an unsegmented virus composed of single-stranded negative sense RNA of approximately 15Kb (**Broor and Bharaj, 2007**).

Coronaviruses (CoV), members of the *Coronaviridae* family, are enveloped ARN virus associated with diseases in humans and animals. Classically associated with respiratory or digestive conditions, these viruses can also be responsible for nervous, hepatic or urogenital diseases. Interest in coronaviruses has increase since the identification of a new coronavirus responsible for severe acute respiratory syndrome (SARS) (**Ksiazek et al., 2003; Peiris et al.,**

2003) and more recently, with the emergence of the agent of middle east respiratory syndrome (MERS) (Cotten et al., 2013).

1. Material and Methods

1.1. Clinical examination and sampling of birds.

All birds selected for this study had showed respiratory signs or/and dropping eggs production. Eleven flocks were sampled from Muscovy and Pekin duck. Ten samples (tracheal swabs) were collected per flock and pooled so one “pool” corresponded to one flock.

1.2. Viral nucleic acid concentration

The preparation of specimens for deep sequencing was performed as previously described (Phan et al., 2011; Shan et al., 2011; Victoria et al., 2009). Briefly, all the swabs received at the National Veterinary School of Toulouse were first vortexed in a 500 µL PBS solution. After centrifugation (10000g, 30 min, 4°C) clinical specimens were treated with 100 U DNase Rnase Free (Invitrogen, Carlsbad, CA, USA) and 20 µg/ml RNase A (Ambion, Austin, TX, USA).

1.3. Libraries construction and illumina MiSeq High-throughput.

250µl of sample were added to 750µl of TRIzol (Invitrogen ®). After five minutes at room temperature, 200µl Chloroform (Invitrogen ®) were added to the mixture which was then centrifuged (11000g 30min, 4°C). The aqueous phase (containing RNA) was collected. Viral RNA was extracted from 150 µL of sample using a RNA Virus Kit (Macherey-Nagel®,

Düren, Germany) according to the manufacturer's protocol. RNA was eluted into 50 µL RNase-free water. RNA amplification was carried out by RT-PCR using primers with randomized 3' ends and tags on the 5' UTR (**Victoria et al., 2009**). We selected PCR amplicons of around 300 bp according to MiSeq illumine maximal size. Purification of bands was performed using NucleoSpin Extract II according to the manufacturer's instructions (Macherey-Nagel®, Düren, Germany). The DNA was eluted into 50µL Buffer NE (Macherey-Nagel®, Düren, Germany). The sequencing was performed, using the constructors' protocol for the MiSeq sequencer (Illumina®, MiSeq).

1.4.Data assembly and processing

We initially generated 3.17 million reads of about 150 nucleotides for this study. The Illumina filtered 12.30% of the generated reads and 2.7 million reads were finally obtained after removal of host sequence reads using BWA (**Li and Durbin, 2009**). Reads were then demultiplexed and short reads were assembled. The contigs and singlet sequences were compared to Genbank database using GAAS metagenomic tools with E scores $< 10^{-5}$ (**Angly et al., 2009**). Visualisation of results was done using Krona tools (**Ondov et al., 2011**).

2.5. Phylogenetic analysis

Phylogenetics analyses were carried out using the Maximum Likelihood method, General Time Reversible model. The tree with the highest log likelihood (-12646.4196) is shown (Figure 6). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.7998)). Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 1930 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

2. Results

2.1. Global metagenomics overview

Around 1.3 million reads corresponding to 43% of the total reads identified as known sequences and associated with Bacteria, Archea or Viruses genome on database. In average, 0.57% of our sequence reads matched a viral sequence referenced in the database. We identified a majority of sequences of (i) bacteriophages (*Siphoviridae*, *Myoviridae*, *Podoviridae*, *Microviridae*) and (ii) animal viruses (*Retroviridae*, *Reoviridae*, *Adenoviridae*, *Poxviridae*, *Picornaviridae* and *Paramyxoviridae*); and observed a few sequences of (iii) plant viruses (*Phycodnaviridae*, *Begomoviridae*) and (iv) insect viruses (*Polydnaviridae*, *Iridoviridae*, and *Baculoviridae*) (Figure 3).

2.2. Libraries overview

The global reads repartition matching with known genomes is represented in Figure 1. In all samples, we observed mainly bacteria reads, followed by a lower proportion of archaea and viruses reads.

The majority of the known sequences were matched on bacterial genome referenced in the database (Figure 1), indeed only 1 of the 11 samples had less than 50% of total reads associated with bacterial genomes. Within these bacteria reads, we identified many different families, with *Enterobacteriaceae* representing more than 30% of bacteria reads for half of our samples. Other families such as *Moraxellaceae*, *Pseudomonaceae* or *Pasteurellaceae* were also identified (Figure 2)

A significant part of reads were identified as *Baculovirus* or *Herpesvirus* sequences (Figure 3). After further analysis, we concluded that these herpesvirus-like sequences blasted all on a small telomerase sequence, suggesting that these sequences are actually probably cellular and were abusively considered as viral. The important representation of insect and plant viral reads could be explained by the environmental and food of contamination.

Figure 1: Average repartition of all generated sequences.

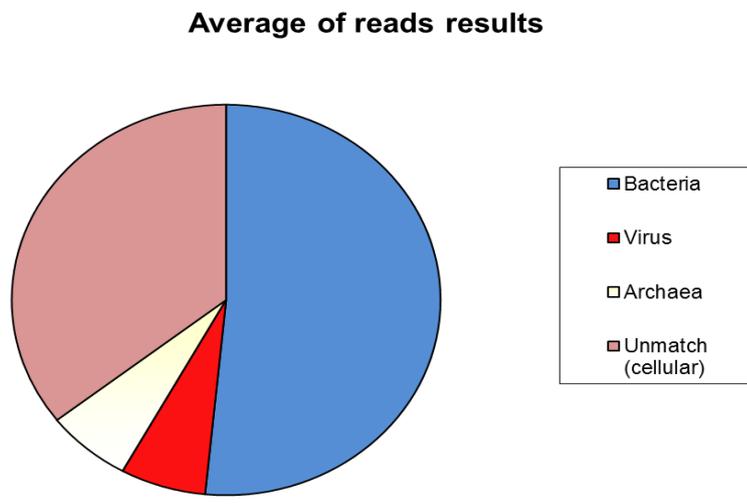


Figure 2: Average repartition of all bacteria sequences.

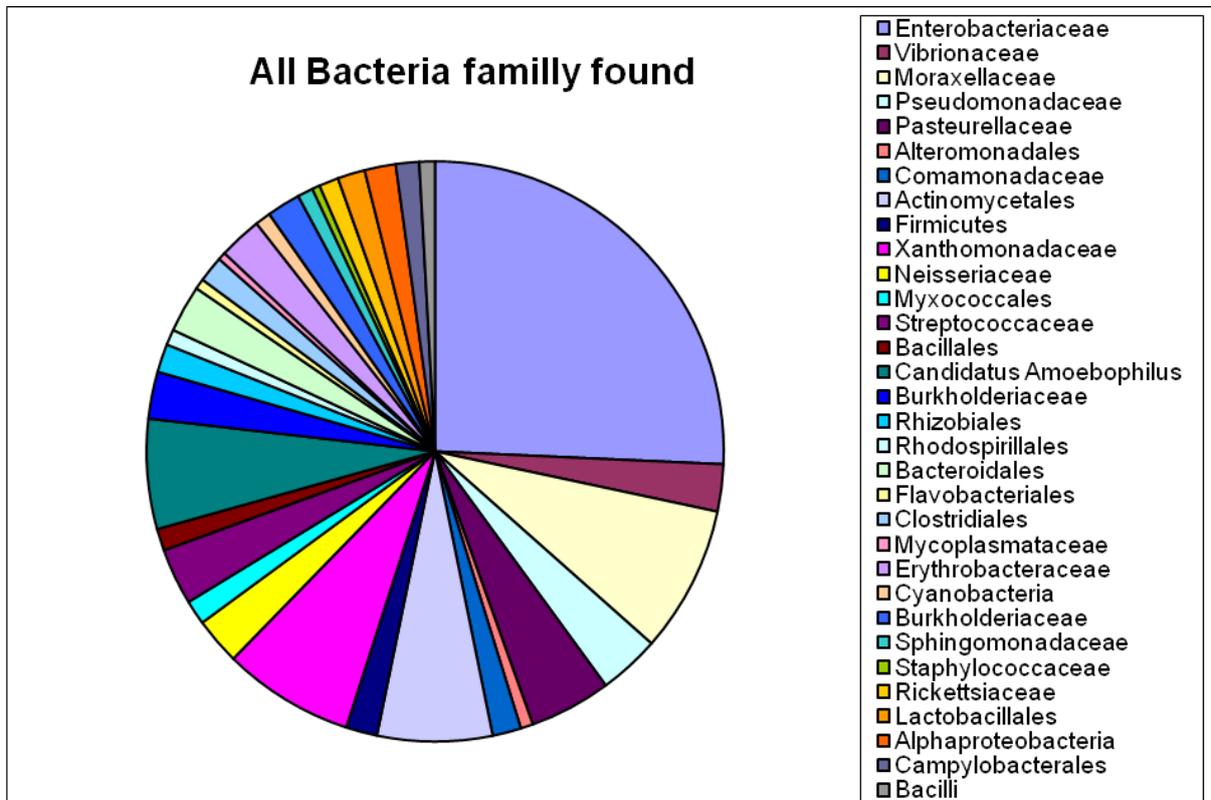
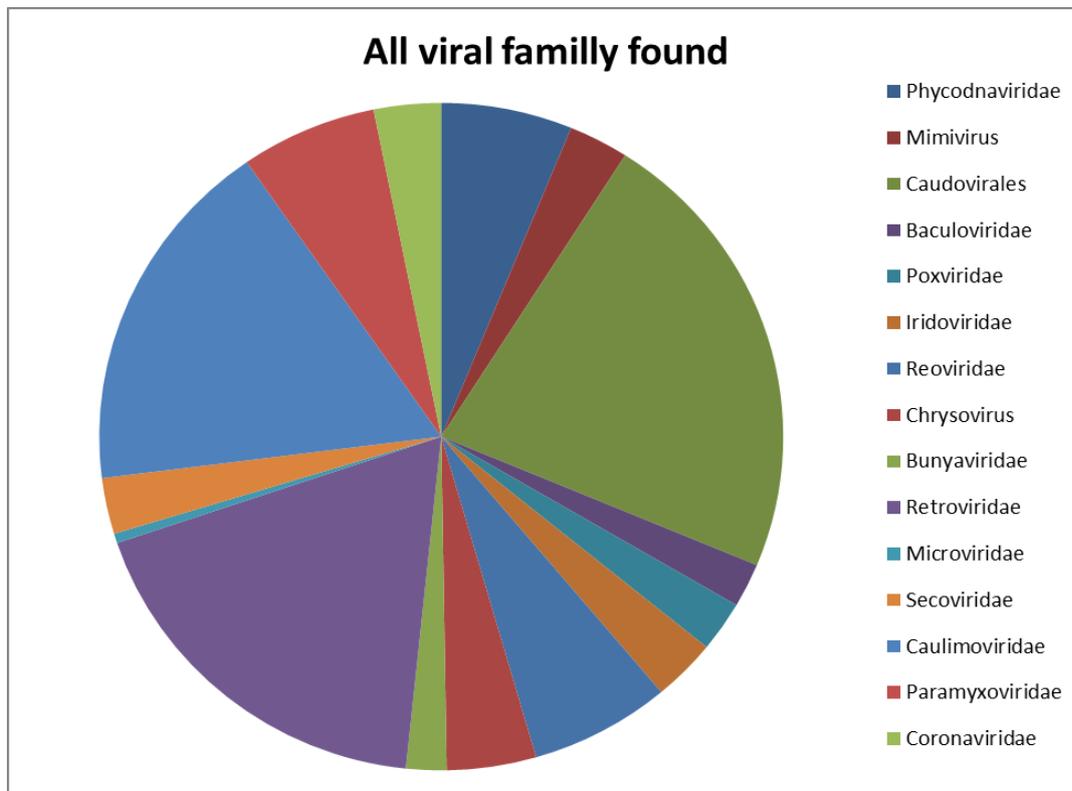


Figure 3. Average repartition of all viral sequences.



2.3.Detection of avian viruses

Within detected the avian viruse: avian metapneumovirus was observed in 5 out of 11 pools, avian coronavirus in 6 out of 11 pools, and avian retrovirus-like in 5 out of 11 pools.

We detected aMPVs in 5 samples pools corresponding to 0.06% to 1.12% of viral reads. But after further analysis, we successfully collected sequences from only 3 of these samples. These sequences were aligned to a reference genome as show on Figure 4. Thanks to Sanger sequencing, we completed the NGS data and obtained the full G gene sequence to performed phylogenetic analysis. These avian metapneumovirus sequences clustered with genotype C viruses (Figure 4). We also found some coronavirus reads but in lower proportion. We recovered coronavirus sequences in 6 samples, representing 0.01% to 1.49% of the viral

reads. We aligned all the coronavirus reads to a reference genome, the recent discovered Duck coronavirus (**Chen *et al.*, 2013**) (Figure 5). We did not succeed in amplifying spike or RdRp gene fragments that would have enabled us to carry out phylogenetic analyses. This may result from a too small amount of viral nucleic acids in the original samples. Interestingly, we found both metapneumovirus and coronavirus in 3 of our samples but with just a few sequence reads.

Figure 4: Map of reads repartition on avian metapneumovirus genome. The full sequence of avian metapneumovirus Colorado strain (AY590688.1) was used as reference genome. The upper part of the figure (labelled “Genome organization”) refers to the metapneumovirus genes. All extracted sequences matching the avian metapneumovirus genome are represented as horizontal grey bands. The “Genome coverage” lines represent the coverage depth at each genome position as vertical bars. Colours indicate differences from the reference strain.



Figure 5: Map of reads repartition on Duck coronavirus genome. The full genome of Duck coronavirus isolate DK/CH/HN/ZZ2004 (JF705860) was used as reference sequence genome. The upper part of the figure maps the duck coronavirus genes. All extracted sequences matching are represented as horizontal grey bands. The “Genome coverage” lines represent the coverage depth at each genome position as vertical bars.

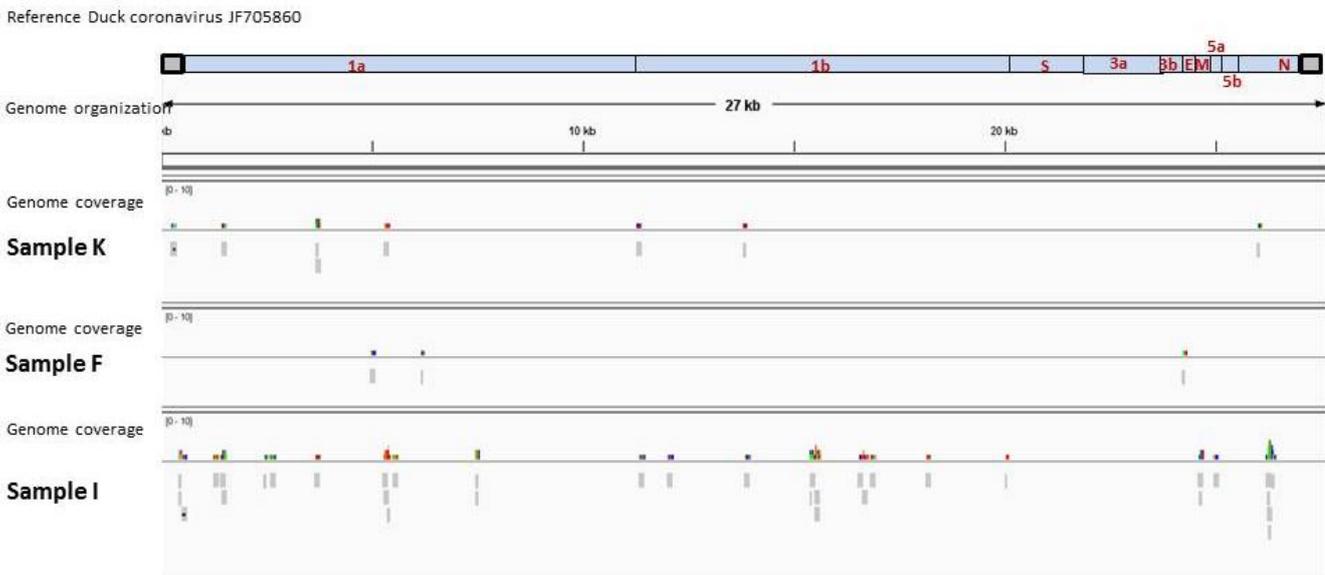
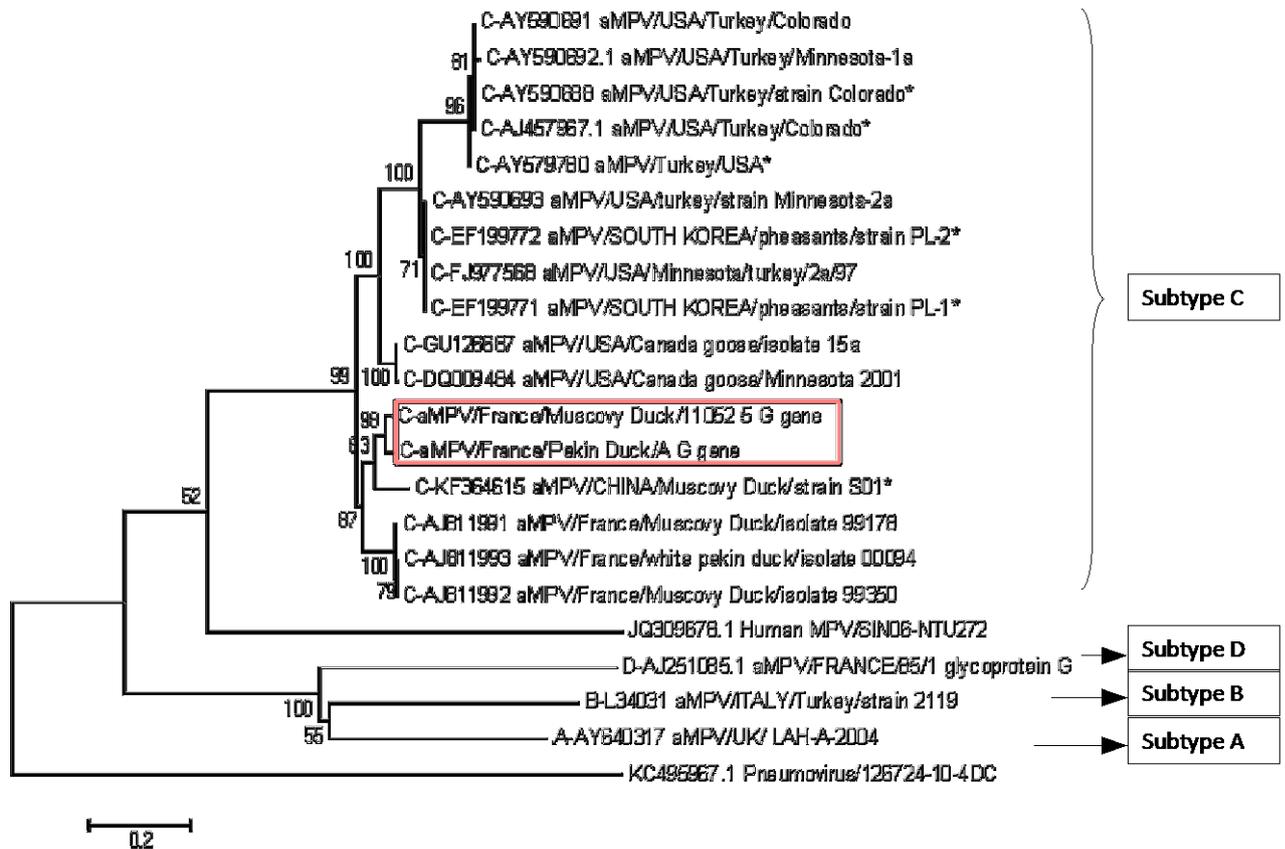


Figure 6: Phylogenetic analysis of the attachment protein gene (G) of avian metapneumoviruses. Our duck aMPV strains are boxed.



2.4. Other viruses

We found a very important part of reads which cluster with insect viruses such as *Baculoviridae* in the Duck respiratory tract. This could be an environmental contamination or coming from the duck feeding. As for the plant viruses this probably came from the poultry food.

3. Discussion

Viruses are the most abundant biological entities present all around the world. While classical methods are poorly efficient to identify these microorganisms, metagenomic approach is a powerful and one of the best ways to identify new viruses and understand their role in our environment (**Edwards and Rohwer, 2005**)

Classical methods of viral detection target known agents and are unsuitable for the identification of new ones. Furthermore, all bacteria and viruses cannot be cultivated in laboratory conditions. During the past few years, the metagenomic approach had an increase of use. Many animals' viromes were identified, but most of them targeted the fecal flora (**Donaldson et al., 2010; Phan et al., 2011; Shan et al., 2011; Wu et al., 2012**), and a few on the respiratory tract (**Wu et al., 2012**). Some other studies on human were published and described respiratory tract (**Greninger et al., 2010; Lysholm et al., 2012; van den Hoogen et al., 2001; Willner et al., 2009; Wootton et al., 2011**), and gastrointestinal tract (**Minot et al., 2011**) viromes.

We successfully obtained virus sequences from the tracheal swabs of ducks showing respiratory signs and/or egg drop syndroms. As previously described in metagenomic analyses (**Donaldson et al., 2010; Lysholm et al., 2012; Phan et al., 2011; Shan et al., 2011; van den Hoogen et al., 2001; Willner et al., 2009; Wu et al., 2012**), we identified known and unknown pathogens (viruses and bacteria). We found in our study some plant and insect viruses which could appear more surprising in the context of respiratory tract. It could be linked to a pollution of swabs by feed material in oropharyngeal cavity during the sampling, or more likely, to air contamination by vegetal debris from straw and/or feeds. Indeed this approach is a very powerful and sensitive diagnosis method so it was very easy to induce some biases while sampling in flock.

Duck respiratory virome showed some similarities with the recently published rodent fecal virome such as plant virus (*Phycodnaviridae*) or other viruses (*Secoviridae* and *Coronavirus*) (**Phan et al., 2011**) and with the human viral microbiome from patients facing respiratory infections as well (**Lysholm et al., 2012; Woo et al., 2005**). Other studies on human respiratory tract evidenced many different viral families, with a major part of bacteriophages (**Willner et al., 2009**). Our results suggested a very important diversity of bacteriophages-like sequences in the trachea. This population may have a major role in immunity and ecological regulation of bacterial population. Due to the close relationship between bacteriophages and bacteria, a viral metagenomic analysis could still give some information on bacterial diversity. Some phage as Chlamydiaphages may have a very important impact on pathogenicity of their host such as *Chlamydia pneumonia* (**Karunakaran et al., 2002**). In other hand, some other Chlamydia such as *Chlamydia caviae* is inhibited by phages infection (**Rank et al., 2009; Śliwa-Dominiak et al., 2013**). In this study, we did not investigated further the bacteriophage populations and decided to focus on viruses putatively associated with pathological

conditions. Further analysis will be necessary to obtain a comprehensive picture of phage populations in the respiratory tract of birds and investigate their functional importance.

Our metagenomic approach can not quantify the microorganisms in the duck respiratory tract, but can only increase our understanding of the microbial population diversity. Our results show the difficulty to appreciate and analyse metagenomic data: indeed, the data obtained after the first estimation of the viral community have to be verified. One of the biggest challenges in NGS studies is the bioinformatics analysis applied to a huge amount of sequence data. The lack of information and standardization of metagenomic approaches and analysis make these analyses still very complicated and time-consuming.

We have conducted a metagenomic analysis of RNA from 10 flocks of Peking ducks or Muscovy ducks. Around 2 million sequence reads were generated, and both contigs and singlet sequences were used for BLAST analysis. Deep sequencing is a powerful approach to detect most infectious agents in a clinical case. It also allows for the detection of the opportunistic pathogen flora, unrelated to clinical signs. This study increases our understanding of the viral diversity in ducks and highlights the complexity of co-infections in poultry respiratory tract. Egg drop and respiratory syndromes observed in duck could therefore be associated with infections with metapneumovirus, identified in the tracheal swabs. Interestingly, we found co-infections in some of our samples with viral pathogens (metapneumovirus and coronavirus).

Acknowledgements

E.L. PhD is supported by a grant from French Ministry of Agriculture and Region Midi-Pyrénées.

References

- Angly, F.E., Willner, D., Prieto-Davó, A., Edwards, R.A., Schmieder, R., Vega-Thurber, R., Antonopoulos, D.A., Barott, K., Cottrell, M.T., Desnues, C., Dinsdale, E.A., Furlan, M., Haynes, M., Henn, M.R., Hu, Y., Kirchman, D.L., McDole, T., McPherson, J.D., Meyer, F., Miller, R.M., Mundt, E., Naviaux, R.K., Rodriguez-Mueller, B., Stevens, R., Wegley, L., Zhang, L., Zhu, B., Rohwer, F., 2009. The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes. *PLoS Computational Biology* 5, e1000593. doi:10.1371/journal.pcbi.1000593
- Broor, S., Bharaj, P., 2007. Avian and Human Metapneumovirus. *Annals of the New York Academy of Sciences* 1102, 66–85. doi:10.1196/annals.1408.005
- Chen, G.-Q., Zhuang, Q.-Y., Wang, K.-C., Liu, S., Shao, J.-Z., Jiang, W.-M., Hou, G.-Y., Li, J.-P., Yu, J.-M., Li, Y.-P., Chen, J.-M., 2013. Identification and Survey of a Novel Avian Coronavirus in Ducks. *PLoS ONE* 8, e72918. doi:10.1371/journal.pone.0072918
- Cherbonnel, M., Lamandé, J., Allée, C., Schmitz, A., Ogor, K., Le Gall-Reculé, G., Le Bras, M.O., Guillemoto, C., Pierre, I., Picault, J.P., Jestin, V., 2007. Virologic findings in selected free-range mule duck farms at high risk for avian influenza infection. *Avian Dis.* 51, 408–413.
- Cotten, M., Lam, T.T., Watson, S.J., Palser, A.L., Petrova, V., Grant, P., Pybus, O.G., Rambaut, A., Guan, Y., Pillay, D., Kellam, P., Nastouli, E., 2013. Full-Genome Deep Sequencing and Phylogenetic Analysis of Novel Human Betacoronavirus. *Emerg Infect Dis* 19, 736–742. doi:10.3201/eid1905.130057
- Donaldson, E.F., Haskew, A.N., Gates, J.E., Huynh, J., Moore, C.J., Frieman, M.B., 2010. Metagenomic Analysis of the Viromes of Three North American Bat Species: Viral Diversity among Different Bat Species That Share a Common Habitat. *Journal of Virology* 84, 13004–13018. doi:10.1128/JVI.01255-10
- Edwards, R.A., Rohwer, F., 2005. Viral metagenomics. *Nature Reviews Microbiology* 3, 504–510. doi:10.1038/nrmicro1163

- Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., Wang, D., 2008. Metagenomic Analysis of Human Diarrhea: Viral Detection and Discovery. *PLoS Pathogens* 4, e1000011. doi:10.1371/journal.ppat.1000011
- Greninger, A.L., Chen, E.C., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G., Kim, E., Pillai, D.R., Guyard, C., Mazzulli, T., Isa, P., Arias, C.F., Hackett, J., Schochetman, G., Miller, S., Tang, P., Chiu, C.Y., 2010. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS ONE* 5, e13381. doi:10.1371/journal.pone.0013381
- Hoffmann, B., Scheuch, M., Höper, D., Jungblut, R., Holsteg, M., Schirrmeyer, H., Eschbaumer, M., Goller, K.V., Wernike, K., Fischer, M., Breithaupt, A., Mettenleiter, T.C., Beer, M., 2012. Novel Orthobunyavirus in Cattle, Europe, 2011. *Emerging Infectious Diseases* 18, 469–472. doi:10.3201/eid1803.111905
- Karunakaran, K.P., Blanchard, J.F., Raudonikiene, A., Shen, C., Murdin, A.D., Brunham, R.C., 2002. Molecular Detection and Seroepidemiology of the Chlamydia pneumoniae Bacteriophage (Cpn1). *Journal of Clinical Microbiology* 40, 4010–4014. doi:10.1128/JCM.40.11.4010-4014.2002
- Kim, J.-K., Negovetich, N.J., Forrest, H.L., Webster, R.G., 2009. Ducks: The “Trojan Horses” of H5N1 influenza. *Influenza and Other Respiratory Viruses* 3, 121–128. doi:10.1111/j.1750-2659.2009.00084.x
- Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.-E., Humphrey, C.D., Shieh, W.-J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.-Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., 2003. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *New England Journal of Medicine* 348, 1953–1966. doi:10.1056/NEJMoa030781
- Laroucau, K., de Barbeyrac, B., Vorimore, F., Clerc, M., Bertin, C., Harkinezhad, T., Verminnen, K., Obeniche, F., Capek, I., Bébéar, C., Durand, B., Zanella, G., Vanrompay, D., Garin-Bastuji, B., Sachse, K., 2009. Chlamydial infections in duck farms associated with human cases of psittacosis in France. *Veterinary Microbiology, Special Issue: Chlamydioses* 135, 82–89. doi:10.1016/j.vetmic.2008.09.048
- Liais, E., Croville, G., Mariette, J., Delverdier, M., Lucas, M.-N., Klopp, C., Lluch, J., Donnadieu, C., Guy, J.S., Corrand, L., Ducatez, M.F., Guérin, J.-L., 2014. Novel avian coronavirus and fulminating disease in guinea fowl, France. *Emerging Infect. Dis.* 20, 105–108. doi:10.3201/eid2001.130774
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, A.M., Persson, B., Allander, T., Andersson, B., 2012. Characterization of the Viral Microbiome in Patients with Severe Lower Respiratory Tract Infections, Using Metagenomic Sequencing. *PLoS ONE* 7, e30875. doi:10.1371/journal.pone.0030875
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., Bushman, F.D., 2011. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* 21, 1616–1625. doi:10.1101/gr.122705.111
- Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 385. doi:10.1186/1471-2105-12-385

- Peiris, J., Lai, S., Poon, L., Guan, Y., Yam, L., Lim, W., Nicholls, J., Yee, W., Yan, W., Cheung, M., Cheng, V., Chan, K., Tsang, D., Yung, R., Ng, T., Yuen, K., 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet* 361, 1319–1325. doi:10.1016/S0140-6736(03)13077-2
- Phan, T.G., Kapusinszky, B., Wang, C., Rose, R.K., Lipton, H.L., Delwart, E.L., 2011. The fecal viral flora of wild rodents. *PLoS pathogens* 7, e1002218.
- Rank, R.G., Bowlin, A.K., Cané, S., Shou, H., Liu, Z., Nagarajan, U.M., Bavoil, P.M., 2009. Effect of Chlamydiophage phiCPG1 on the course of conjunctival infection with “Chlamydia caviae” in guinea pigs. *Infect. Immun.* 77, 1216–1221. doi:10.1128/IAI.01109-08
- Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A., Delwart, E., 2011. The Fecal Virome of Pigs on a High-Density Farm. *Journal of Virology* 85, 11697–11708. doi:10.1128/JVI.05217-11
- Śliwa-Dominiak, J., Suszyńska, E., Pawlikowska, M., Deptuła, W., 2013. Chlamydia bacteriophages. *Arch. Microbiol.* 195, 765–771. doi:10.1007/s00203-013-0912-8
- Stanislawek, W.L., Wilks, C.R., Meers, J., Horner, G.W., Alexander, D.J., Manvell, R.J., Kattenbelt, J.A., Gould, A.R., 2002. Avian paramyxoviruses and influenza viruses isolated from mallard ducks (*Anas platyrhynchos*) in New Zealand. *Arch. Virol.* 147, 1287–1302. doi:10.1007/s00705-002-0818-2
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28, 2731–2739. doi:10.1093/molbev/msr121
- Tong, S., Li, Y., Rivaller, P., Conrardy, C., Castillo, D.A.A., Chen, L.-M., Recuenco, S., Ellison, J.A., Davis, C.T., York, I.A., Turmelle, A.S., Moran, D., Rogers, S., Shi, M., Tao, Y., Weil, M.R., Tang, K., Rowe, L.A., Sammons, S., Xu, X., Frace, M., Lindblade, K.A., Cox, N.J., Anderson, L.J., Rupprecht, C.E., Donis, R.O., 2012. A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci U S A* 109, 4269–4274. doi:10.1073/pnas.1116200109
- Van den Hoogen, B.G., de Jong, J.C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R.A., Osterhaus, A.D., 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* 7, 719–724. doi:10.1038/89098
- Victoria, J.G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S., Delwart, E., 2009. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. *Journal of Virology* 83, 4642–4651. doi:10.1128/JVI.02301-08
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., Rohwer, F., 2009. Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals. *PLoS ONE* 4, e7370. doi:10.1371/journal.pone.0007370
- Woo, P.C.Y., Lau, S.K.P., Chu, C., Chan, K., Tsoi, H., Huang, Y., Wong, B.H.L., Poon, R.W.S., Cai, J.J., Luk, W., Poon, L.L.M., Wong, S.S.Y., Guan, Y., Peiris, J.S.M., Yuen, K., 2005. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *Journal of Virology* 79, 884–895. doi:10.1128/JVI.79.2.884-895.2005
- Wootton, S.C., Kim, D.S., Kondoh, Y., Chen, E., Lee, J.S., Song, J.W., Huh, J.W., Taniguchi, H., Chiu, C., Boushey, H., Lancaster, L.H., Wolters, P.J., DeRisi, J., Ganem, D., Collard, H.R., 2011. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 183, 1698–1702. doi:10.1164/rccm.201010-1752OC

- Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., Zhang, J., Dong, J., Sun, L., Du, J., Liu, L., Xue, Y., Wang, J., Yang, F., Zhang, S., Jin, Q., 2012. Virome Analysis for Identification of Novel Mammalian Viruses in Bat Species from Chinese Provinces. *J. Virol.* 86, 10999–11012. doi:10.1128/JVI.01394-12
- Yu, X., Jin, T., Cui, Y., Pu, X., Li, J., Xu, J., Liu, G., Jia, H., Liu, D., Song, S., Yu, Y., Xie, L., Huang, R., Ding, H., Kou, Y., Zhou, Y., Wang, Y., Xu, X., Yin, Y., Wang, J., Guo, C., Yang, X., Hu, L., Wu, X., Wang, H., Liu, J., Zhao, G., Zhou, J., Pan, J., Gao, G.F., Yang, R., Wang, J., 2014. Influenza H7N9 and H9N2 Viruses: Coexistence in Poultry Linked to Human H7N9 Infection and Genome Characteristics. *J. Virol.* 88, 3423–3431. doi:10.1128/JVI.02059-13

Discussion générale

I. Perspectives du diagnostic par séquençage à haut débit

En médecine humaine comme en médecine vétérinaire, le diagnostic est indispensable. L'objectif de ce travail de thèse a été de mettre au point un mode opératoire sans *a priori* permettant de répondre à la recherche d'un pathogène précis d'une part, et permettant de mettre en évidence un ensemble de microorganismes d'autre part. Ce travail a permis de confirmer la puissance du séquençage dernière génération comme outil de diagnostic, mais également l'importance de l'analyse complémentaire des métadonnées (données générées par le séquençage). Les études menées ont abouti à l'estimation de l'ensemble des virus ARN lors d'un épisode respiratoire et à l'identification d'un nouveau pathogène responsable de la maladie foudroyante de la pintade. La découverte du coronavirus de la pintade a conduit à des études complémentaires de caractérisation et d'isolement qui seront discutées par la suite.

Le séquençage à très haut débit a connu un engouement très important ces 10 dernières années avec les avancées technologiques qui ont ouvert de nouvelles questions mais aussi de nouvelles opportunités d'analyse. La quantité de séquences générées permet désormais de parler de « profondeur » de lecture (correspondant au nombre de lectures sur une même position). Cette profondeur tolère maintenant les études de sous populations et de recherche de mutations, qui sont très importantes dans la lutte contre les résistances aux vaccins ou antiviraux. Un aspect très important aussi est le coût d'un séquençage. En effet, la quantité de données générées a permis de réduire de façon très importante le prix d'une séquence, rendant très accessible le NGS. Cependant le taux d'erreur d'un séquençage haut débit reste plus élevé que celui d'un séquençage Sanger classique. C'est pourquoi de nombreux programmes sont développés, comme le montre la figure 34 (106), servant à éliminer ces erreurs. L'évolution des méthodes d'analyse et le manque de recul rendent difficile la mise en place de standard. Cette lacune induit forcément des biais dans la comparaison entre les différentes études. Ceci rend l'analyse et la mise au point d'une méthode de routine délicate.

Dans le cadre de ces travaux de thèse, nous avons utilisé un seul et même mode opératoire pour répondre aux deux questions :

- Une recherche ciblant un pathogène en particulier, responsable d'une maladie.

- Une recherche « exhaustive » des virus ARN présents chez des animaux présentant des signes cliniques respiratoires ou des chutes de ponte.

Les données produites pour répondre à ces deux questions doivent être analysées de façon différente. En effet, dans un contexte de recherche d'un agent pathogène responsable de signes cliniques tels que ceux de la maladie foudroyante de la pintade, il est important dans un premier temps de cibler les familles très représentées et ayant un pouvoir pathogène connu. Dans le cadre de la maladie foudroyante de la pintade, nous avons analysé deux échantillons ce qui a permis la sélection des agents présents uniquement dans ces deux points d'analyses. Dans le contexte de recherche d'agent responsable de signe respiratoire, plusieurs agents peuvent en être la cause, il est donc nécessaire d'avoir une vision d'ensemble. Pour cette étude, nous avons choisi de nous focaliser sur les virus à ARN, ce qui peut se discuter au regard de la diversité des virus ADN. Néanmoins, il est important de savoir que le traitement des échantillons post séquençage n'est pas le même. Malgré la présence de séquences correspondant potentiellement à des virus à ADN dans nos échantillons, nous n'avons pas pu confirmer leur identification ; le génome complet du canard n'a été que récemment publié, ce qui rend l'élimination des séquences génomiques plus délicate (227). En outre, certains virus à ADN comme les herpesvirus ont intégré au cours de leur évolution des séquences génomiques de leur hôte au sein de leur propre génome. Cette caractéristique peut induire des erreurs d'estimation des populations virales, surtout si aucune analyse complémentaire n'est réalisée après l'étape de classification des séquences générées.

Pour résumer il est très important de définir plusieurs critères avant de mettre au point une analyse métagénomique.

- La question scientifique à laquelle on souhaite répondre peut nécessiter une analyse ciblée (*avec a priori*) ou aléatoire (*sans a priori*).
- Le traitement des échantillons en amont du séquençage peut lui aussi avoir un impact très important sur les résultats, et donc sur la réponse obtenue. En effet, selon que l'on souhaite obtenir des informations concernant les bactéries, les virus (ARN, ADN ou les deux), l'hôte ou l'ensemble des acteurs, alors le traitement des échantillons de base doit être différent.

- La question posée nécessite-t-elle une très grande quantité de séquence ou des séquences de grande taille ? La recherche de mutations d'une sous population ne nécessite pas la même quantité de séquences que la recherche d'un pathogène pour établir un diagnostic.
- La durée de la réaction de séquençage étant variable selon le séquenceur utilisé, il est indispensable de définir si la réponse peut attendre quelques heures, jours ou semaines. En effet, dans un contexte de diagnostic des appareils comme le séquenceur MiSeq (Illumina) ou les séquenceurs Ion Torrent seront privilégiés aux dépens de séquenceurs comme le HiSeq ayant une durée de réaction avoisinant les deux semaines.

II. Perspectives de l'identification du coronavirus de la pintade comme agent étiologique de la maladie foudroyante.

Grâce à cette approche de séquençage massif, nous avons pu identifier, l'agent étiologique de la maladie foudroyante de la pintade. Il s'agit d'un coronavirus aviaire proche du coronavirus de la dinde. Pour la suite des études visant à caractériser ce virus, il serait indispensable de l'isoler pour permettre la reproductibilité des futures études. La culture des coronavirus, comme un grand nombre de virus aviaires, peut être réalisée sur œuf comme sur cellules. Cependant les essais sont jusqu'à présent restés infructueux, tant en ovoculture qu'avec les lignées cellulaires testées. En nous basant sur les études menées sur d'autres coronavirus aviaires, nous savons que les virus de la bronchite infectieuse (IBV) se multiplient dans le liquide allantoïque d'œuf embryonné et que les coronavirus de la dinde se répliquent quant à eux dans le liquide amniotique. Suite à l'échec de ces deux voies d'inoculation sur œuf de poule, afin d'accroître les chances d'isolements, des œufs de pintades devraient être privilégiés. En vue de la contrainte morphologique des œufs de pintades (apparence mouchetée et opaque) et de la similitude génétique avec les TCoV, un isolement par inoculation par voie amniotique dans un œuf de dinde peut être envisagé. En effet, l'étroite proximité génétique entre le gfCoV et le TCoV nous a permis de mettre en évidence la présence du gfCoV au niveau des cellules épithéliales du duodénum de pintade

infectées grâce à un marquage avec des anticorps anti-TCoV (155). Il est donc possible que cette proximité permette aussi au virus de se développer dans un système d'œuf de dinde.

Afin de mieux comprendre le coronavirus de la pintade, une étude de la dose létale 50 (DL50) sera conduite. La DL50 correspond à la dose nécessaire pour obtenir 50% de mortalité lors d'une infection expérimentale. Une étude préliminaire nous a permis d'estimer entre une dilution de 1/100 et 1/1000 la DL50 de notre « stock » de virus. Durant cette expérience, nous avons également réalisé des prélèvements journaliers afin d'apprécier la cinétique d'excrétion virale. Les résultats ont montré un pic d'excrétion dès le deuxième jour ; cependant ces résultats demandent à être confirmés.

Les résultats obtenus semblent démontrer que ce virus entérique est transmis par voie fécale-orale. Lors des reproductions expérimentales en isolateur, nous avons placé deux lots d'oiseaux dans le même isolateur ; un groupe infecté et un groupe sain. La totalité des oiseaux a présenté des signes cliniques dans les 3 jours post infection et sont tous morts en moins d'une semaine. Il est donc probable que le gCoV soit transmis par contact direct ou fécal-oral. Cependant il est possible que ce virus puisse également être transmis par aérosol. Afin de mieux appréhender cette question une nouvelle expérimentation animale devra être conduite. Pour cela deux lots d'animaux seront séparés avec une barrière physique excluant les contacts directs mais permettant les contacts par aérosols.

Les coronavirus sont des virus ARN ayant un fort pouvoir de mutation. Ce caractère partagé avec d'autres virus ARN (comme le virus influenza) lui permet de passer les barrières d'espèce et d'échapper aux défenses immunitaires. Une étude sera conduite afin d'observer l'apparition de mutations lors de passages en série lors d'infections expérimentales en isolateur. Ainsi en séquençant le virus parental et les différents passages nous pourrions peut-être observer des mutations. Il sera crucial que le virus parental ne puisse infecter qu'un seul lot d'oiseaux afin de ne pas introduire un biais dans l'expérience. Afin de comprendre l'origine de ce virus aviaire, une étude complémentaire sera dirigée sur la capacité de ce gCoV à infecter des espèces aviaires d'origine différente (exemple: dinde, canard, poulet).

L'étude du génome complet a mis en évidence une possible recombinaison au niveau du gène de spicule, comme le montre la figure 2 de l'article « *Full genome sequence of French guinea fowl coronavirus associated with fulminating disease* ». Sur cette figure, on ne remarque pas d'importante différence entre le coronavirus de la dinde et le coronavirus de

la bronchite infectieuse, en comparaison avec le coronavirus de la pintade. La distance varie entre 100% et 74% de similitude sur l'ensemble du génome à l'exception d'une partie du génome, correspondant au gène de spicule. En effet, on observe une similitude plus importante entre le gfCoV et le TCov qu'avec l'IBV. Cette recombinaison, déjà observée par Jackwood et ses collaborateurs (183), n'a à l'heure actuelle pas pu être vérifiée. Cependant plusieurs hypothèses sont envisagées : il n'est pas impossible que cette divergence de génome provienne en fait d'une suite de mutations liées à la pression de sélection et l'adaptation à un nouvelle hôte. En effet, une importante diversité moléculaire du gène de spicule est observée au sein d'une même espèce de coronavirus. Une étude de « screening » visant à rechercher de nouveaux coronavirus aviaires pourrait être mis en place afin d'obtenir de nouvelles données. Ces données nous permettraient d'avancer dans la compréhension de cette famille virale et ainsi, de lever le voile sur l'historique de cette recombinaison.

Ces études futures ont pour but de mieux appréhender et éventuellement prédire l'apparition de cas dans les fermes avicoles. De plus, elles permettront de proposer des moyens de prévention. En effet, actuellement la seule conduite à tenir lors d'un diagnostic de maladie foudroyante dans un élevage est l'euthanasie du lot. Il reste donc encore beaucoup à faire sur ce virus tout nouvellement identifié.

III. Le diagnostic à haut débit: quel avenir?

Les agents responsables de troubles respiratoires sont nombreux en médecine vétérinaire, et induisent une perte économique importante. Dans cette étude visant à mettre au point un outil de diagnostic rapide et très efficace, nous nous sommes intéressés aux virus ARN présents lors d'épisode de toux et/ou de chute de ponte. Notre étude nous a permis de mettre en évidence une diversité importante dans des écouvillons trachéaux d'oiseaux présentant des signes cliniques respiratoires. Une telle variété a déjà été observée dans des études similaires chez des patients humains atteints de troubles respiratoires divers (15,17,126). Parmi les familles virales retrouvées l'analyse métagénomique nous a permis d'identifier notamment deux candidats pathogènes: un métapneumovirus aviaire et un coronavirus.

Des études complémentaires nous ont permis de caractériser le métapneumovirus aviaire. Ce virus fait partie des métapneumovirus aviaires de groupe C, groupe classiquement détecté en France et aux États-Unis. L'étude phylogénétique a permis de montrer une différence entre les virus observés dans nos échantillons et les virus français déjà répertoriés. Cette étude a été menée sur le gène de glycoprotéine d'enveloppe (G) responsable de l'attachement du virus. Le but de ce travail étant principalement descriptif le génome n'a pas été entièrement séquencé. De plus, les métapneumovirus sont connus en médecine (humaine et vétérinaire) pour être responsables de troubles respiratoires (236). La présence de ce virus dans nos échantillons est donc cohérente.

Le nombre de séquences associées aux coronavirus retrouvées dans nos échantillons n'a pas permis de réaliser une étude phylogénétique. Nous n'avons pas pu amplifier les ARNs viraux de ce virus par les méthodes de biologie moléculaire classique. Ceci peut être dû à un manque de sensibilité des méthodes traditionnelles, à une mauvaise qualité du matériel génétique de nos échantillons ou à l'incapacité des amorces à se fixer au génome. Cette impossibilité de réaliser une analyse complémentaire sur ces séquences peut soulever l'interrogation sur la présence effective de ce virus. Cependant, le nombre de séquences observées dans les échantillons ne peut pas correspondre à une contamination d'environnement.

Nous avons choisi de ne pas approfondir l'étude sur l'ensemble des virus. Dans le cadre de ce travail de thèse, notre objectif était la mise au point d'une technique et la validation de ce

nouvel outil dans une optique de diagnostic. Nous nous sommes donc concentrés sur les candidats potentiels pouvant répondre à une maladie donnée, tout en gardant une vision globale de la population de virus. Parmi les familles virales identifiées par l'analyse des métadonnées, on peut retrouver un grand nombre de virus d'insectes et de plantes. Ces virus correspondent très vraisemblablement à des pollutions lors de l'échantillonnage, soit par contact de l'écouvillon dans la cavité buccale des canards soit par les poussières ambiantes des élevages. Toutefois, malgré l'origine de ces virus il est clair que cette méthode de diagnostic est capable d'illustrer l'environnement dans lequel évoluent les animaux de cette étude.

Lors de l'analyse de cette étude nous avons ciblé les virus ARN car le protocole d'extraction et d'amplification des acides nucléiques ne permettaient pas d'amplifier de manière optimale les séquences des virus d'ADN. En effet, une étape de « Klenow » aurait dû être utilisée (étape qui a été rajoutée dans le mode opératoire pour les études de métagénomique réalisées au sein du laboratoire). La « Klenow » est un fragment issu de la digestion de l'ADN polymérase I. Son activité enzymatique est comparable à celle d'une ADN polymérase I sauf qu'elle est dépourvue d'activité d'exonucléase 5' → 3' (responsable de la dégradation du brin néosynthétisé). Cette étape aboutit à la création d'un brin complémentaire à partir d'un ADN simple brin. Malgré ce choix initial dans notre protocole, nous avons pu obtenir un grand nombre de séquences potentielles de virus à ADN qui, pour la plupart, ont été associés à la famille des *Herpesviridae*. Nous avons analysé ces séquences d'herpesvirus : après visualisation d'un alignement sur un génome de référence il est apparu que l'ensemble de nos séquences correspondent au gène de la télomérase (cellulaire). Ces séquences ont donc été associées à tort à la famille *Herpesviridae* et sont en réalité des séquences génomiques de l'hôte. Cet exemple d'erreur d'affiliation des séquences générées par le séquençage à haut débit montre l'importance de réaliser une analyse complémentaire des données afin de vérifier la véracité des résultats obtenus. Ceci est particulièrement important pour les virus à ADN et les rétrovirus, les rétrovirus étant connus pour avoir intégré des gènes de leur hôte au cours de leur évolution.

L'analyse des données produites par le séquençage constitue un enjeu crucial pour donner du sens aux séquences générées. En effet, aujourd'hui le verrou majeur de ces analyses est la valorisation des informations. La quantité d'information produite par les machines de dernières générations rend le traitement individuel des séquences inconcevable. Il existe un

grand nombre d'algorithmes, chacun avec des propriétés différentes. Pour cela le nombre d'outils permettant l'analyse des métadonnées a considérablement augmenté. De plus ces algorithmes permettent également de minimiser les potentielles erreurs de séquençage. Cependant la seule répartition des séquences entre les différentes familles n'est pas suffisante. Comme nos résultats ont pu le démontrer, ces algorithmes possèdent encore des lacunes que l'expertise humaine doit combler. Des analyses complémentaires seront donc nécessaires pour donner du sens aux premiers résultats générés par ces algorithmes.

Références

1. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature*. 2004 Jul 8;430(6996):242–9.
2. Koo B-S, Lee H-R, Jeon E-O, Jang H-S, Han M-S, Mo I-P. An unusual case of concomitant infection with chicken astrovirus and group A avian rotavirus in broilers with a history of severe clinical signs. *J Vet Sci*. 2013;14(2):231–3.
3. Mor SK, Sharafeldin TA, Abin M, Kromm M, Porter RE, Goyal SM, et al. The occurrence of enteric viruses in Light Turkey Syndrome. *Avian Pathol J WVPA*. 2013;42(5):497–501.
4. Roussan DA, Shaheen IA, Khawaldeh GY, Totanji WS, Al-Rifai RH. Simultaneous detection of astrovirus, rotavirus, reovirus and adenovirus type I in broiler chicken flocks. *Pol J Vet Sci*. 2012;15(2):337–44.
5. Roussan DA, Shaheen IA, Totanji WS, Khawaldeh GY, Rifai RHA. Simultaneous detection of *Clostridium perfringens* and *Clostridium colinum* by duplex-polymerase chain reaction. *Poult Sci*. 2012 Jan 12;91(12):3080–5.
6. Alnassan AA, Kotsch M, Shehata AA, Krüger M, Dausgshies A, Bangoura B. Necrotic enteritis in chickens: development of a straightforward disease model system. *Vet Rec*. 2014 May 31;174(22):555.
7. Staheli JP, Boyce R, Kovarik D, Rose TM. CODEHOP PCR and CODEHOP PCR primer design. *Methods Mol Biol Clifton NJ*. 2011;687:57–73.
8. Culley AI, Lang AS, Suttle CA. High diversity of unknown picorna-like viruses in the sea. *Nature*. 2003 Aug 28;424(6952):1054–7.
9. Zlateva KT, Crusio KM, Leontovich AM, Lauber C, Claas E, Kravchenko AA, et al. Design and validation of consensus-degenerate hybrid oligonucleotide primers for broad and sensitive detection of corona- and toroviruses. *J Virol Methods*. 2011 Nov;177(2):174–83.
10. Zlateva KT, Coenjaerts FEJ, Crusio KM, Lammens C, Leus F, Viveen M, et al. No novel coronaviruses identified in a large collection of human nasopharyngeal specimens using family-wide CODEHOP-based primers. *Arch Virol*. 2013 Jan 1;158(1):251–5.
11. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. *Nat Genet*. 1999 Jan 1;21:5–9.
12. Wang D, Coscoy L, Zylberberg M, Avila PC, Boushey HA, Ganem D, et al. Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A*. 2002 Nov 26;99(24):15687–92.
13. Chen EC, Miller SA, DeRisi JL, Chiu CY. Using a Pan-Viral Microarray Assay (Virochip) to Screen Clinical Samples for Viral Pathogens. *J Vis Exp [Internet]*. 2011 Apr 27 [cited 2014 Jul 15];(50). Available from: <http://www.jove.com/index/Details.stp?ID=2536>
14. Chiu CY, Rouskin S, Koshy A, Urisman A, Fischer K, Yagi S, et al. Microarray Detection of Human Parainfluenzavirus 4 Infection Associated with Respiratory Failure in an Immunocompetent Adult. *Clin Infect Dis*. 2006 Oct 15;43(8):e71–6.
15. Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, et al. Pan-Viral Screening of Respiratory Tract Infections in Adults With and Without Asthma Reveals Unexpected Human Coronavirus and Human Rhinovirus Diversity. *J Infect Dis*. 2007 Sep 15;196(6):817–25.

16. Chiu CY, Alizadeh AA, Rouskin S, Merker JD, Yeh E, Yagi S, et al. Diagnosis of a Critical Respiratory Illness Caused by Human Metapneumovirus by Use of a Pan-Virus Microarray. *J Clin Microbiol.* 2007 Jul 1;45(7):2340–3.
17. Chiu CY, Urisman A, Greenhow TL, Rouskin S, Yagi S, Schnurr D, et al. Utility of DNA microarrays for detection of viruses in acute respiratory tract infections in children. *J Pediatr.* 2008 Jul;153(1):76–83.
18. Nicholson TL, Kukielka D, Vincent AL, Brockmeier SL, Miller LC, Faaberg KS. Utility of a Panviral Microarray for Detection of Swine Respiratory Viruses in Clinical Samples. *J Clin Microbiol.* 2011 Apr 1;49(4):1542–8.
19. Kruger D, Schneck P, Gelderblom H. Helmut Ruska and the visualisation of viruses. *The Lancet.* 2000 May 13;355(9216):1713–7.
20. Ruska H, Borries B v, Ruska E. Die Bedeutung der Übermikroskopie für die Virusforschung. *Arch Für Gesamte Virusforsch.* 1939 Feb 1;1(1):155–69.
21. Kausche GA, Pfankuch E, Ruska H. Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften.* 1939 May 1;27(18):292–9.
22. Borries B von, Ruska E, Ruska H. Bakterien und Virus in Übermikroskopischer Aufnahme. *Klin Wochenschr.* 1938 Jul 1;17(27):921–5.
23. Luria SE, Anderson TF. The Identification and Characterization of Bacteriophages with the Electron Microscope. *Proc Natl Acad Sci U S A.* 1942 Apr;28(4):127.
24. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. Visualization by Immune Electron Microscopy of a 27-nm Particle Associated with Acute Infectious Nonbacterial Gastroenteritis. *J Virol.* 1972 Nov;10(5):1075.
25. Melnick JL, Phillips CA. Enteroviruses: vaccines, epidemiology, diagnosis, classification. *CRC Crit Rev Clin Lab Sci.* 1970 Jan;1(1):87–118.
26. Welch AB. Purification, Morphology and Partial Characterization of a Reovirus-like Agent Associated with Neonatal Calf Diarrhea. *Can J Comp Med.* 1971 Jul;35(3):195.
27. Gust ID, Kaldor J, Cross GF, Waugh M, Ferris AA. Virus-like particles associated with a faecal antigen from hepatitis patients and with Australia antigen. *Aust J Exp Biol Med Sci.* 1971 Feb;49(1):1–9.
28. Feinstone SM, Kapikian AZ, Purceli RH. Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness. *Science.* 1973 Dec 7;182(4116):1026–8.
29. Dane DS, Cameron CH, Briggs M. VIRUS-LIKE PARTICLES IN SERUM OF PATIENTS WITH AUSTRALIA-ANTIGEN-ASSOCIATED HEPATITIS. *The Lancet.* 1970 Apr 4;295(7649):695–8.
30. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science.* 1983 May 20;220(4599):868–71.
31. Nagler FPO, Rake G. The Use of the Electron Microscope in Diagnosis of Variola, Vaccinia, and Varicella. *J Bacteriol.* 1948 Jan 1;55(1):45–51.

32. Friedlaender M, Moore DH, Love R, Brown RA, Koprowski H. STUDIES WITH THE ELECTRON MICROSCOPE OF VIRUS-HOST RELATIONSHIPS IN EHRlich ASCITES TUMOR CELLS : I. THE IDENTIFICATION AND STRUCTURE OF ANOPHELES A VIRUS. *J Exp Med.* 1955 Sep 30;102(4):361.
33. Friedlaender M, Moore DH, Koprowski H. STUDIES WITH THE ELECTRON MICROSCOPE OF VIRUS-HOST RELATIONSHIPS IN EHRlich ASCITES TUMOR CELLS : II. THE LOCALIZATION AND POSSIBLE DEVELOPMENT OF ANOPHELES A VIRUS WITHIN THE ENDOPLASMIC RETICULUM OF THE HOST CELL. *J Exp Med.* 1955 Sep 30;102(4):371.
34. Almeida JD. A Classification of Virus Particles Based on Morphology. *Can Med Assoc J.* 1963 Oct 19;89(16):787.
35. Almeida JD, Waterson AP. Some implications of a morphologically oriented classification of viruses. *Arch Virol.* 1970;32(1):66–72.
36. CDC. Negative Staining Electron Microscopic Protocol for Rash Illness [Internet]. [cited 2014 Jul 29]. Available from: <http://www.bt.cdc.gov/agent/smallpox/lab-testing/pdf/em-rash-protocol.pdf>
37. Goldsmith CS, Miller SE. Modern Uses of Electron Microscopy for Detection of Viruses. *Clin Microbiol Rev.* 2009 Oct 1;22(4):552–63.
38. Murray K, Selleck P, Hooper P, Hyatt A, Gould A, Gleeson L, et al. A morbillivirus that caused fatal disease in horses and humans. *Science.* 1995;268(5207):94–7.
39. Chua KB, Goh KJ, Wong KT, Kamarulzaman A, Tan PSK, Ksiazek TG, et al. Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *The Lancet.* 1999 Oct 9;354(9186):1257–9.
40. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N Engl J Med.* 2003;348(20):1967–76.
41. Watson JD, Crick FHC. Molecular Structure of Deoxyribose Nucleic Acids [Internet]. 1953 [cited 2014 May 13]. Available from: <http://www.nature.com/nature/dna50/watsoncrick.pdf>
42. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A.* 1977 Feb;74(2):560–4.
43. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975 May 25;94(3):441–8.
44. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977 Dec;74(12):5463–7.
45. Fiers W, Contreras R, Haegemann G, Rogiers R, Van de Voorde A, Van Heuverswyn H, et al. Complete nucleotide sequence of SV40 DNA. *Nature.* 1978 May 11;273(5658):113–20.
46. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage λ DNA. *J Mol Biol.* 1982 Dec 25;162(4):729–73.
47. Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature.* 1984 Jul 19;310(5974):207–11.

48. Bilofsky HS, Burks C, Fickett JW, Goad WB, Lewitter FI, Rindone WP, et al. The GenBank genetic sequence databank. *Nucleic Acids Res.* 1986 Jan 10;14(1):1–4.
49. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A.* 1988 Apr;85(8):2444–8.
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 Oct 5;215(3):403–10.
51. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51 Pt 1:263–73.
52. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature.* 1986 Jun 12;321(6071):674–9.
53. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science.* 1987 Oct 16;238(4825):336–41.
54. Ju J, Ruan C, Fuller CW, Glazer AN, Mathies RA. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc Natl Acad Sci U S A.* 1995 May 9;92(10):4347–51.
55. Metzker ML, Lu J, Gibbs RA. Electrophoretically uniform fluorescent dyes for automated DNA sequencing. *Science.* 1996 Mar 8;271(5254):1420–2.
56. Takahashi S. Multiple Sheath-Flow Gel Capillary-Array Electrophoresis for Multicolor Fluorescent DNA Detection [Internet]. 1994 [cited 2014 Jul 2]. Available from: <http://pubs.acs.org/doi/pdf/10.1021/ac00079a016>
57. Adams MD, Fields C, Venter JC. *Automated DNA Sequencing and Analysis.* Elsevier; 1994. 400 p.
58. Sulston J, Du Z, Thomas K, Wilson R, Hillier L, Staden R, et al. The *C. elegans* genome sequencing project: a beginning. *Nature.* 1992 Mar 5;356(6364):37–41.
59. Ewing B, Green P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res.* 1998 Jan 3;8(3):186–94.
60. Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* 1998 Jan 3;8(3):175–85.
61. Nyrén P, Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem.* 1985 Dec;151(2):504–9.
62. Nyrén P. Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal Biochem.* 1987 Dec;167(2):235–8.
63. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Anal Biochem.* 1996 Nov 1;242(1):84–9.
64. Ronaghi M, Uhlén M, Nyrén P. A Sequencing Method Based on Real-Time Pyrophosphate. *Science.* 1998 Jul 17;281(5375):363–5.

65. Ronaghi M, Nygren M, Lundeberg J, Nyrén P. Analyses of secondary structures in DNA by pyrosequencing. *Anal Biochem.* 1999 Feb 1;267(1):65–71.
66. Alderborn A, Kristofferson A, Hammerling U. Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.* 2000 Aug;10(8):1249–58.
67. Ahmadian A, Lundeberg J, Nyrén P, Uhlén M, Ronaghi M. Analysis of the p53 tumor suppressor gene by pyrosequencing. *BioTechniques.* 2000 Jan;28(1):140–4, 146–7.
68. Garcia CA, Ahmadian A, Gharizadeh B, Lundeberg J, Ronaghi M, Nyrén P. Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene.* 2000 Aug 8;253(2):249–57.
69. Nordström T, Ronaghi M, Forsberg L, de Faire U, Morgenstern R, Nyrén P. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. *Biotechnol Appl Biochem.* 2000 Apr;31 (Pt 2):107–12.
70. Cabal A, Strunk M, Domínguez J, Lezcano MA, Vitoria MA, Ferrero M, et al. Single nucleotide polymorphism (SNP) analysis used for the phylogeny of the Mycobacterium tuberculosis complex based on a pyrosequencing assay. *BMC Microbiol.* 2014;14:21.
71. Lindstrom A, Odeberg J, Albert J. Pyrosequencing for Detection of Lamivudine-Resistant Hepatitis B Virus. *J Clin Microbiol.* 2004 Oct;42(10):4788–95.
72. Gharizadeh B, Kalantari M, Garcia CA, Johansson B, Nyrén P. Typing of human papillomavirus by pyrosequencing. *Lab Investig J Tech Methods Pathol.* 2001 May;81(5):673–9.
73. Honkavuori KS, Shivaprasad HL, Williams BL, Quan P-L, Hornig M, Street C, et al. Novel Borna Virus in Psittacine Birds with Proventricular Dilatation Disease. *Emerg Infect Dis.* 2008 Dec;14(12):1883–6.
74. Eriksson N, Pachter L, Mitsuya Y, Rhee S-Y, Wang C, Gharizadeh B, et al. Viral Population Estimation Using Pyrosequencing. *PLoS Comput Biol* [Internet]. 2008 May [cited 2014 Jul 7];4(5). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323617/>
75. Shcherbik SV, Pearce NC, Levine ML, Klimov AI, Villanueva JM, Bousse TL. Rapid Strategy for Screening by Pyrosequencing of Influenza Virus Reassortants - Candidates for Live Attenuated Vaccines. *PLoS ONE* [Internet]. 2014 Mar 19 [cited 2014 Jul 7];9(3). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3960276/>
76. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* [Internet]. 2005 Jul 31 [cited 2014 Jul 2]; Available from: <http://www.ncbi.nlm.nih.gov.gate1.inist.fr/pmc/articles/PMC1464427/>
77. Bennett S. Solexa Ltd. *Pharmacogenomics.* 2004 Jun;5(4):433–8.
78. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science.* 2009 Feb 1;323(5910):133–8.
79. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature.* 2011 Jul 21;475(7356):348–52.

80. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 2012 May;30(5):434–9.
81. Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, et al. Use of Four Next-Generation Sequencing Platforms to Determine HIV-1 Coreceptor Tropism. Mammano F, editor. *PLoS ONE*. 2012 Nov 14;7(11):e49602.
82. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol*. 2012 Sep;93(Pt 9):1853–68.
83. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clin Microbiol Infect*. 2013 Jan 1;19(1):15–22.
84. Barzon L, Lavezzo E, Costanzi G, Franchin E, Toppo S, Palù G. Next-generation sequencing technologies in diagnostic virology. *J Clin Virol*. 2013 Oct;58(2):346–50.
85. Barba M, Czosnek H, Hadidi A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses*. 2014 Jan 6;6(1):106–36.
86. Gibson RM, Schmotzer CL, Quiñones-Mateu ME. Next-Generation Sequencing to Help Monitor Patients Infected with HIV: Ready for Clinical Use? *Curr Infect Dis Rep*. 2014 Apr;16(4):401.
87. Voelkerding KV, Dames S, Durtschi JD. Next Generation Sequencing for Clinical Diagnostics- Principles and Application to Targeted Resequencing for Hypertrophic Cardiomyopathy. *J Mol Diagn*. 2010 Sep;12(5):539–51.
88. Rusk N. Torrents of sequence. *Nat Methods*. 2011 Jan;8(1):44–44.
89. Fang W, Radovich M, Zheng Y, Fu C-Y, Zhao P, Mao C, et al. “Druggable” alterations detected by Ion Torrent in metastatic colorectal cancer patients. *Oncol Lett*. 2014 Jun;7(6):1761–6.
90. Bai X, Zhang E, Ye H, Nandakumar V, Wang Z, Chen L, et al. PIK3CA and TP53 gene mutations in human breast cancer tumors frequently detected by ion torrent DNA sequencing. *PloS One*. 2014;9(6):e99306.
91. Xu Z, Huo X, Ye H, Tang C, Nandakumar V, Lou F, et al. Genetic mutation analysis of human gastric adenocarcinomas using ion torrent sequencing platform. *PloS One*. 2014;9(7):e100442.
92. Nikiforova MN, Wald AI, Roy S, Durso MB, Nikiforov YE. Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. *J Clin Endocrinol Metab*. 2013 Nov;98(11):E1852–60.
93. Jang JY, Lim HI, Park HW, Choi H-J, Kim T-W, Kang M, et al. Draft Genome Sequence of *Lactobacillus plantarum* wikim18, Isolated from Korean Kimchi. *Genome Announc*. 2014;2(3).
94. Taillebois L, Crook D, Saunders T, Ovenden J. The complete mitochondrial genome of the golden snapper *Lutjanus johnii* (Perciformes: Lutjanidae). *Mitochondrial DNA*. 2014 May 20;1–2.
95. Hause BM, Ducatez M, Collin EA, Ran Z, Liu R, Sheng Z, et al. Isolation of a Novel Swine Influenza Virus from Oklahoma in 2011 Which Is Distantly Related to Human Influenza C Viruses. Fouchier RAM, editor. *PLoS Pathog*. 2013 Feb 7;9(2):e1003176.

96. Hall-Mendelin S, Allcock R, Kresoje N, van den Hurk AF, Warrilow D. Detection of Arboviruses and Other Micro-Organisms in Experimentally Infected Mosquitoes Using Massively Parallel Sequencing. *PLoS ONE*. 2013 Feb 27;8(2):e58026.
97. Luo G, Angelidaki I. Analysis of bacterial communities and bacterial pathogens in a biogas plant by the combination of ethidium monoazide, PCR and Ion Torrent sequencing. *Water Res*. 2014 Sep 1;60:156–63.
98. Roux S, Faubladiere M, Mahul A, Paulhe N, Bernard A, Debroas D, et al. Metavir: a web server dedicated to virome analysis. *Bioinformatics*. 2011 Sep 11;27(21):3074–5.
99. Wommack KE, Bhavsar J, Polson SW, Chen J, Dumas M, Srinivasiah S, et al. VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand Genomic Sci*. 2012 Jul 27;6(3):427–39.
100. Mohammed MH, Ghosh TS, Singh NK, Mande SS. SPHINX—an algorithm for taxonomic binning of metagenomic sequences. *Bioinformatics*. 2011 Jan 1;27(1):22–30.
101. Srinivasan SM, Guda C. MetaID: A novel method for identification and quantification of metagenomic samples. *BMC Genomics*. 2013 Dec 9;14(Suppl 8):S4.
102. Laserson J, Jojic V, Koller D. Genovo: De Novo Assembly for Metagenomes. *J Comput Biol*. 2011 Mar 1;18(3):429–43.
103. Ye Y, Tang H. AN ORFOME ASSEMBLY APPROACH TO METAGENOMICS SEQUENCES ANALYSIS. *J Bioinform Comput Biol*. 2009 Jun;7(3):455–71.
104. Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, et al. The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes. Stormo GD, editor. *PLoS Comput Biol*. 2009 Dec 11;5(12):e1000593.
105. Xia LC, Cram JA, Chen T, Fuhrman JA, Sun F. Accurate Genome Relative Abundance Estimation Based on Shotgun Metagenomic Reads. *PLoS ONE* [Internet]. 2011 Dec 6 [cited 2014 Aug 15];6(12). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232206/>
106. Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research. *Virology*. 2012 Dec;434(2):162–74.
107. Mori A, Deola S, Xumerle L, Mijatovic V, Malerba G, Monsurrò V. Next generation sequencing: new tools in immunology and hematology. *Blood Res*. 2013;48(4):242.
108. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet*. 2010 Oct;42(10):833–9.
109. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011 Feb;43(2):101–8.
110. Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, et al. Complete Khoisan and Bantu genomes from southern Africa. *Nature*. 2010 Feb 18;463(7283):943–7.
111. Chen PE, Willner KM, Butani A, Dorsey S, George M, Stewart A, et al. Rapid Identification of Genetic Modifications in *Bacillus anthracis* Using Whole Genome Draft Sequences Generated by 454 Pyrosequencing. *PLoS ONE* [Internet]. 2010 Aug 25 [cited 2014 Jul 8];5(8). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2928293/>

112. Seth-Smith HMB, Harris SR, Rance R, West AP, Severin JA, Ossewaarde JM, et al. Genome Sequence of the Zoonotic Pathogen *Chlamydophila psittaci*. *J Bacteriol.* 2011 Mar;193(5):1282–3.
113. Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, et al. Identification of a Salmonellosis Outbreak by Means of Molecular Sequencing. *N Engl J Med.* 2011;364(10):981–2.
114. Ren X, Yang F, Hu Y, Zhang T, Liu L, Dong J, et al. Full Genome of Influenza A (H7N9) Virus Derived by Direct Sequencing without Culture. *Emerg Infect Dis [Internet].* 2013 Nov [cited 2014 Oct 13];19(11). Available from: http://wwwnc.cdc.gov/eid/article/19/11/13-0664_article.htm
115. Abolnik C, de Castro M, Rees J. Full genomic sequence of an African avian paramyxovirus type 4 strain isolated from a wild duck. *Virus Genes.* 2012 Dec;45(3):537–41.
116. Han Y, Zhang Y, Mei Y, Wang Y, Liu T, Guan Y, et al. Analysis of hepatitis B virus genotyping and drug resistance gene mutations based on massively parallel sequencing. *J Virol Methods.* 2013 Nov;193(2):341–7.
117. Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, et al. Assessing the Diversity and Specificity of Two Freshwater Viral Communities through Metagenomics. Martin DP, editor. *PLoS ONE.* 2012 Mar 14;7(3):e33641.
118. Wan X-F, Barnett JL, Cunningham F, Chen S, Yang G, Nash S, et al. Detection of African swine fever virus-like sequences in ponds in the Mississippi Delta through metagenomic sequencing. *Virus Genes.* 2013 Jun;46(3):441–6.
119. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome. *ISME J [Internet].* 2014 Aug 5 [cited 2014 Aug 14]; Available from: <http://www.nature.com/gate1.inist.fr/ismej/journal/vaop/ncurrent/full/ismej2014143a.html>
120. Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *Proc Natl Acad Sci.* 2014 Jul 22;111(29):10714–9.
121. Fancello L, Raoult D, Desnues C. Computational tools for viral metagenomics and their application in clinical research. *Virology.* 2012 Dec;434(2):162–74.
122. Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, et al. A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *N Engl J Med.* 2008;358(10):991–8.
123. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, et al. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. *J Virol.* 2009 Feb;83(9):4642–51.
124. Holtz LR, Finkbeiner SR, Kirkwood CD, Wang D. Identification of a novel picornavirus related to cosaviruses in a child with acute diarrhea. *Virol J.* 2008;5:159.
125. Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One.* 2010;5(10):e13381.

126. Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, et al. Characterization of the Viral Microbiome in Patients with Severe Lower Respiratory Tract Infections, Using Metagenomic Sequencing. Highlander SK, editor. PLoS ONE. 2012 Feb 15;7(2):e30875.
127. Abeles SR, Pride DT. Molecular Bases and Role of Viruses in the Human Microbiome. J Mol Biol [Internet]. [cited 2014 Aug 13]; Available from: <http://www.sciencedirect.com/science/article/pii/S0022283614003301>
128. Zhang W, Li L, Deng X, Kapusinszky B, Pesavento PA, Delwart E. THE FECAL VIROME OF CATS IN AN ANIMAL SHELTER. J Gen Virol. 2014 Jul 30;
129. Ng TFF, Mesquita JR, Nascimento MSJ, Kondov NO, Wong W, Reuter G, et al. Feline fecal virome reveals novel and prevalent enteric viruses. Vet Microbiol. 2014 Jun 25;171(1–2):102–11.
130. Li L, Pesavento PA, Shan T, Leutenegger CM, Wang C, Delwart E. Viruses in diarrhoeic dogs include novel kobuviruses and sapoviruses. J Gen Virol. 2011 Jan 11;92(11):2534–41.
131. Bermingham A, Chand MA, Brown CS, Aarons E, Tong C, Langrish C, et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2012;17(40):20290.
132. Lau SKP, Woo PCY, Li KSM, Huang Y, Tsoi H-W, Wong BHL, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci. 2005 Sep 27;102(39):14040–5.
133. Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. Metagenomic Analysis of the Viromes of Three North American Bat Species: Viral Diversity among Different Bat Species That Share a Common Habitat. J Virol. 2010 Oct 6;84(24):13004–18.
134. Baker KS, Leggett RM, Bexfield NH, Alston M, Daly G, Todd S, et al. Metagenomic study of the viruses of African straw-coloured fruit bats: Detection of a chiropteran poxvirus and isolation of a novel adenovirus. Virology. 2013 Jul 5;441(2):95–106.
135. Dacheux L, Cervantes-Gonzalez M, Guigon G, Thiberge J-M, Vandenbogaert M, Maufrais C, et al. A Preliminary Study of Viral Metagenomics of French Bat Species in Contact with Humans: Identification of New Mammalian Viruses. Tse H, editor. PLoS ONE. 2014 Jan 29;9(1):e87194.
136. Nelson MI, Holmes EC. The evolution of epidemic influenza. Nat Rev Genet. 2007 Mar;8(3):196–205.
137. Horimoto T, Kawaoka Y. Influenza: lessons from past pandemics, warnings from current incidents. Nat Rev Microbiol. 2005 Aug;3(8):591–600.
138. Forrest HL, Webster RG. Perspectives on influenza evolution and the role of research. Anim Health Res Rev. 2010 Jun;11(01):3–18.
139. Sachsenröder J, Twardziok SO, Scheuch M, Johne R. The General Composition of the Faecal Virome of Pigs Depends on Age, but Not on Feeding with a Probiotic Bacterium. Meng X-J, editor. PLoS ONE. 2014 Feb 19;9(2):e88888.
140. Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E. The Fecal Virome of Pigs on a High-Density Farm. J Virol. 2011 Sep 7;85(22):11697–708.

141. Lager KM, Ng TF, Bayles DO, Alt DP, Delwart EL, Cheung AK. Diversity of viruses detected by deep sequencing in pigs from a common background. *J Vet Diagn Invest*. 2012 Nov 1;24(6):1177–9.
142. Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. *PLoS Pathog*. 2011;7(9):e1002218.
143. Phan TG, Vo NP, Boros Á, Pankovics P, Reuter G, Li OTW, et al. The Viruses of Wild Pigeon Droppings. Vijaykrishna D, editor. *PLoS ONE*. 2013 Sep 4;8(9):e72787.
144. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol*. 2013 Jun;31(6):533–8.
145. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever-Associated Arenavirus from Southern Africa. *PLoS Pathog* [Internet]. 2009 May [cited 2014 Jul 8];5(5). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680969/>
146. Quan P-L, Junglen S, Tashmukhamedova A, Conlan S, Hutchison SK, Kurth A, et al. Moussa virus: a new member of the Rhabdoviridae family isolated from *Culex decens* mosquitoes in Côte d'Ivoire. *Virus Res*. 2010 Jan;147(1):17–24.
147. Honkavuori KS, Shivaprasad HL, Briese T, Street C, Hirschberg DL, Hutchison SK, et al. Novel Picornavirus in Turkey Poults with Hepatitis, California, USA. *Emerg Infect Dis*. 2011 Mar;17(3):480–7.
148. Buckwalter MR, Nga PT, Gouilh MA, Fiette L, Bureau J-F, Laird ME, et al. Identification of a Novel Neuropathogenic Theiler's Murine Encephalomyelitis Virus[∇]. *J Virol*. 2011 Jul;85(14):6893–905.
149. Kapoor A, Simmonds P, Gerold G, Qaisar N, Jain K, Henriquez JA, et al. Characterization of a canine homolog of hepatitis C virus. *Proc Natl Acad Sci U S A*. 2011 Jul 12;108(28):11608–13.
150. Towner JS, Sealy TK, Khristova ML, Albarino CG, Conlan S, Reeder SA, et al. Newly Discovered Ebola Virus Associated with Hemorrhagic Fever Outbreak in Uganda. *PLoS Pathog* [Internet]. 2008 Nov [cited 2014 Jul 8];4(11). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2581435/>
151. Hoffmann B, Scheuch M, Höper D, Jungblut R, Holsteg M, Schirrmeyer H, et al. Novel Orthobunyavirus in Cattle, Europe, 2011. *Emerg Infect Dis*. 2012 Mar;18(3):469–72.
152. Quan P-L, Firth C, Street C, Henriquez JA, Petrosov A, Tashmukhamedova A, et al. Identification of a Severe Acute Respiratory Syndrome Coronavirus-Like Virus in a Leaf-Nosed Bat in Nigeria. *mBio* [Internet]. 2010 Oct 12 [cited 2014 Jul 8];1(4). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2975989/>
153. Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, et al. Fruit bats as reservoirs of Ebola virus. *Nature*. 2005 Dec 1;438(7068):575–6.
154. Epstein JH, Field HE, Luby S, Pulliam JRC, Daszak P. Nipah virus: Impact, origins, and causes of emergence. *Curr Infect Dis Rep*. 2006 Feb 1;8(1):59–65.

155. Liais E, Croville G, Mariette J, Delverdier M, Lucas M-N, Klopp C, et al. Novel avian coronavirus and fulminating disease in guinea fowl, France. *Emerg Infect Dis*. 2014 Jan;20(1):105–8.
156. Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, Van Blaricum AJ, et al. Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proc Natl Acad Sci U S A*. 2013 Apr 9;110(15):E1407–15.
157. Guérin J-L, Balloy D, Villate DFH. MALADIES DES VOLAILLES. Édition : 3e édition. Paris: France Agricole; 2012. 576 p.
158. Kovács GM, LaPatra SE, D'Halluin JC, Benkó M. Phylogenetic analysis of the hexon and protease genes of a fish adenovirus isolated from white sturgeon (*Acipenser transmontanus*) supports the proposal for a new adenovirus genus. *Virus Res*. 2003 Dec;98(1):27–34.
159. CHERKI S. ETUDE BIBLIOGRAPHIQUE ET EPIDEMIOLOGIQUE PRELIMINAIRE D'UN SYNDROME ENTERIQUE CHEZ LA VOLAILLE APPELLANT A D'ULTERIEURS TRAVAUX DE LABORATOIRE [Internet]. 2014 [cited 2015 Jan 29]. Available from: http://s3.amazonaws.com/academia.edu.documents/35361489/these_final-libre.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1422525992&Signature=NHbDpsMTTjShtioUcl4dqmJBuPw%3D
160. Koci MD, Schultz-Cherry S. Avian astroviruses. *Avian Pathol*. 2002 Jun 1;31(3):213–27.
161. Dodet B, Heseltine E, Mary C, Saliou P. [Rotaviruses in human and veterinary medicine]. *Santé Montrouge Fr*. 1997 Jun;7(3):195–9.
162. Benavente J, Martínez-Costas J. Avian reovirus: Structure and biology. *Virus Res*. 2007 Feb;123(2):105–19.
163. Ganar K, Das M, Sinha S, Kumar S. Newcastle disease virus: Current status and our understanding. *Virus Res*. 2014 May 12;184:71–81.
164. Fleury HJ, Morere G, Demarquez N, Vuillaume A. Unidentified viral particles could be associated with enteritis of various commercial bird species. *Ann Inst Pasteur Virol*. 1988 Dec;139(4):449–53.
165. Brahem A, Demarquez N, Beyrie M, Vuillaume A, Fleury HJ. A highly virulent togavirus-like agent associated with the fulminating disease of guinea fowl. *Avian Dis*. 1992 Mar;36(1):143–8.
166. Andral B, Lagadic M, Louzis C, Guillou JP, Gourreau JM. Fulminating disease in Guinea fowl: etiological investigation. *Point Veterinaire* [Internet]. 1987 Oct [cited 2013 May 22];v. 19(108) p. 515-520. Available from: <http://agris.fao.org/agris-search/search/display.do?f=1988/FR/FR88011.xml;FR880163788>
167. Toffan A, Catania S, Salviato A, De Battisti C, Vascellari M, Toson M, et al. Experimental infection of poults and guinea fowl with genetically distinct avian astroviruses. *Avian Pathol J WVPA*. 2012 Oct;41(5):429–35.
168. González JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol*. 2003 Nov;148(11):2207–35.

169. Gorbalenya AE, Snijder EJ, Spaan WJM. Severe Acute Respiratory Syndrome Coronavirus Phylogeny: toward Consensus. *J Virol*. 2004 Jan 8;78(15):7863–6.
170. Woo PCY, Huang Y, Lau SKP, Yuen K-Y. Coronavirus Genomics and Bioinformatics Analysis. *Viruses*. 2010 Aug 24;2(8):1804–20.
171. Cao J, Wu C-C, Lin TL. Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. *Virus Res*. 2008 Sep;136(1-2):43–9.
172. Dong BQ, Liu W, Fan XH, Vijaykrishna D, Tang XC, Gao F, et al. Detection of a Novel and Highly Divergent Coronavirus from Asian Leopard Cats and Chinese Ferret Badgers in Southern China. *J Virol*. 2007 Jul;81(13):6920–6.
173. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res*. 2006;66:193–292.
174. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N Engl J Med*. 2003 May 15;348(20):1953–66.
175. Peiris J, Lai S, Poon L, Guan Y, Yam L, Lim W, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet*. 2003 Apr;361(9366):1319–25.
176. Weiss SR, Leibowitz JL. Chapter 4 - Coronavirus Pathogenesis. In: Karl Maramorosch AJS and FAM, editor. *Advances in Virus Research* [Internet]. Academic Press; 2011 [cited 2014 Aug 6]. p. 85–164. Available from: <http://www.sciencedirect.com/science/article/pii/B9780123858856000092>
177. Boheemen S van, Graaf M de, Lauber C, Bestebroer TM, Raj VS, Zaki AM, et al. Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans. *mBio* [Internet]. 2012 Dec 31 [cited 2013 May 24];3(6). Available from: <http://mbio.asm.org/content/3/6/e00473-12>
178. Perera RA, Wang P, Gomaa MR, El-Shesheny R, Kandeil A, Bagato O, et al. Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull*. 2013;18(36):pii=20574.
179. Reusken CB, Haagmans BL, Müller MA, Gutierrez C, Godeke G-J, Meyer B, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect Dis*. 2013 Oct;13(10):859–66.
180. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science*. 2003 Oct 10;302(5643):276–8.
181. Chastel C. Le « syndrome respiratoire du Moyen-Orient » (MERS) : qui est responsable, les chauves-souris ou le dromadaire ? *Bull Société Pathol Exot*. 2014 May;107(2):69–73.
182. Jackwood MW, Boynton TO, Hilt DA, McKinley ET, Kissinger JC, Paterson AH, et al. Emergence of a group 3 coronavirus through recombination. *Virology*. 2010 Mar 1;398(1):98–108.
183. Lai MM. Recombination in large RNA viruses: coronaviruses. *seminars in VIROLOGY* [Internet]. Elsevier; 1996 [cited 2015 Jan 30]. p. 381–8. Available from: <http://www.sciencedirect.com/science/article/pii/S1044577396900463>

184. Pillet S. Le coronavirus du SRAS : mosaïque de recombinaisons génétiques j. *Virologie*. 2004 Mar 1;8(2):157–157.
185. Stipp DT, Barry AF, Alfieri AF, Bodnar L, Alfieri AA. Isolation in HRT-18 cells and molecular analysis of a BCoV strain from diarrheic feces of naturally infected calves. *Braz Arch Biol Technol*. 2009 Nov;52(SPE):51–6.
186. Ziegler T, Matikainen S, Rönkkö E, Österlund P, Sillanpää M, Sirén J, et al. Severe Acute Respiratory Syndrome Coronavirus Fails To Activate Cytokine-Mediated Innate Immune Responses in Cultured Human Monocyte-Derived Dendritic Cells. *J Virol*. 2005 Jan 11;79(21):13800–5.
187. Yen Y-T, Liao F, Hsiao C-H, Kao C-L, Chen Y-C, Wu-Hsieh BA. Modeling the Early Events of Severe Acute Respiratory Syndrome Coronavirus Infection In Vitro. *J Virol*. 2006 Mar 15;80(6):2684–93.
188. Yan H, Xiao G, Zhang J, Hu Y, Yuan F, Cole DK, et al. SARS coronavirus induces apoptosis in Vero E6 Cells. *J Med Virol*. 2004 Jul 1;73(3):323–31.
189. Hansa A, Rai RB, Dhama K, Wani MY, Saminathan M, Ranganath GJ. Isolation of bovine coronavirus (bcov) in vero cell line and its confirmation by direct FAT and RT-PCR. *Pak J Biol Sci PJBs*. 2013 Nov 1;16(21):1342–7.
190. L'HARIDON R, Scherrer R, Vautherot JF, La Bonnardière C, Laporte J, Cohen J, et al. Adaptation d'un isolement de coronavirus enterique bovin a la culture cellulaire et caracterisation de la souche obtenue. *Annales de Recherches Vétérinaires* [Internet]. 1981 [cited 2014 Oct 14]. p. 243–51. Available from: <http://hal.archives-ouvertes.fr/docs/00/90/13/31/PDF/hal-00901331.pdf>
191. Lin X, O'Reilly KL, Storz J. Infection of polarized epithelial cells with enteric and respiratory tract bovine coronaviruses and release of virus progeny. *Am J Vet Res*. 1997 Oct;58(10):1120–4.
192. Guy JS. Isolation and propagation of coronaviruses in embryonated eggs. *Methods Mol Biol Clifton NJ*. 2008;454:109–17.
193. Kim J-K, Negovetich NJ, Forrest HL, Webster RG. Ducks: The “Trojan Horses” of H5N1 influenza. *Influenza Other Respir Viruses*. 2009 Jul;3(4):121–8.
194. Laroucau K, de Barbeyrac B, Vorimore F, Clerc M, Bertin C, Harkinezhad T, et al. Chlamydial infections in duck farms associated with human cases of psittacosis in France. *Vet Microbiol*. 2009 Mar 16;135(1–2):82–9.
195. Chu DKW, Leung CYH, Gilbert M, Joyner PH, Ng EM, Tse TM, et al. Avian Coronavirus in Wild Aquatic Birds. *J Virol*. 2011 Dec 1;85(23):12815–20.
196. Chen G-Q, Zhuang Q-Y, Wang K-C, Liu S, Shao J-Z, Jiang W-M, et al. Identification and Survey of a Novel Avian Coronavirus in Ducks. Woo PCY, editor. *PLoS ONE*. 2013 Aug 30;8(8):e72918.
197. Buckwalter MR, Nga PT, Gouilh MA, Fiette L, Bureau J-F, Laird ME, et al. Identification of a Novel Neuropathogenic Theiler's Murine Encephalomyelitis Virus[∇]. *J Virol*. 2011 Jul;85(14):6893–905.

198. Epstein JH, Quan P-L, Briese T, Street C, Jabado O, Conlan S, et al. Identification of GBV-D, a Novel GB-like Flavivirus from Old World Frugivorous Bats (*Pteropus giganteus*) in Bangladesh. *PLoS Pathog* [Internet]. 2010 Jul [cited 2014 Jul 8];6(7). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2895649/>
199. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic Analysis of Human Diarrhea: Viral Detection and Discovery. Holmes EC, editor. *PLoS Pathog*. 2008 Feb 29;4(2):e1000011.
200. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, et al. Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. *J Virol*. 2009 Feb 11;83(9):4642–51.
201. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic Analysis of Respiratory Tract DNA Viral Communities in Cystic Fibrosis and Non-Cystic Fibrosis Individuals. Gold JA, editor. *PLoS ONE*. 2009 Oct 9;4(10):e7370.
202. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011 Jan;29(1):24–6.
203. Broor S, Bharaj P. Avian and Human Metapneumovirus. *Ann N Y Acad Sci*. 2007 Apr 1;1102(1):66–85.
204. Juhasz K, Easton AJ. Extensive Sequence Variation in the Attachment (G) Protein Gene of Avian Pneumovirus: Evidence for Two Distinct Subgroups. *J Gen Virol*. 1994 Jan 11;75(11):2873–80.
205. Easton AJ, Domachowske JB, Rosenberg HF. Animal Pneumoviruses: Molecular Genetics and Pathogenesis. *Clin Microbiol Rev*. 2004 Apr;17(2):390–412.
206. Wei L, Zhu S, Yan X, Wang J, Zhang C, Liu S, et al. Avian Metapneumovirus Subgroup C Infection in Chickens, China. *Emerg Infect Dis*. 2013 Jul;19(7):1092–4.
207. Rivera-Benitez JF, Martínez-Bautista R, Ríos-Cambre F, Ramírez-Mendoza H. Molecular detection and isolation of avian metapneumovirus in Mexico. *Avian Pathol*. 2014 Mar 12;43(3):217–23.
208. Sun S, Chen F, Cao S, Liu J, Lei W, Li G, et al. Isolation and characterization of a subtype C avian metapneumovirus circulating in Muscovy ducks in China. *Vet Res*. 2014 Jul 25;45(1):74.
209. Cadman HF, Kelly PJ, Zhou R, Davelaar F, Mason PR. A serosurvey using enzyme-linked immunosorbent assay for antibodies against poultry pathogens in ostriches (*Struthio camelus*) from Zimbabwe. *Avian Dis*. 1994 Sep;38(3):621–5.
210. Lee E ho, Song M-S, Shin J-Y, Lee Y-M, Kim C-J, Lee YS, et al. Genetic characterization of avian metapneumovirus subtype C isolated from pheasants in a live bird market. *Virus Res*. 2007 Sep;128(1–2):18–25.
211. Martin ET, Fairchok MP, Stednick ZJ, Kuypers J, Englund JA. Epidemiology of Multiple Respiratory Viruses in Childcare Attendees. *J Infect Dis*. 2013 Mar 15;207(6):982–9.
212. Schneider H, Adams O, Weiss C, Merz U, Schrotten H, Tenenbaum T. Clinical characteristics of children with viral single- and co-infections and a petechial rash. *Pediatr Infect Dis J*. 2013 May;32(5):e186–91.

213. Roux S, Krupovic M, Poulet A, Debroas D, Enault F. Evolution and Diversity of the Microviridae Viral Family through a Collection of 81 New Complete Genomes Assembled from Virome Reads. Dutilh BE, editor. PLoS ONE. 2012 Jul;7(7):e40418.
214. Cherbonnel M, Lamandé J, Allée C, Schmitz A, Ogor K, Le Gall-Reculé G, et al. Virologic findings in selected free-range mule duck farms at high risk for avian influenza infection. *Avian Dis.* 2007 Mar;51(1 Suppl):408–13.
215. Kim J-K, Negovetich NJ, Forrest HL, Webster RG. Ducks: The “Trojan Horses” of H5N1 influenza. *Influenza Other Respir Viruses.* 2009 Jul;3(4):121–8.
216. Stanislawek WL, Wilks CR, Meers J, Horner GW, Alexander DJ, Manvell RJ, et al. Avian paramyxoviruses and influenza viruses isolated from mallard ducks (*Anas platyrhynchos*) in New Zealand. *Arch Virol.* 2002 Jul 1;147(7):1287–302.
217. Yu X, Jin T, Cui Y, Pu X, Li J, Xu J, et al. Influenza H7N9 and H9N2 Viruses: Coexistence in Poultry Linked to Human H7N9 Infection and Genome Characteristics. *J Virol.* 2014 Mar 15;88(6):3423–31.
218. Cotten M, Lam TT, Watson SJ, Palser AL, Petrova V, Grant P, et al. Full-Genome Deep Sequencing and Phylogenetic Analysis of Novel Human Betacoronavirus. *Emerg Infect Dis.* 2013 May;19(5):736–42.
219. Tong S, Li Y, Rivallier P, Conrardy C, Castillo DAA, Chen L-M, et al. A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci U S A.* 2012 Mar 13;109(11):4269–74.
220. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N Engl J Med.* 2003;348(20):1953–66.
221. Peiris J, Lai S, Poon L, Guan Y, Yam L, Lim W, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet.* 2003 Apr 19;361(9366):1319–25.
222. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009 Jul 15;25(14):1754–60.
223. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics.* 2011;12:385.
224. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.* 2011 Oct;28(10):2731–9.
225. Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol.* 2005 Jun;3(6):504–10.
226. Wu Z, Ren X, Yang L, Hu Y, Yang J, He G, et al. Virome Analysis for Identification of Novel Mammalian Viruses in Bat Species from Chinese Provinces. *J Virol.* 2012 Oct 15;86(20):10999–1012.
227. Van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA, et al. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med.* 2001 Jun;7(6):719–24.

228. Wootton SC, Kim DS, Kondoh Y, Chen E, Lee JS, Song JW, et al. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011 Jun 15;183(12):1698–702.
229. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*. 2011 Aug 31;21(10):1616–25.
230. Woo PCY, Lau SKP, Chu C, Chan K, Tsoi H, Huang Y, et al. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J Virol*. 2005 Jan;79(2):884–95.
231. Karunakaran KP, Blanchard JF, Raudonikiene A, Shen C, Murdin AD, Brunham RC. Molecular Detection and Seroepidemiology of the Chlamydia pneumoniae Bacteriophage (Cpn1). *J Clin Microbiol*. 2002 Nov 1;40(11):4010–4.
232. Rank RG, Bowlin AK, Cané S, Shou H, Liu Z, Nagarajan UM, et al. Effect of Chlamydiaphage phiCPG1 on the course of conjunctival infection with “Chlamydia caviae” in guinea pigs. *Infect Immun*. 2009 Mar;77(3):1216–21.
233. Śliwa-Dominiak J, Suszyńska E, Pawlikowska M, Deptuła W. Chlamydia bacteriophages. *Arch Microbiol*. 2013 Nov;195(10-11):765–71.
234. Huang Y, Li Y, Burt DW, Chen H, Zhang Y, Qian W, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet*. 2013 Jun 9;45(7):776–83.
235. Broor S, Bharaj P. Avian and Human Metapneumovirus. *Ann N Y Acad Sci*. 2007 Apr 1;1102(1):66–85.

AUTHOR: LIAIS Etienne

TITLE: Identification and characterisation of avian viruses using high throughput sequencing

SUPERVISORS: GUÉRIN Jean-Luc and DUCATEZ Mariette

Abstract

Infectious diseases are considered the most prevalent cause of mortality in humans as well as other animals worldwide. Since the advent of high throughput sequencing technologies, diagnostic methods for these conditions have quickly changed and evolved, as the continuously decreasing cost of mass sequencing is making this tool available to larger numbers of people. As part of my thesis project, an Illumina®-based sequencing method (on a MiSeq machine) was designed for diagnostic purposes in clinical cases in poultry. We first used this method to identify the causative agent of the fulminating disease of guinea fowl. This validated the use of our protocol to identify the pathogenic infectious agent behind a specific condition. This newly identified Coronavirus was further analysed and characterised. In a second study we used an unbiased mass sequencing approach to describe the RNA virus populations present in the duck respiratory tract during clinical episodes (respiratory illness or egg drops). Data showed an important viral diversity and we identified some candidate pathogens.

Taken together, these results validate the use of high throughput sequencing as a powerful diagnostic tool.

KEYS WORDS: high throughput sequencing, diagnostic tools, avian pathology, Coronavirus, Metapneumovirus, bioinformatics.

LABORATORY ADDRESS:

INRA UMR 1225 IHAP-ENVT 23 Chemin des Capelles, 31076 Toulouse cedex

AUTEUR : LIAIS Etienne

TITRE : Identification et caractérisation de virus aviaires par des approches de séquençage haut débit

DIRECTEUR DE THÈSE : GUÉRIN Jean-Luc et DUCATEZ Mariette

LIEU ET DATE DE SOUTENANCE : École National Vétérinaire de Toulouse le ../12/14

RÉSUMÉ en français.

En médecine humaine et vétérinaire, les agents pathogènes représentent la cause de mortalité principale à travers la planète. Les méthodes de diagnostic de ces pathogènes ont considérablement changé et évolué particulièrement depuis l'apparition du séquençage haut débit. Les nouvelles méthodes de séquençage massif ont considérablement diminué le prix d'une séquence permettant de rendre accessible cette technologie révolutionnaire. Dans le cadre de mes travaux de thèse, nous avons mis en place un protocole pour l'utilisation du séquençage Illumina® (avec le séquenceur MiSeq) comme méthode de diagnostic lors de différents cas pathologiques aviaires. L'utilisation de cette méthode nous a permis dans un premier temps d'identifier l'agent étiologique de la maladie foudroyante de la pintade. Cette étude nous a permis de valider l'utilisation de ce genre de méthode pour des cas ciblés, ici lors d'un épisode clinique particulier n'impliquant vraisemblablement qu'un seul candidat pathogène. Ce nouveau coronavirus a fait l'objet d'études complémentaires afin de le caractériser. Nous avons élargis les cibles recherchées en analysant dans un deuxième temps l'ensemble des virus ARN chez le canard lors d'épisodes cliniques respiratoires et/ou de chute de ponte. L'analyse des données a mis en évidence une importante diversité virale et a permis d'identifier des candidats responsables potentiels. L'ensemble des résultats obtenus nous permet de valider l'utilisation du séquençage à haut débit comme un outil puissant de diagnostic.

MOTS CLÉS : Séquençage à haut débit, outil de diagnostic, pathologie aviaire, Coronavirus, Métapneumovirus, bio-informatique.

DISCIPLINE ADMINISTRATIVE :

ED SEVAB : Pathologie, Toxicologie, Génétique et Nutrition

INTITULÉ OU ADRESSE DU LABORATOIRE :

INRA UMR 1225 IHAP-ENVT 23 Chemin des Capelles, 31076 Toulouse cedex