



HAL
open science

**Séquençage et PCR à haut débit : application à la
détection et la caractérisation d'agents pathogènes
respiratoires aviaires et au contrôle de pureté
microbiologique des vaccins**

Guillaume Croville

► **To cite this version:**

Guillaume Croville. Séquençage et PCR à haut débit : application à la détection et la caractérisation d'agents pathogènes respiratoires aviaires et au contrôle de pureté microbiologique des vaccins. Biologie moléculaire. Université Paris sciences et lettres, 2017. Français. NNT : 2017PSLEP028 . tel-02106661

HAL Id: tel-02106661

<https://theses.hal.science/tel-02106661>

Submitted on 23 Apr 2019

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

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

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École Pratique des Hautes Études

Séquençage et PCR à haut débit : Application à la détection et
la caractérisation d'agents pathogènes respiratoires aviaires
et au contrôle de pureté microbiologique des vaccins

École doctorale de l'EPHE – ED 472

Spécialité : BIOLOGIE CELLULAIRE ET MOLÉCULAIRE
ET SCIENCES DE LA SANTÉ

COMPOSITION DU JURY :

M. Jacques IZOPET
Université de Toulouse
Président de jury

Mme Thérèse COMMES
Université Montpellier 2
Rapporteur

M. Étienne THIRY
Université de Liège
Rapporteur

Mme Fanny BEILVERT
Filavie S.A.S
Membre invité

Soutenue par :

Guillaume CROVILLE

le 30 octobre 2017

Dirigée par :

Thierry DUPRESSOIR

Jean-Luc GUÉRIN



École Pratique
des Hautes Études



REMERCIEMENTS

Mes remerciements s'adressent en premier lieu aux membres du jury, merci à Jacques Izopet d'avoir accepté de présider ce jury, merci Thérèse Commes et Etienne Thiry d'avoir accepté de juger ce manuscrit et d'en avoir été les rapporteurs. Merci Fanny Beilvert d'avoir accepté de siéger dans ce jury en qualité de membre invité.

Merci à Thierry Dupressoir, co-directeur de cette thèse, merci de m'avoir accompagné et de me faire confiance depuis 2010, l'année où tout a commencé avec le diplôme EPHE. Je garde un excellent souvenir de ta manière d'enseigner et de ton humour pas toujours simple à capter. Je te souhaite plein de bonnes choses pour la « suite ».

Merci Jean-Luc, mon second co-directeur, professeur presque ordinaire, je ne compterai pas les années car ça me ficherait un coup de vieux. Merci pour tout ce que tu fais et tout ce que tu as fait pour ma vie professionnelle. Longue vie à notre binôme de choc dans les NGS, je ne comprendrai jamais comment tu arrives à mener tant de projets de front tout en restant si pertinent et informé sur chaque sujet. Heureusement que tu es l'incarnation même de l'expression *comme une poule qui a trouvé un couteau* lorsqu'il s'agit d'utiliser un ordinateur et la ligne de commande, ça me permet de me sentir un minimum indispensable dans notre collaboration. Je suis impatient de la suite, cette cellule de diagnostic est une superbe aventure à venir. Et, enfin, bravo pour avoir rentré le million.

Merci à Dominique Fournier, Odile Le Minor et Fanny Beilvert de Filavie. C'est grâce à votre confiance que ce sujet de thèse est né. La CIFRE est terminée mais la collaboration continue, le contrat est rempli !

Merci aux plateformes de génomique et de bioinformatique Génotoul. La qualité de votre travail est sans conteste, c'est une chance de pouvoir travailler avec vous. Mes pensées vont particulièrement vers Catherine pour son travail sur le MinION (à toi de m'apprendre à l'utiliser maintenant !), Sophie pour l'Illumina et Clémence pour le 454 (à l'époque !), Maxime pour la bioinfo MinION. Merci à Christophe et Jérôme, les bioinfos qui nous ont mis le pied à l'étrier il y a bien longtemps.

Le temps est venu de remercier mes collègues du quotidien, la liste est longue et je ne saurais m'attarder sur chacun d'entre vous. Néanmoins, pour des raisons politiques, je me dois de remercier mes « co-bureau » de l'extrême : Mariette et son imitation inimitable du petit cheval

et Christelle, notre Madame Dumas et ses relations obscures avec Eurogentec. Sache que Mariette et moi on sait tout.

Elias, mon pur-sang arabe, futur premier ministre du Liban. Merci pour tous ces moments passés ensemble et ce langage codé que nous avons développé pendant tout ce temps, on est connecté. Tu vas beaucoup nous manquer à Pipou et moi. Habibi khouya.

Une pensée pour Guillaume, Romain et Florian, mes compères de grimpe et de trails nocturnes. Pour Stéphane notre fournisseur de cafetières et Gillou notre râleur hors-classe. Pour Sylvie parce que je te fais toujours marcher, Brigitte parce que FO c'est la vie et Josy mon ambulancière, pour Kunta et son équipe de petits chinois. Une pensée pour Cécile et notre trafic botanique, Marie la maman de Sirocco le kakapo, Julien et Hadrien nos docteurs ès troubleshooting, Angélique l'artiste-pâtissière, Thomas et son café du matin, Justine et son chat-drone, Charlotte et ses 800 cookies (et les plaques biomark, accessoirement), Jessica et ses amis les escargots, Mattias et son scooby doo.

Une pensée pour les autres doctorants ou ex-doctorants de l'école : Ahmad, Yoann, Cécile, Diane, Marion, Patrick, Justine, Elias, Marie et Mattias. Ne lâchez rien !

Papa, Maman, Claire, je suis tant heureux et fier à la fois que l'on puisse à nouveau partager des instants en famille. Je vous aime fort.

Adeline, ma petite *Mother of cats*, ta présence et ton amour me sont vitaux. Merci pour ton soutien indéfectible dans nos projets les plus fous. Merci de supporter les *Pink floyd* et mes élucubrations sur l'harmonie. Merci pour la façon que tu as d'aimer et de faire grandir notre fils. L'avenir nous appartient, à nous d'en écrire les plus belles pages.

Alban, mon fiston, ma petite belette, mon petit bout de mec. Tu as rendu ma vie magique, tu es ce qu'il y a de plus cher à mes yeux, grâce à toi je connais cet indescriptible amour que l'on ressent pour son enfant. Je donnerais tout pour toi, je t'aime mon chéri.

Merci Pipou, tu es formicat. Merci David Gilmour, tu donnes un sens à mon ouïe.

Enfin, parce que les écrits restent, je ne peux m'empêcher de citer les gens que j'aime, partis trop loin : mes grands-pères, mes grands-mères, Denise, Bruno, Marcelle & Charlot, Yves.

Liste des abréviations

Ac : anticorps

ADN : acide désoxyribonucléique

ADNc : acide désoxyribonucléique complémentaire

Ag : antigène

aMPVc : métapneumovirus aviaire de type C

ARN : acide ribonucléique

BLAST : *Basic Local Alignment Search Tool*

BWA : *burrows-wheeler aligner*

CIFRE : convention industrielle de formation par la recherche en entreprise

Cq : cycle de quantification

DIVA : différenciation des animaux infectés de ceux vaccinés

dNTP : nucléoside triphosphate

ECP : effet cytopathique

ELISA : *enzyme linked immunosorbent assay*

EOPS : exempt d'organismes pathogènes spécifiques

EQA : *external quality assessment*

ESI-TOF: *electron spray ionization – time of flight*

EVE : éléments viraux endogènes

GAAS : *genome relative abundance and average size*

gfCoV : : coronavirus de la pintade

Gb : gigabase

HEV : virus de l'hépatite E

HIV : *human immunodeficiency virus*

HRM : *high resolution melting*

IAHP : influenza aviaire hautement pathogène

IGV : *integrative genomics viewer*

IQA : *internal quality assessment*

IQC : *internal quality control*

Kb : kilobase

MALDI-TOF: *matrix-assisted laser desorption/ionization - time of flight*

ME : microscopie électronique

MEB : microscopie électronique à balayage

MET : microscopie électronique à transmission

MFP : maladie foudroyante de la pintade

MIQE : *minimum information for publication of quantitative digital pcr experiments*

NCBI : *national center for biotechnology information*

NGS : nouvelles générations de séquençage

OIE : organisation mondiale de la santé animale

ONT : *Oxford nanopore technologies*

nt : nucléotide

pb : paire de bases

PCR : *polymerase chain reaction*

PGM : *personal genome machine*

pH : potentiel hydrogène

Ph. Eur. : pharmacopée européenne

qPCR : *quantitative PCR*

RT-qPCR : *real-time quantitative PCR*

SARS : *severe acute respiratory syndrome*

SNP : *single nucleotide polymorphism*

SR : syndrome respiratoire

TCID : *tissue culture infectious dose*

TCoV : *turkey coronavirus*

VIP : *virus identification pipeline*

Index des Illustrations

FIGURE 1 APERÇU DE L'OFFRE DES NGS.....	14
FIGURE 2 LOI DE FLATLEY ET SA SUPERPOSITION A LA LOI DE MOORE.....	15
FIGURE 3 INTERPRETATION DES RESULTATS ISSUS DE CONTROLE QUALITE.....	21
FIGURE 4 APERÇU DE L'OFFRE EN SEQUENCEURS NGS AU FIL DU TEMPS.....	28
FIGURE 5 ÉVOLUTION DES PARUTIONS DE PUBLICATIONS TRAITANT DE BIO-INFORMATIQUE ET DE METAGENOMIQUE	30
FIGURE 6 CAPTURE D'ECRAN D'UN EXEMPLE DE RESULTATS GENERE PAR VIP (LI ET AL 2016).....	32
FIGURE 7 RESULTATS DE RT-PCRQ BIOMARK.....	39
FIGURE 8 PRINCIPE DE FONCTIONNEMENT DE LA TECHNOLOGIE ILLUMINA.....	68
FIGURE 9 SIGNAL FLUORESCENT GENERE LORS DE SEQUENCAGE ILLUMINA.....	69
FIGURE 10 KRONA GENERE PAR L'OUTIL D'ANALYSE BIOINFORMATIQUE GAAS.....	71
FIGURE 11 SEQUENCEUR MINION	79
FIGURE 12 ILLUSTRATION DU SEQUENÇAGE MINION.....	80
FIGURE 13 REPRESENTATION SCHEMATIQUE DE LA PREPARATION DES LIBRAIRIES MINION	82
FIGURE 14 AFFICHAGE VIA LE PROGRAMME IGV D'UN L'ALIGNEMENT DE SEQUENCES	83

Liste des annexes

ANNEXE 1 CONTROLES A METTRE EN ŒUVRE POUR LA VALIDATION DE PURETE MICROBIOLOGIQUE DE PREPARATIONS VACCINALES.....	111
----------------------------------------------------------------------------------------------------------------------	-----

Table des matières

Introduction générale	1
1 Méthodes de détection des agents infectieux	3
1.1 Méthodes sérologiques.....	5
1.2 Méthodes de détection de micro-organismes directes.....	7
1.2.1 Culture cellulaire	7
1.2.2 Microscopie électronique	8
1.3 Méthodes de détection de micro-organismes directes et ciblées.....	8
1.3.1 Détection des acides nucléiques.....	8
1.3.1.1 PCR	9
1.3.1.2 PCR multiplex	10
1.3.1.3 PCR à haut débit.....	11
1.3.2 Séquençage d'ADN.....	11
1.4 Méthodes directes ciblant un large panel d'agents infectieux.....	15
1.4.1 Spectrométrie de masse à temps de vol.....	15
1.4.1.1 ESI-TOF	15
1.4.2 Puces de type <i>virochips</i>	16
1.4.3 <i>E-probes</i> : séquençage NGS sans <i>a priori</i> avec résolution ciblée par bioinformatique...	17
1.4.4 Technologie <i>AmpliSeq</i> (Ion Torrent)	17
1.4.5 Technologie <i>Luminex</i>	18
1.5 Détection de micro-organismes sans <i>a priori</i>	18
2 Validation des tests moléculaires et contrôle qualité	19
3 Le cas particulier du contrôle de sureté des produits biologiques.....	22
3.1 Tests de pureté	24
3.1.1 Recherche directe spécifique :.....	24
3.1.2 Recherche directe non spécifique :.....	24
3.1.3 Recherche indirecte spécifique :.....	24
4 Verrous méthodologiques liés à la biologie moléculaire moderne et à la manipulation de séquences.....	25
Objectifs de recherche Introduction à la partie expérimentale	33
Première partie expérimentale	35

A broad-range panel assay for avian respiratory pathogens on a nanofluidic pcr platform.....	42
Deuxième partie expérimentale	65
Novel Avian Coronavirus and Fulminating Disease in Guinea Fowl, France	73
Troisième partie expérimentale	77
Rapid Whole-Genome Based Typing and Surveillance of Avipoxviruses Using Nanopore Sequencing	86
Discussion générale et perspectives.....	97
Références Bibliographiques.....	105

Introduction générale

Outils modernes de génomique :

Leur impact sur le diagnostic virologique et le contrôle des produits biologiques vétérinaires.

Les maladies infectieuses chez les animaux domestiques ou sauvages ont un impact sévère sur l'économie d'un pays voire sur la santé publique puisque 60% des maladies infectieuses humaines sont d'origine zoonotique (Jones et al., 2008).

D'après Jones et al, les maladies émergentes peuvent avoir trois types d'origine :

- Une évolution de souches microbiennes avec, par exemple, l'acquisition de résistance aux antibiotiques ou antiviraux.
- Le contact pour la première fois de micro-organismes avec l'espèce humaine (exemple : HIV-1, SARS).
- L'infection historique de la population humaine à bas bruit par des agents pathogènes dont l'incidence a significativement augmenté récemment (exemple : maladie de Lyme).

Cette étude indique que 54.3% des micro-organismes impliqués dans les maladies émergentes sont des bactéries ou rickettsies alors que les virus ou les prions sont responsables de 25.4% des maladies émergentes, que la part restante est répartie entre champignons, protozoaires et helminthes. Les zoonoses émergentes sont quant à elles causées à 71.8% par des micro-organismes issus de la faune sauvage.

En raison de la hausse de la demande en protéines d'origine animale (viandes de volailles et œufs essentiellement), la productivité des élevages de volaille a considérablement augmenté ces dernières décennies. Selon Bagust TJ, le développement de l'élevage intensif de volailles dans les pays en voie de développement est directement relié à la croissance de la taille et de la quantité des petites et moyennes exploitations de volailles commerciales (Bagust, 2008). Bien que les maladies nutritionnelles et métaboliques soient considérables, le contrôle des maladies d'origine infectieuse est une priorité. Au niveau global, des différences dans la répartition géographique des agents pathogènes sont claires avec des variants génétiques associés à certaines régions du monde. L'émergence de souches pathogènes hautement

virulentes est d'ailleurs favorisée par les échanges commerciaux depuis l'entrée des pays émergents dans le marché de l'exportation.

Les agents pathogènes aviaires les plus susceptibles d'émerger sont des virus à génome ARN (acide ribonucléique) et précisément ceux responsables de la bronchite infectieuse, de la maladie de Newcastle, de la maladie de Gumboro et de l'influenza aviaire.

Les règles de biosécurité et le suivi vétérinaire des élevages sont fondamentaux et leur inobservance peut être à l'origine de l'introduction d'un agent pathogène au sein d'un ou plusieurs élevage(s) de volaille. Un tel évènement peut représenter la menace pérenne d'une maladie endémique au sein d'une région géographique.

Sur le plan économique, une étude réalisée en 2007 par l'Université de Géorgie (États-Unis), a estimé la perte brute de l'industrie de la volaille à 28.8 milliards de dollars américains (USD) dont 8.2% (2.3 milliards) correspondent aux pertes liées aux maladies. À chaque millier de USD perdu suite à la mortalité, deux autres milliers sont perdus par ailleurs en raison de la diminution de production causée par l'épisode clinique.

Les changements climatiques ont un impact sur les habitats de la faune sauvage et notamment chez les oiseaux migrateurs. Il a été démontré que les rudes hivers européens sont responsables du maintien de canards migrateurs au sein d'une zone géographique, créant ainsi les conditions nécessaires à la diffusion et la persistance de virus influenza A dans cette même zone (Keller et al., 2009; Reperant et al., 2010). L'accroissement des précipitations sur la côte Est des États-Unis est à l'origine de zones humides davantage étendues qui favorisent l'accumulation de canards sauvages, espèces réservoirs naturels de virus influenza aviaires (Vandegrift et al., 2010).

À titre d'illustration de cette évolution de la situation épidémiologique à l'échelle globale, ces dernières années, notre pays et en particulier le sud-ouest de la France a payé un lourd tribut du fait de l'émergence et la diffusion successive de plusieurs virus influenza aviaires hautement pathogènes (IAHP).

1 Méthodes de détection des agents infectieux

La surveillance et le contrôle des maladies animales nécessite une forte capacité de détection des agents pathogènes et ce à l'aide de tests rapides, peu coûteux, spécifiques et

sensibles. La détection et l'identification des agents présents dans n'importe quel type de prélèvement sont les objectifs à atteindre en matière de diagnostic en médecine vétérinaire. La démarche classique appliquée au diagnostic repose sur des techniques robustes, validées et pour lesquelles les laboratoires de diagnostic possèdent une expérience significative et des procédures de contrôle qualité.

La détection, directe ou indirecte, d'agents infectieux est mise en œuvre dans deux grands types d'applications :

- La **surveillance** des maladies d'intérêt économique ou liées à la santé publique.
- Le **diagnostic** dans le cas d'une suspicion d'une maladie.

Les signes cliniques présentés par un individu ou un élevage sont dans un premier temps explorés par le vétérinaire qui, après avoir suspecté l'agent ou les agents responsable(s), déclenchera le processus d'analyse. Les analyses qui en découlent sont aujourd'hui des tests ciblés pour une recherche directe de l'agent suspecté.

Les bases régissant la relation de cause à effet liant un agent pathogène à une maladie sont toujours fortement inspirées du postulat initial de Koch selon lequel, par exemple, un agent infectieux sera toujours présent dans un organe touché par une maladie mais absent d'un organe sain. Selon Koch, l'agent responsable de la maladie doit également être cultivable en laboratoire et doit reproduire la maladie après son inoculation à un hôte sain.

Dans leur revue de 2016 traitant du postulat initial de Koch, Byrd et Segre se basent sur des études réalisées en bactériologie pour proposer un nouvel angle sous lequel le postulat devrait être abordé (Byrd and Segre, 2016). Ils rappellent également qu'une valence supplémentaire, non incluse par Koch lui-même, énonce que l'agent infectieux doit pouvoir être cultivé après une infection expérimentale. Selon les auteurs, des germes commensaux ont la capacité de protéger leur hôte contre d'autres agents pathogènes, on qualifie cela de « colonisation de résistance ». Des études ont modernisé le postulat initial de Koch en prenant en compte les communautés bactériennes dans leur ensemble. Buffie *et al* ont par exemple démontré le rôle joué par *Clostridium scindens* dans la résistance à *Clostridium difficile* en agissant sur l'acidité gastrique (Buffie et al., 2015).

La distinction entre une infection virale ou bactérienne doit pouvoir se faire au stade le plus précoce possible afin que le clinicien puisse prescrire, ou non, un traitement antibiotique et éliminer le risque d'une infection bactérienne grave. Les effets d'un traitement antibiotique

sont aujourd'hui connus et reconnus, les risques vont de l'émergence de résistance aux antibiotiques à des réactions allergiques voire une toxicité induite par le traitement. Un traitement antibiotique peut également fausser le résultat d'un test diagnostique. Afin de contourner la culture de bactéries sur milieu gélosé, qui est une technique longue et peu sensible, le diagnostic des infections bactériennes peut être effectuée par PCR en ciblant l'ARN ribosomal 16S. Ainsi, l'identification de bactéries pathogènes chez des patients sous traitement antibiotique est efficacement réalisée (Rampini et al., 2011). L'ARN 16S présente l'avantage d'être suffisamment conservé entre les genres bactériens pour pouvoir amplifier et détecter les séquences de différentes espèces en une réaction. Les virus ne présentent cependant pas cette caractéristique de conservation d'un gène discriminant et leurs génomes sont davantage soumis à des variations génétiques. Les tests moléculaires sont donc plus lourds et permettent plus difficilement d'effectuer un criblage large.

1.1 Méthodes sérologiques

Les méthodes indirectes ne sont pas basées sur la recherche de l'agent pathogène. Ce sont des techniques sérologiques qui consistent en la recherche d'antigènes (Ag) issus du micro-organisme infectant l'hôte ou bien la recherche d'anticorps (Ac) sécrétés par l'hôte en réponse à l'infection.

Une technique de référence employée en diagnostic sérologique est le test ELISA pour la recherche dans l'échantillon d'antigènes (Ag) de l'agent infectieux ou d'anticorps (Ac) spécifiques de l'agent infectieux potentiellement présents dans le sérum de l'hôte. Le test ELISA existe sous forme de test rapide directement utilisable sur le terrain par le vétérinaire ou bien sous une forme plus complète (permettant de titrer le taux d'Ag ou d'Ac) à utiliser en laboratoire.

D'une façon plus générale, les tests sérologiques de type agglutination, hémagglutination ou inhibition de l'hémagglutination, séroneutralisation, immuno-diffusion sur gélose ou encore ELISA présentent la particularité d'être des tests sensibles et spécifiques.

Comme définit par Saah et Hoover, la sensibilité analytique représente la plus petite quantité d'une substance ou organisme quantifiable avec précision dans un échantillon ; la spécificité analytique est quant à elle définie par la capacité d'un test à détecter un organisme ou substance particulier(e). Les sensibilité et spécificité sont cependant distinctes selon qu'elles

appartiennent au secteur analytique ou diagnostic. La sensibilité diagnostique est représentée par le pourcentage d'individus soupçonnés porteurs d'une maladie et identifiés par le test comme étant effectivement porteurs de la maladie. Enfin, la spécificité diagnostique est le pourcentage d'individus non malades et identifiés comme tels par le test (Saah and Hoover, 1998).

Lors de l'interprétation de résultats issus de tests sérologiques, l'attention du biologiste doit être portée sur le statut de l'animal testé qui, pour un stade d'infection précoce, ne présenterait pas systématiquement d'Ac spécifiques contre l'agent responsable de la maladie. Les tests de détection directe de l'agent infectieux ou de son génome seront dans ce cas plus pertinents. Par ailleurs, le sérum d'un animal vacciné pourra être faussement positif si le test ne permet pas de discriminer les Ac dirigés contre une souche vaccinale de ceux dirigés contre la souche circulante (Day, 2015).

De sorte à contourner la problématique de différenciation des Ac issus d'une vaccination de ceux issus d'une infection, le développement de vaccins peut suivre la stratégie nommée DIVA (Différenciation des Animaux Infectés de ceux Vaccinés). Cette stratégie repose sur une association de l'Ag contre lequel on souhaite vacciner à un autre Ag qui n'entre pas dans la protection immunitaire souhaitée. Un test de diagnostic permettant de détecter la réponse sérologique contre l'antigène contenu dans le vecteur vaccinal et de détecter la réponse sérologique contre un antigène viral non contenu dans le vecteur vaccinal permettra de différencier les animaux infectés de ceux vaccinés. Dans le cadre de l'influenza aviaire, un vaccin DIVA sera composé d'un sous-type d'hémagglutinine homologue à celui de la souche circulante tandis que le sous-type de neuraminidase sera hétérologue. Ainsi, la différenciation entre animaux infectés et animaux vaccinés se fera sur la nature des Ac anti-neuraminidase (Capua et al., 2003). L'efficacité de l'utilisation d'un virus différent comme vecteur vaccinal a également été démontrée, notamment lors de l'utilisation du virus de la maladie de Newcastle pour vacciner contre le sous-type H5 de l'influenza aviaire (Veits et al., 2008). Cependant, cette distance antigénique entre la souche vaccinale et la souche circulante ou l'expression d'un seul type d'antigène, par exemple à la surface d'un baculovirus, peut être à l'origine d'une efficacité de vaccination réduite (Beer et al., 2007).

Tous ces tests restent des méthodes de référence utiles pour la validation de nouvelles générations de techniques. On les qualifie alors de *gold standard*.

1.2 Méthodes de détection de micro-organismes directes

1.2.1 Culture cellulaire

La culture, en laboratoire, d'agents infectieux est une méthode de choix incontournable pour certaines applications nécessitant de multiplier l'agent pour suivre son évolution ou pour en faire des stocks. La culture de cellules est apparue au début du XX^{ème} siècle avec la culture d'organes entiers puis a évolué en culture de cellules primaires (issues d'animaux de laboratoire ou de prélèvements humains) ou de cellules immortalisées qui présentent l'avantage de croître indéfiniment. Le premier isolement viral date de 1949 lorsque John Enders et son équipe parviennent à cultiver un Poliovirus sur des cellules primaires humaines.

Du fait de leur propriété de parasite intracellulaire obligatoire, les virus se cultivent *in vitro* sur cellules ou *in vivo* sur œufs embryonnés (concernant les virus d'origine aviaire) ou animaux vivants. La production de virus en laboratoire est nécessaire pour mettre au point ou produire du stock de souche vaccinale (atténuée après passages sur cellules par exemple), réaliser des études de structure virale, des études d'interactions avec les facteurs cellulaires, détecter ou titrer un virus en suivant l'effet cytopathique ou encore isoler un variant viral. Des cinétiques de croissance virale nécessitent également la mise en culture de virus.

Le titrage de particules virales est une des applications majeures de la culture cellulaire, elle est issue des travaux de Renato Dulbecco qui, en 1952, a été le premier à quantifier une suspension virale par la technique des plages de lyse. Cette méthode a ensuite été supplantée par les techniques en dilutions limites telles que la TCID₅₀ (*tissue culture infectious dose*) qui sont des techniques de titrage statistiquement plus robustes du fait de leur mise en œuvre à plus grande échelle. Cependant, les plages de lyse restent, pour certains virus, nécessaires pour isoler physiquement des clones viraux. Suite à l'infection d'une cellule par une particule virale, sous milieu gélosé, les virions émanant de la cellule infectée diffuseront uniquement dans les cellules voisines. Il sera alors possible de sélectionner spécifiquement la plage de lyse et les particules virales clonales qui y sont présentes.

La culture des bactéries se réalise quant à elle sur milieux gélosés en respectant des conditions de nutrition, de pH ou encore de respiration des différentes espèces bactériennes. Leur culture peut être nécessaire pour leur identification qui est réalisée selon le milieu sur lequel elles croissent, leurs caractères morphologiques, physico-chimiques ou biochimiques. Néanmoins,

leur mise en culture se révélera parfois infructueuse compte tenu de l'existence de bactéries non cultivables en conditions de laboratoire.

1.2.2 Microscopie électronique

Dans le cadre du diagnostic, la culture virale sera réalisée en amont de la microscopie électronique qui permettra l'identification de la famille virale selon des critères morphologiques. La microscopie électronique a été développée dans les années 1930 pour surpasser les limites de la microscopie optique et la première observation d'un virus par microscopie électronique date de 1939. Dans les années qui suivirent, les taux de grossissement ont atteint 100 000x. Deux grands principes de microscopie électronique existent : la microscopie électronique à transmission (MET) et la microscopie électronique à balayage (MEB). La microscopie électronique est employée pour le comptage de particules virales ou pour l'étude de la structure virale ; elle peut également être utilisée pour du diagnostic mais l'interprétation de l'imagerie doit dans ce cas être très précautionneuse. En effet, des organites cellulaires comme les ribosomes peuvent être confondus avec des particules virales.

1.3 Méthodes de détection de micro-organismes directes et ciblées

1.3.1 Détection des acides nucléiques

Les tests moléculaires permettent la détection directe du génome de l'agent suspecté et ce directement après avoir extrait le matériel génétique de l'échantillon à analyser. L'avantage de pouvoir travailler à partir d'acides nucléiques réside dans leur stabilité lorsqu'ils sont conservés à basse température et dans leur caractère non infectieux évitant tout risque de biosécurité après extraction des acides nucléiques, voire avant par utilisation de milieux de prélèvement permettant l'inactivation immédiate des agents (fixateurs RNA later/Trizol ou papiers traités). Dans la catégorie des papiers traités, le papier FTA Whatman™ fait office de référence, le FTA est un traitement chimique qui entraîne la lyse cellulaire après l'application des échantillons sur le papier. Il s'agit d'un procédé permettant de récupérer, stocker ou transporter des acides nucléiques sans précaution particulière.

Les tests moléculaires sont utiles pour la recherche d'un agent infectieux donné mais ils permettent également une caractérisation plus fine par la recherche de signatures génétiques

spécifiques de résistance aux antibiotiques ou antiviraux ou encore des marqueurs de pathogénicité (Araujo et al., 2017; Assoumou et al., 2017; Zhang et al., 2017).

1.3.1.1 PCR

Les tests moléculaires les plus utilisés aujourd'hui sont basés sur l'amplification du génome cible, ils reposent sur la technique de PCR (Mullis et al., 1992). L'apport de la PCR dans le domaine du diagnostic est considérable au vu de ses performances en matière de sensibilité et de spécificité. Le principe d'amplification d'une région génomique permet de multiplier une molécule d'ADN ou d'ARN initialement présente en trop faibles quantités pour être détectée ou pour d'autres applications nécessitant beaucoup de matériel.

La spécificité de la PCR est liée à l'utilisation d'oligonucléotides complémentaires de la séquence nucléotidique ciblée. La conception des oligonucléotides est capitale dans la spécificité de la réaction, on devra choisir une zone du génome peu variable et conservée pour s'assurer de l'hybridation des amorces au sein d'une même famille virale. En cas de variations nucléotidiques connues, l'utilisation d'oligonucléotides dégénérés représente une alternative de choix. Cependant, bien qu'un résultat positif soit robuste et fiable, un résultat pourra être faussement négatif si l'on a affaire à un couple d'amorce trop strict.

La PCR conventionnelle fournit un résultat qualitatif voire semi-quantitatif, elle se voit aujourd'hui remplacée au profit de la PCR quantitative en temps réel (qPCR). La qPCR offre une information quantitative absolue ou relative. Cela permet de quantifier le génome d'un micro-organisme ou encore de quantifier l'expression d'un gène (RT-qPCR) de l'hôte impliqué, par exemple, dans la réponse à l'infection.

La RT-qPCR est aujourd'hui démocratisée autour de deux grands formats de fluorescence : un premier type qui utilise les propriétés d'un fluorophore spécifique de l'ADN double brin, le SYBR Green et un second type basé sur l'utilisation de sondes d'hydrolyse ou sondes Taqman.

La chimie SYBR Green est basée sur l'utilisation de molécules intercalantes de l'ADN qui émettent de la fluorescence une fois incorporées dans un double brin d'ADN et ce même pour de faibles quantité d'ADN. Cette technologie présente une sensibilité accrue permettant potentiellement la détection de seulement 20 pg d'ADN de génome initialement présents dans un échantillon (Karlsen et al., 1995). La spécificité de la réaction sera également assurée par

l'interprétation de la courbe de fusion qui apporte une information relative à la composition en bases azotées du fragment d'ADN amplifié.

L'utilisation d'une sonde Taqman accroît considérablement la spécificité de la réaction puisqu'il s'agit d'une séquence nucléotidique complémentaire de la zone d'ADN à amplifier. En complément des amorces, l'hybridation de la sonde sera nécessaire pour la détection de l'agent ciblé. De plus, la fluorescence ne sera émise que si l'ADN polymérase synthétise le brin d'ADN complémentaire de l'ADN recherché.

Le choix entre la chimie SYBR Green et la chimie Taqman se fera selon la spécificité désirée. Une PCR Taqman sera préférée lorsqu'un motif particulier ne présentant pas de risque de variabilité est recherché. La chimie SYBR Green présente quant à elle l'avantage de pouvoir amplifier une zone potentiellement variable, à condition que les sites d'hybridation des amorces soient stables. L'analyse de la courbe de fusion offerte par la technologie SYBR Green apporte par exemple la possibilité d'identifier la présence de mutants ou de motifs pertinents et ce, notamment, grâce à la technologie HRM (*High Resolution Melting*).

La PCR est une technologie de choix pour la recherche d'un agent pathogène précis. Cependant, le domaine du diagnostic requiert parfois, selon le contexte clinique, la recherche de plusieurs agents dans un même échantillon et en une seule réaction. L'objectif de la manœuvre est d'une part économique mais cela permet également d'économiser du temps et de préserver les échantillons. On parle alors de PCR multiplex.

1.3.1.2 PCR multiplex

La PCR multiplex est aujourd'hui largement commercialisée sous forme de kits employant la technologie Taqman. En effet, une réaction dédiée à la recherche de plusieurs agents pathogènes se fera grâce au marquage de différentes sondes avec des fluorophores dont les longueurs d'ondes d'émission se situent dans des spectres distincts. Le facteur limitant de la capacité de multiplexage est en fait déterminé par la disponibilité et la diversité des fluorochromes ainsi que par la capacité du thermocycleur à pouvoir discriminer les différents signaux de fluorescence en temps réel.

La rapidité et la simplicité de mise en œuvre de l'outil PCR en ont naturellement fait un outil incontournable et largement démocratisé dans la majorité des laboratoires d'analyse

médicale. Toutefois, il est nécessaire de souligner que cette technologie fournit une information sur la présence du matériel génétique d'un micro-organisme mais en aucun cas sur la capacité du micro-organisme à se multiplier.

1.3.1.3 PCR à haut débit

La PCR reste un outil phare au vu de ses qualités de spécificité et de sensibilité, l'enjeu réside donc dans la possibilité de pouvoir rechercher un large panel de micro-organismes dans un grand nombre d'échantillons, simultanément. Le développement des méthodes de diagnostic innovantes s'accompagne d'une diminution des volumes réactionnels, les systèmes microfluidiques couplés à des appareils de détection de signal reposant sur les nanotechnologies font désormais partie du paysage de l'analyse moléculaire (Coelho et al., 2017). Cette diminution de volumes permet par conséquent une multiplication des réactions.

Le format classique de plaques de 96 puits a été complété par un format de 384 puits afin d'augmenter la capacité de criblage, cependant il s'agit toujours de qPCR à capacité de multiplexage limitée. L'analyse proposée par le Biomark (Fluidigm) est qualifiée de test à haut débit car elle permet le criblage de 48 ou 96 échantillons vis-à-vis de 48 ou 96 couples d'amorces en une seule réaction de qPCR. Il s'agit cependant toujours d'une recherche ciblée nécessitant une connaissance préalable des agents recherchés et de leurs génomes, le tout sous réserve qu'une modification génétique n'entrave pas l'analyse.

1.3.2 Séquençage d'ADN

Le séquençage d'ADN par la technique de Sanger (Sanger et al., 1992) est un procédé de lecture d'une suite de bases azotées au sein d'une molécule d'ADN, c'est une technique de séquençage par terminaison de réaction. Cette technologie innovante a notamment été à l'origine du projet génome humain entrepris en 1990 et achevé en 2003, ces treize années de travaux ont permis d'obtenir une séquence et une cartographie de génome humain mais aussi d'autres espèces de mammifères comme la souris et le rat. Le séquençage Sanger est la technique de choix pour un besoin ponctuel, rapide et ciblé lorsqu'il n'est pas nécessaire d'avoir recours au séquençage à haut débit. La méthode de Maxam, Gilbert (Maxam and Gilbert, 1992) a été, quant à elle, nettement moins démocratisée.

Dans le domaine de l'identification génétique ou du diagnostic, le séquençage apporte une information capitale puisque la séquence génomique comporte l'information source du micro-organisme. L'interprétation de la séquence obtenue se fait par comparaison à des bases de données publiques ou personnelles. Classiquement, les biologistes chargent leurs séquences à analyser *via* l'interface *web* de l'outil BLAST (*Basic Local Alignment Search Tool*) (Altschul et al., 1990) qui repose sur la base de données du NCBI (pubmeddev). Afin d'outrepasser les délais d'attente (qui dépendent du réveil des biologistes du Nouveau Monde et qui s'allongent avec la taille des jeux de données à analyser), l'outil BLAST peut être utilisé sur un centre de calcul par l'intermédiaire d'un terminal.

Néanmoins, l'information de séquence génomique ne se suffit pas systématiquement à elle-même. Le génome de tout organisme ou micro-organisme ne lui est pas propre dans son intégralité, en effet, de nombreux gènes peuvent être conservés d'une famille à l'autre ou être très variables d'une espèce à l'autre. Le choix de la séquence à amplifier avant séquençage se fera selon le niveau de précision souhaité dans l'identification. Le séquençage d'un gène bactérien codant pour les ARN ribosomiques 16S sera par exemple difficilement exploitable pour identifier une espèce bactérienne en raison de la faible vitesse d'évolution de ces gènes.

Les techniques traditionnelles de mise en culture restent alors incontournables pour étudier la viabilité d'un agent et ne peuvent, jusqu'à présent, pas être supplantées par les techniques moléculaires.

Une fois l'hypothèse du clinicien validée par les tests biologiques, le diagnostic est validé et un traitement est mis en place.

Les techniques évoquées ci-dessus correspondent à de la recherche ciblée sur un ou quelques agent(s), ce qui implique d'avoir une suspicion précise de l'agent causal de la maladie. Dans le cas de maladies dont l'agent causal demeure inconnu, les techniques précédemment citées nécessiteront un déploiement d'efforts considérable, coûteux, long et potentiellement inefficace. De plus, les maladies d'origine multifactorielle complexifient la démarche, le dogme « une maladie/un agent » est aujourd'hui obsolète puisque les coinfections, le microbiote, le patrimoine génétique, l'alimentation ou encore l'environnement dans lequel l'hôte évolue sont des paramètres primaires impliqués dans le développement des maladies. Ian Lipkin évoque notamment dans sa revue de 2013 les liens entre *Helicobacter pylori* et l'ulcère de l'estomac, le papillomavirus humain et le cancer du col de l'utérus ou encore l'association du polyomavirus et le carcinome à cellules de Merkel (Lipkin, 2013). Toutes ces maladies étaient

encore considérées comme idiopathiques il y a quelques années, c'est la preuve qu'une exploration large à l'échelle de l'hôte dans son intégralité est une démarche à ne pas négliger.

Jusqu'aux années 2000, la recherche d'agents jusqu'alors inconnus (mais suspectés) pouvait se faire par microscopie électronique (Roingard, 2008) ou par une amplification aléatoire des ADNs et ARNs présents suivie d'un clonage (Zou et al., 2003) puis d'un séquençage par la technique de Sanger. Seul un nombre limité de clones étaient séquencés et les séquences produites ne dépassaient guère 1 kb. Suite à l'analyse des données, les séquences pertinentes pouvaient offrir une information sur la famille virale ou bactérienne à rechercher. De la marche sur génome ou de la PCR conventionnelle ciblée suivie de séquençage étaient ensuite effectuées pour approfondir l'exploration du génome suspecté.

Aujourd'hui, les techniques de biologie moléculaire moderne permettent d'envisager une démarche de diagnostic radicalement différente de celle décrite précédemment. Les capacités de production de données et leur analyse progressent considérablement (Figure 1) et placent les techniques modernes en tête de l'arbre décisionnel suivi dans le domaine du diagnostic moléculaire. Les technologies de séquençage faisant appel à l'informatique et l'électronique de pointe, on peut supposer que la loi de Moore s'y applique.

Cette loi, initialement énoncée par Gordon Moore en 1965, propose que tous les 18 mois, la puissance des dispositifs informatiques double, tant au niveau des capacités de stockage que des puissances de calcul. De la même sorte, les tarifs diminuent d'un facteur 2 tous les 18 mois. Ces dernières années, les lois de la physique, de la consommation et de l'économie ont diminué la vitesse d'accroissement des capacités informatiques. En effet, selon les théoriciens il est peu probable de pouvoir stocker plus d'un bit par atome de matière ; de plus le coût de l'investissement aujourd'hui nécessaire pour maintenir le niveau de progrès augmente de façon exponentielle (loi de Rock, (Stoffan, 2004)) et, enfin, les consommateurs renouvellent moins leurs ordinateurs et consacrent d'avantage leur budget aux téléphones ou aux tablettes.

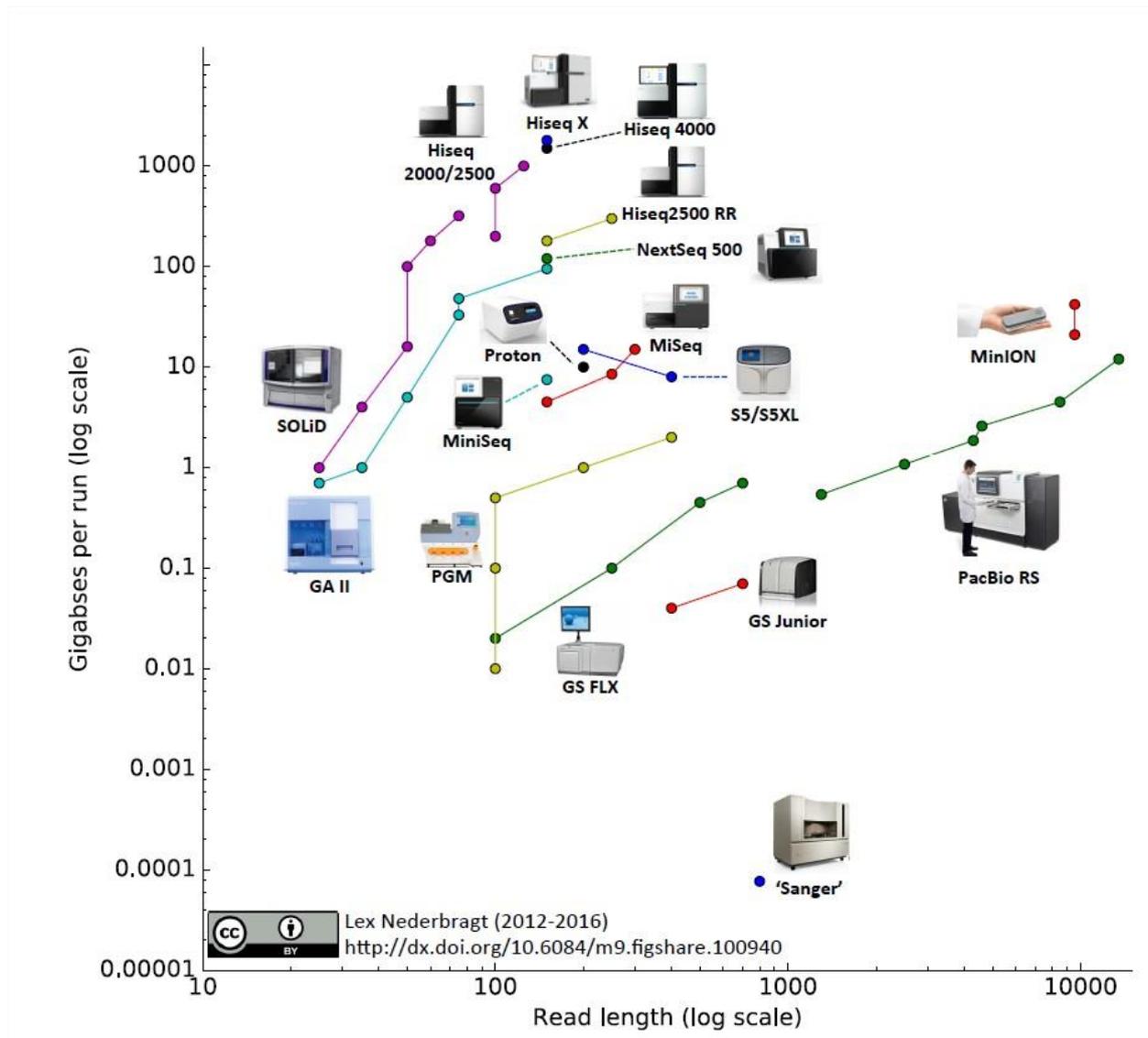


Figure 1 Aperçu de l'offre des NGS, de la quantité de nucléotides lus et de la longueur des séquences générées par réaction de séquençage.

Dès 2005, G. Moore annonçait que les limites de sa loi étaient proches :

« Cela ne peut durer indéfiniment. Le propre d'une exponentielle est de vous entraîner hors des limites raisonnables jusqu'à un éventuel désastre. La taille d'un transistor approche celle d'un atome, qui est une barrière fondamentale. Nous l'atteindrons d'ici deux ou trois générations [de microprocesseurs]. »

L'évolution des performances et des tarifs du séquençage NGS suit la loi de Flatley (ancien PDG d'Illumina) qui est basée sur la chute des coûts de séquençage. À titre d'exemple, le séquençage d'un génome humain est passé de 100 millions USD à 1000 USD en 13 ans soit un facteur de réduction 1 000 fois plus important que celui de la loi de Moore (Figure 2).

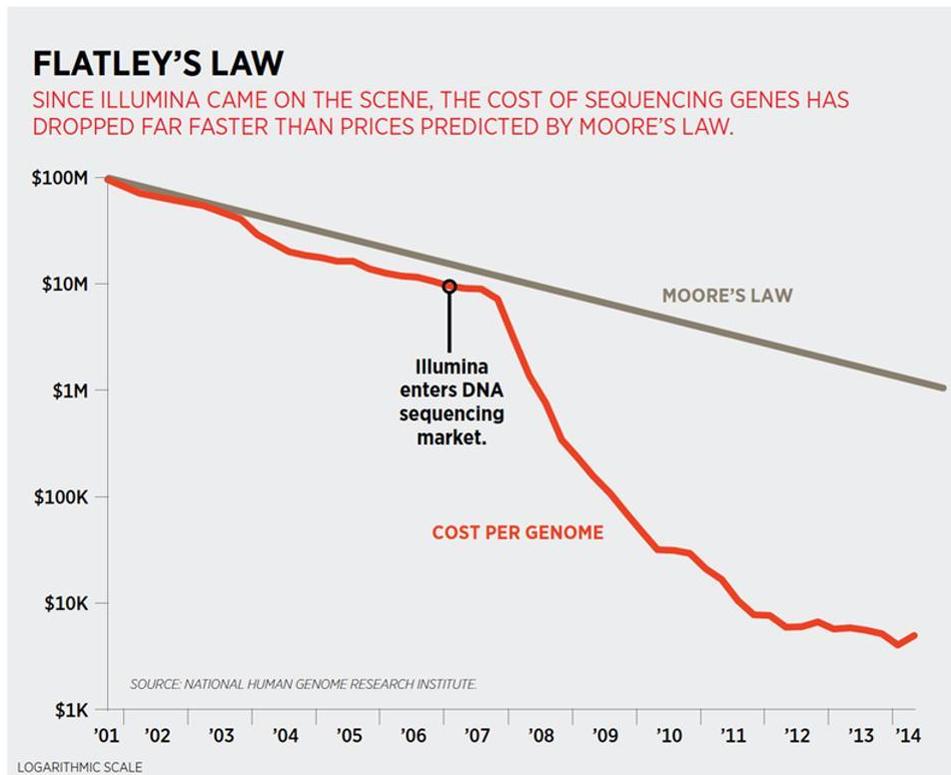


Figure 2 Loi de Flatley et sa superposition à la loi de Moore.

Ces techniques ne sont pas systématiquement employées puisqu'elles nécessitent encore d'importantes ressources en matériel de laboratoire et en informatique mais elles tendent à le devenir grâce à une miniaturisation accompagnée d'une simplification des procédés et d'une externalisation des ressources de calcul (*cluster* informatique, *cloud*). Cependant, dans le cas d'analyses à haut débit, de description de microbiotes ou d'échec des techniques traditionnelles, les techniques modernes sont nécessaires.

1.4 Méthodes directes ciblant un large panel d'agents infectieux

1.4.1 Spectrométrie de masse à temps de vol

1.4.1.1 ESI-TOF

La spectrométrie de masse ESI-TOF (*Electron Spray Ionization – Time Of Flight*) est une méthode de détection d'acides nucléiques après leur amplification par des amorces aléatoires à large spectre ciblant des régions génétiques conservées encadrant des régions variables. Le temps de vol au sein du tube de vol du spectromètre des acides nucléiques est

mesuré et permettra l'identification de la composition en bases de l'acide nucléique soumis à l'analyse. En effet, le temps de vol de chaque amplicon possiblement amplifié par le panel d'amorces aléatoires lui est propre, l'analyse et l'identification de l'agent reposent sur la comparaison du temps de vol de l'échantillon à la base de données des cibles recherchées. Cette technologie a été validée pour le diagnostic d'infections respiratoires à virus influenza A en médecine humaine et a fait preuve de bons niveaux de sensibilité et de spécificité (Mengelle et al., 2013). Cette méthode présente l'avantage de ne pas avoir à cibler spécifiquement la recherche d'agent infectieux en utilisant des amorces à large spectre, cependant l'analyse totale reste longue à mettre en place, demande beaucoup de ressources de laboratoire et nécessite une amplification préalable des acides nucléiques. De plus, l'encombrement de l'équipement, son entretien et le coût d'utilisation sont des paramètres qui rendent son utilisation exclusive et difficilement applicable au diagnostic de routine.

1.4.2 Puces de type *virochips*

Les puces à ADN ne nécessitent pas d'amplification du génome et sont basées sur l'utilisation de sondes. Les sondes sont des séquences nucléotidiques complémentaires du génome recherché et sont fixées sur le support d'analyse sur lequel sera chargé l'échantillon à analyser. Grâce au couplage préalable des acides nucléiques de l'échantillon à un fluorochrome, une hybridation spécifique d'un ADN ou ADNc cible à sa sonde sera révélée par la fluorescence.

Cette technique présente l'avantage d'être extrêmement spécifique puisque les sondes peuvent atteindre plusieurs centaines de nucléotides. La spécificité accrue de cet outil en augmente la rigueur, ce qui peut présenter un inconvénient en cas de variabilité génétique mais cela en fait un outil remarquablement efficace pour la recherche de SNP (*Single Nucleotide Polymorphism*). La spécificité élevée de cet outil est cependant contournée par la possibilité d'hybrider plusieurs dizaines de milliers de sondes sur la puce, si la variabilité génétique peut être anticipée, un criblage large est envisageable. Chen et al ont présenté en 2011 une version de puce contenant environ 36 000 sondes développées à partir de 1 500 références virales de la base de données Genbank (Chen et al., 2011).

1.4.3 *E-probes* : séquençage NGS sans *a priori* avec résolution ciblée par bioinformatique

Les techniques détaillées dans la section à venir « Détection sans *a priori* » reposent sur du séquençage à très haut débit sans *a priori* dédié à une analyse exhaustive non orientée de métagénomes. Or, des données NGS peuvent également être soumises à une analyse ciblée. Une méthode innovante, que l'on pourrait qualifier de PCR *in silico*, consiste à rechercher dans des données de NGS des motifs nucléotidiques précis qualifiés de e-sondes (Visser et al., 2016). Le contournement de l'analyse « tout venant » de métagénomes constitue un avantage certain du fait de la complexité de l'analyse métagénomique, tant du point de vue de la complexité de l'analyse bioinformatique que des ressources et du temps nécessaires ainsi que de l'interprétation des résultats. Les e-sondes permettent une recherche précise directe sur des données générées sans *a priori* de séquences pertinentes, cela accroît la rapidité d'analyse et présente l'avantage de fournir un résultat directement corrélé aux agents infectieux d'intérêt. Cependant, il est apparu que cet outil est extrêmement stringent et ne permet pas d'erreur ni de variabilité entre la sonde de recherche et les séquences du jeu de données (Delpuech, 2017). Les génomes viraux et notamment à ARN pouvant présenter d'importantes variabilités génétiques, cet outil n'est pas adapté à leur recherche.

1.4.4 Technologie *AmpliSeq* (Ion Torrent)

Un des défis à surmonter dans la recherche d'agents pathogènes par métagénomique sans *a priori* est la suppression du génome de l'hôte (qui est généralement présent en abondance) en amont du séquençage. De sorte à s'affranchir des fastidieuses méthodes d'élimination du génome de l'hôte (qui sont potentiellement délétères), la société ThermoFisher scientifique propose une méthode d'enrichissement en génomes d'intérêt par PCR avant le séquençage NGS, c'est la technique *ION Ampliseq*. Le fabricant annonce la possibilité de réaliser une PCR *multiplex* ciblant plus de 20 000 régions en une seule réaction avant de réaliser un séquençage NGS sur le panel d'amplicons générés. Cette méthode est extrêmement innovante et basée sur un développement fin d'amorces de PCR (le développement d'amorce peut être réalisé par le fabricant). Cette application reste finalement principalement employée dans la recherche de mutations génétiques impliquées dans le développement de tumeurs en santé humaine. En effet, le principal avantage est de pouvoir réaliser un séquençage « profond » sur des régions génomiques identifiées comme étant porteuses de mutations d'intérêt.

L'utilisation de cette technique dans un contexte de diagnostic est complexe puisqu'elle nécessite une très bonne connaissance préalable des génomes recherchés. La possibilité de multiplexer à un très haut niveau peut compenser la variabilité nucléotidique mais cela implique de connaître les régions génomiques conservées, ce qui n'est pas le cas pour tous les virus à rechercher.

1.4.5 Technologie *Luminex*

La technologie Luminex est basée sur la cytométrie de flux, des microbilles et de la lumière laser. Cette méthode permet la détection de protéines (protéines impliquées dans la réponse à l'infection, par exemple) grâce au couplage d'anticorps sur les billes, la détection d'anticorps grâce au couplage de protéines sur les billes ou bien la détection d'acides nucléiques grâce au couplage d'oligonucléotides sur les billes. La fluorescence des microbilles varie selon la cible à détecter ; on déterminera à l'avance que telle ou telle fluorescence correspondra à telle ou telle recherche afin d'identifier les microbilles par cytométrie en flux. Le contrôle du couplage de la molécule recherchée se fera grâce à un second anticorps fluorescent qui se fixera sur la protéine ou l'anticorps déjà capté(e) par la bille ou bien par un fluorochrome qui se fixera sur la molécule d'ADN hybridée à la bille. Des kits dédiés à la santé humaine sont commercialisés et permettent la détection de virus respiratoires classiquement recherchés (Influenza virus, virus respiratoire syncytial, coronavirus, parainfluenza virus, entérovirus, métag pneumovirus). L'atout de cette technique est la possibilité de multiplexer la recherche en recherchant jusqu'à 19 cibles dans 94 échantillons en 4 heures de réaction (Ko et al., 2017).

1.5 Détection de micro-organismes sans *a priori*

L'aspect révolutionnaire de la démarche est basé sur le fait que la première étape consiste à générer des données génomiques pour ensuite proposer un diagnostic.

De sorte à pallier la contrainte imposée par la recherche ciblée de micro-organismes, les tests dits « sans *a priori* » promettent d'obtenir une vision riche et descriptive du milieu étudié.

La recherche sans *a priori* consiste à caractériser tout acide nucléique présent dans un échantillon suite à son extraction puis son amplification totalement aléatoire suivie d'un séquençage massif ou bien en réalisant un séquençage à très haut débit directement à partir du matériel génétique extrait. Dès lors, on peut supposer qu'une analyse sans *a priori* directement

réalisée sur une lésion ou un organe malade permettra de décrire la diversité des micro-organismes présents dans l'échantillon, leurs caractéristiques génétiques et leur statut phylogénétique. Cette approche idéale présente l'avantage de générer une quantité et une diversité d'informations précieuses en des temps réduits et à des coûts raisonnables. Cependant, ces outils sont encore très rarement employés en routine compte tenu des difficultés techniques à surmonter lors de la préparation de l'échantillon, de son séquençage et lors de l'analyse des jeux de données massifs et totalement dématérialisés.

L'analyse des jeux de données issus de NGS représente un verrou majeur dans la démarche de métagénomique. À titre d'exemple, un *run* de séquençage *Miseq* (Illumina) peut produire jusqu'à 30 millions de séquences nucléotidiques de 500 paires de bases chacune, ce qui représente un jeu de 15 milliards de caractères. Au 26 août 2017, la base données *Virus Pathogen Resource* contenait 717 910 références virales, comparer chacune des 30 millions de séquences à une telle base de données représente intuitivement un travail considérable. De sorte à rendre les analyses bioinformatiques réalisables dans des délais raisonnables, des centres de calcul ont été développés autour de chaque pôle scientifique. À Toulouse, la plateforme de bioinformatique de l'INRA met à disposition les ressources nécessaires pour traiter nos données, le temps de calcul alloué aux utilisateurs est illimité.

Lorsque ce sera nécessaire, le diagnostic moléculaire de demain sera manifestement basé sur une analyse fine de la liste des agents pathogènes proposés par le test moléculaire et possiblement impliqués dans la maladie. Une fois la liste proposée, il appartiendra au clinicien d'en extraire les informations les plus pertinentes à ses yeux en fonction du contexte clinique et de son expérience avant de poser un diagnostic et d'en déduire le traitement adéquat.

2 Validation des tests moléculaires et contrôle qualité

Le développement de nouveaux tests de diagnostic vétérinaire doit être associé à une approche robuste de validation des méthodes. Les tests de diagnostic *in vitro* en santé humaine sont soumis à des cahiers des charges de validation qui diffèrent selon le pays de commercialisation du produit ; par exemple, les produits commercialisés sur les continents européen et américain doivent répondre respectivement aux règles de la commission européenne et de la *Food and drug administration*. En médecine vétérinaire, en ce qui concerne

les maladies à plan d'urgence et les maladies à déclaration obligatoire, « les normes, lignes directrices et recommandations s'appliquant aux tests de diagnostic sont publiées dans le Manuel des tests de diagnostic et des vaccins pour les animaux terrestres (Manuel terrestre) et dans le Manuel des tests de diagnostic pour les animaux aquatiques (Manuel aquatique) » (<http://www.oie.int>).

L'OIE (Organisation mondiale de la santé animale) propose une liste regroupant les objectifs que les kits de diagnostic doivent remplir :

- Prouver le statut indemne d'un pays, d'une zone géographique ou d'un troupeau vis-à-vis d'un agent infectieux, avec ou sans vaccination.
- Certifier l'absence d'infection ou d'agent infectieux chez un individu ou un produit commercial.
- Participer à l'éradication de maladies.
- Confirmer le diagnostic de cas cliniques suspects.
- Estimer la prévalence d'une maladie infectieuse ou son exposition afin de faciliter l'analyse.
- Déterminer le statut immunitaire d'individus ou de populations après vaccination.

La validation d'un test moléculaire doit être réalisée avant son utilisation en routine, cependant elle devient caduque et doit être à nouveau réalisée en cas de modification du processus d'analyse, avec par exemple l'utilisation d'un nouvel équipement présentant des caractéristiques différentes de celui employé pour la validation initiale.

Le contrôle qualité d'un test consiste en la combinaison d'un contrôle qualité interne (IQC : *Internal Quality Control*), d'un contrôle interne de répétabilité (IQA : *Internal Quality Assessment*) et d'un contrôle externe de performance (EQA : *External Quality Assessment*). La Figure 3 illustre l'interprétation des résultats issus de contrôle qualité.

- Le contrôle IQC consiste en la recherche systématique, lors du test, d'un matériel connu dont la quantité a été calibrée, cela permet de déterminer en temps réel le niveau de performances du test et de valider les analyses quotidiennes. Les contrôles internes aux

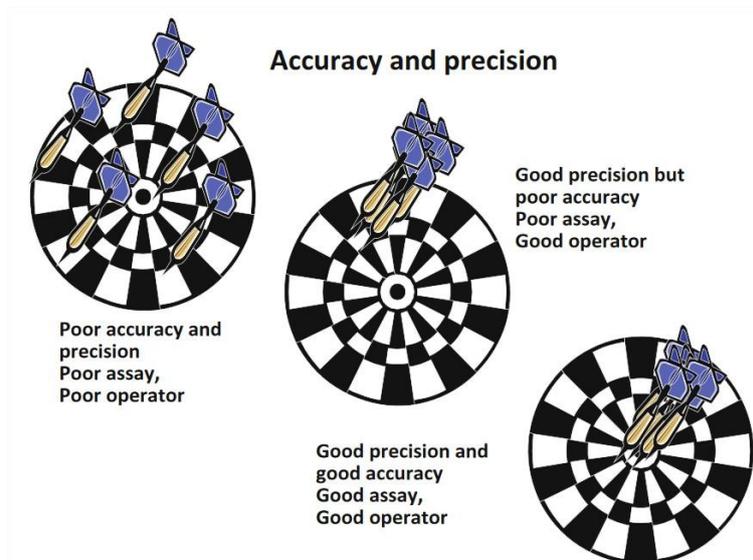


Figure 3 Interprétation des résultats issus de contrôle qualité en matière de performance analytique. Illustration issue de Methods in molecular biology 1247.

kits commerciaux pouvant varier d'un lot à l'autre, il est conseillé que le laboratoire d'analyse utilise un second contrôle indépendant du kit et produit en interne.

- Le contrôle IQA permet de s'assurer qu'un test réalisé deux fois sur un même échantillon générera des résultats cohérents. Typiquement les tests IQA doivent représenter entre 0.5 et 1% du volume total d'analyse. Toutes les 100 ou 200 réceptions, un échantillon sera divisé en deux parties, une partie pour l'analyse prescrite et une partie pour l'analyse IQA (les deux tests étant identiques). La comparaison des résultats fournira une information sur les capacités de répétabilité du test.
- Le contrôle EQA est quant à lui organisé par un laboratoire extérieur à ceux effectuant les tests diagnostiques. Des échantillons strictement identiques sont envoyés pour analyses aux laboratoires de diagnostic et les résultats provenant des différents sites seront comparés. Chaque participant pourra ainsi situer son niveau de performance et adapter sa méthodologie en cas de variations majeures vis-à-vis du résultat attendu.

Ces trois tests permettent de donner une indication sur la validité du test tout juste réalisé (IQC), sur la performance de chaque étape analytique (IQA) et sur le positionnement du laboratoire vis-à-vis d'autres en matière de précision d'analyse (EQA).

3 Le cas particulier du contrôle de sûreté des produits biologiques

Le contrôle et la prévention des maladies animales repose, comme en santé humaine, sur la vaccination des populations qui reste la première méthode à mettre en place. Le statut sanitaire d'un élevage dépend en partie de son programme de vaccination qui doit être réalisé à partir de vaccins vétérinaires de bonne qualité, efficaces et présentant une innocuité parfaite.

En France, la fabrication des vaccins et médicaments immunologiques vétérinaires est soumise aux normes élaborées par la Commission européenne de Pharmacopée. Cette commission élabore la pharmacopée européenne (*Ph. Eur.*) qui est le recueil des normes communes qui régissent, à l'échelle européenne, le contrôle de la qualité des vaccins à usage humain ou vétérinaire. L'analyse de risque des vaccins vétérinaires est basée sur de l'assurance qualité lors du processus de fabrication des vaccins. Il existe un risque de contamination des vaccins par des agents infectieux potentiellement introduits par l'importation d'animaux vivants ou de produits animaux. Le contrôle de chaque support de culture et de tout produit animal est nécessaire pour éviter tout risque de contamination durant le processus de fabrication.

Ce contrôle de qualité concerne notamment l'identité de la souche vaccinale et de son intégrité génétique pour s'assurer de l'absence de dérive génétique. De même, la suspension vaccinale ne devra présenter aucune contamination par un micro-organisme étranger autre que celui constituant le vaccin.

Le développement d'un vaccin viral vétérinaire est réalisé à partir d'un lot de semence virale qui constitue la matrice initiale du vaccin contenant l'agent vaccinal. Le lot de semence virale est en fait le substrat de production servant de base à la constitution du principe actif qui a son tour servira à la production du lot de vaccin.

Le développement initial du vaccin est composé d'étapes clés consistant en :

- La validation de l'identité de la souche vaccinale
- La confirmation de l'atténuation de la souche dans le cas d'un vaccin vivant
- La vérification de la pureté microbiologique de la préparation avec la recherche de 21 agents étrangers pour les vaccins aviaires

En plus de tests biologiques, l'intégrité de l'identité de la souche vaccinale est assurée par la quantité de passages sur culture cellulaire qui est limitée au nombre de cinq, cela permet de conserver une uniformité de production.

Certains vaccins étant produits sur cultures cellulaires, ces dernières sont également soumises à des contraintes d'utilisation répertoriées dans le Manuel Terrestre de l'OIE de 2008 :

- En règle générale le nombre de passages autorisés est compris entre 20 et 40
- Chaque stock de cellules doit être caractérisé sur le plan génétique
- Des études *in vivo* doivent confirmer l'absence de toute activité oncogénique
- L'absence de virus, bactéries, champignons et mycoplasmes des stocks de cellules doit être contrôlée.

Des cellules primaires peuvent également servir de support de culture de virus vaccinal, les cellules sont qualifiées de primaires lorsqu'elles sont issues d'un tissu « frais » et ont subi moins de 10 passages. Dans le cas de produits dédiés à la gamme aviaire, les cellules primaires sont issues d'œufs exempts d'organismes pathogènes spécifiques (EOPS) provenant d'élevages respectant des règles de biosécurité strictes. Les cellules primaires provenant d'animaux non EOPS sont quant à elles soumises à une analyse de pureté microbiologique.

Les animaux servant de support de production de virus vaccinal doivent également présenter le statut EOPS, cela concerne le virus de la maladie hémorragique du lapin qui est cultivé sur foies de lapins vivants.

L'innocuité de la préparation doit être vérifiée dès les premiers stades de développement du vaccin. Cette vérification est réalisée suite à l'injection d'une dose unique de vaccin, d'une surdose et de l'injection répétée de doses uniques. Une surdose est dix fois supérieure à une dose unique pour un vaccin vivant et deux fois supérieure à une dose unique pour un vaccin inactivé.

Les vaccins vivants présentent le risque d'être excrétés par l'hôte vacciné puis transmis aux animaux en contacts. Cette transmission peut être à l'origine du développement de la maladie chez l'animal contact si la préparation vaccinale contient des sous-populations non suffisamment atténuées. Afin de vérifier l'absence de populations virulentes, un groupe d'animaux cibles doit être inoculé avec le lot de semence primaire, puis le virus est récupéré dans les tissus avant d'être inoculé à nouveau sur un groupe d'animaux et ainsi de suite. Ce

passage en série doit être réalisé quatre fois au minimum puis l'isolat est récupéré sur les animaux vaccinés et est caractérisé de la même façon que pour le lot de semence primaire.

3.1 Tests de pureté

La pureté est déterminée par la recherche d'un panel de micro-organismes. Ces recherches sont effectuées sur les lots de semence primaire, les cellules primaires, les œufs embryonnés et tous les constituants d'origine animale ne pouvant pas être stérilisés.

Le contrôle de pureté virale dans le cadre des vaccins aviaires est à réaliser selon les directives de la Pharmacopée Européenne dont les détails sont répertoriés en annexe 1.

Brièvement, sur les matières premières d'origine animale, les lots de semence ou le produit fini, la recherche d'agents étrangers se fait selon les techniques suivantes :

3.1.1 Recherche directe spécifique :

Culture sur cellules spécifiques de l'agent recherché puis recherche d'Ag spécifique par ELISA, fixation du complément ou immunomarquage.

3.1.2 Recherche directe non spécifique :

Culture sur œufs de poules embryonnés (suivi de mortalité embryonnaire, recherche d'agents hémagglutinants) ou sur cellules rénales de poulets (suivi d'effet cytopathique (ECP)).

3.1.3 Recherche indirecte spécifique :

Inoculation de dix poussins EOPS âgés de deux semaines puis suivi pendant cinq semaines avec recherche d'Ac par immunodiffusion sur gélose, ELISA, inhibition de l'hémagglutination, séroneutralisation ou immunomarquage.

Les techniques recommandées par la Ph. Eur. pour réaliser les contrôles de qualité cités ci-dessus sont critiquables dans la mesure où ce sont des techniques ancestrales dont la sensibilité et la spécificité présentent des niveaux de performance inférieurs à ceux des tests moléculaires modernes. De plus, ces techniques ne respectent pas le principe de bioéthiques des

3R, elles sont longues à mettre en place et sont onéreuses. Après avoir gradué les souffrances subies par les animaux de laboratoire d'Angleterre, W.M.S. Russell et R.L. Burch ont, en 1959, proposé un programme de mise en place et de développement de lignes directrices dites "humaines", appelé la "règle des 3 R" (Russell and Burch, 1959). Elle comprend les points suivants :

Reduce (Réduire) le nombre d'animaux en expérimentation

Refine (Raffiner) la méthodologie utilisée, ce qui implique la notion de points limites (critères d'interruption)

Replace (Remplacer) les modèles animaux.

Le remplacement des techniques classiques par des tests moléculaires apparaît évident. Les efforts déployés pour le développement des tests moléculaires dédiés au diagnostic vétérinaire sont applicables à l'industrie du vaccin étant donné que les besoins en détection de micro-organismes des deux secteurs sont convergents.

4 Verrous méthodologiques liés à la biologie moléculaire moderne et à la manipulation de séquences.

Comme évoqué précédemment, les verrous méthodologiques liés aux tests moléculaires à haut-débit sont nombreux. Les techniques étant très récentes et évoluant très rapidement, il est difficile d'acquérir suffisamment de recul pour standardiser les protocoles. Dès lors que l'expérience est acquise, il arrive que la technique soit déjà obsolète. De plus, du fait que l'on retrouve des verrous méthodologiques à chaque étape de la démarche, l'identification des points critiques reste complexe.

Ci-après, les difficultés identifiées sont listées selon la chronologie de la démarche d'analyse virologique :

- Le stockage des prélèvements est d'autant plus important qu'il constitue la première étape à gérer avec attention. Du soin apporté au prélèvement découlera la qualité du résultat. Les tests modernes d'analyse rapide (de type bandelette) ont fait de l'analyse biologique un procédé simple et robuste utilisable sur le terrain sans avoir à prêter attention aux conditions d'échantillonnage. Cependant, il ne faut pas perdre de vue que des tests puissants et fins

nécessitent que le prélèvement soit fait avec le plus de précautions possible et qu'il soit extemporanément stocké et acheminé vers le laboratoire dans des conditions de froid irréprochables. L'exposition à des nucléases et protéases d'un prélèvement pauvre en micro-organismes aura des conséquences irréversibles sur la qualité de l'analyse. Aujourd'hui, des techniques de conditionnement en milieu de conservation et stabilisation existent, le transport sécurisé d'échantillons en azote liquide est également possible grâce aux *dry-shippers* (bombe contenant un matériau qui absorbe l'azote liquide).

- La quantité initiale en particules virales est un facteur primordial puisqu'une faible concentration dans un grand volume d'échantillon nécessitera un traitement fastidieux pour tenter de concentrer les particules virales. Des protocoles basés sur de l'ultracentrifugation (Liais et al., 2014) ou de la filtration tangentielle (Cai et al., 2015) ont déjà été décrits, cependant de tels procédés suscitent légitimement des soupçons sur la perte potentielle de matériel d'intérêt. Dans leur étude, Cai et al. annonce des taux de rendement autour de 40% concernant la récupération des particules virales suite à la filtration tangentielle.

- L'extraction du matériel génétique suit deux grandes techniques largement employées dans tous les laboratoires. La première est basée sur l'utilisation de phénol:chloroforme, la seconde fait appel à des colonnes d'affinité pour les acides nucléiques. L'extraction d'acides nucléiques en utilisant un mélange de phénol-chloroforme est une méthode ancestrale qui a fait ses preuves dans bon nombre d'applications (Bergallo et al., 2006). Ce procédé sépare les acides nucléiques et les protéines puisque, l'ADN et l'ARN étant solubles, ils vont être retrouvés dans la phase aqueuse du mélange alors que les protéines seront dénaturées et seront dans la phase organique phénol:chloroforme. Les acides nucléiques seront ensuite précipités avant d'être lavés pour pouvoir être analysés. Cette méthode présente l'avantage de produire des acides nucléiques de très bonne qualité et intègres, cette dernière notion est capitale si l'on souhaite conserver des génomes de plusieurs centaines de kb. Toutefois, cette technique est fastidieuse, longue à réaliser et fait appel à des solvants organiques à manipuler avec précautions. De plus, les étapes de lavage sont primordiales car la présence de résidus d'éthanol ou de phénols est à proscrire pour des analyses de biologie moléculaire.

Afin de simplifier les protocoles d'extraction, les industriels ont développé des techniques d'extraction sur colonnes de silice que l'on retrouve aujourd'hui dans tous les laboratoires. Cette procédure est simple, rapide et adaptée pour l'automatisation. Les extractions sur

colonnes peuvent être réalisées dans de nombreuses applications mais elles seront à éviter si l'on cherche à obtenir de longs fragments d'ADN (jusqu'à plusieurs centaines de kb) de très bonne qualité. De plus, les colonnes de silice utilisées dans les colonnes peuvent être source de contamination en acides nucléiques (Smuts et al., 2014).

- Lorsque les ADN et/ou ARN sont extraits, deux voies s'ouvrent. La première voie est minimaliste et consiste à soumettre directement les ADN à la préparation des banques à séquencer. Si le projet porte sur les ARNs, une transcription inverse peut-être réalisée. Cette option est idéale dans le sens où le risque d'introduction de biais reste très faible car il n'y a potentiellement qu'une transcription inverse. Cependant, pour pouvoir emprunter cette voie il faudra que l'ADN à séquencer soit en quantités suffisantes pour pouvoir s'affranchir d'une étape d'amplification, les quantités d'ADN à fournir pour la préparation des banques dépend de la chimie de séquençage. Par exemple pour un protocole Nextera d'Illumina, la quantité d'ADN de départ requise est de 50 ng tandis qu'un protocole Oxford Nanopore nécessite 1 µg d'ADN (<https://nanoporetech.com/>).

- Deux cas de figure nécessiteront une amplification des ADN : lorsque la quantité d'ADN de départ est insuffisante et/ou lorsqu'il y a un besoin de multiplexage des échantillons avec l'introduction d'étiquettes nucléotidiques aux extrémités des amplicons à séquencer. Il faut cependant garder à l'esprit que l'amplification des ADN par PCR est une source d'introduction de biais liée au taux d'erreur de l'ADN polymérase employée (Acinas et al 2005).

- La préparation des banques d'ADN pour le séquençage à haut débit est soumise aux recommandations du constructeur, chaque fabricant de séquenceur propose ses propres kits de préparation de banques d'ADN. D'autres fabricants de kits de biologie moléculaire proposent également des kits de préparation de banques d'ADN à séquencer. Les prérequis nécessaires à l'utilisation de ces kits définiront la méthode de préparation des ADN en amont de cette étape. En effet, la quantité d'ADN ou d'ARN ou bien la synthèse d'ADNc simple brin ou double brin sont des conditions imposées par le kit et peuvent représenter un obstacle méthodologique.

- Le séquençage à haut débit est une étape véritablement technique, spécifique de chaque séquenceur et nécessitant une expertise dans l'utilisation des séquenceurs NGS qui sont basés sur des technologies extrêmement pointues. D'autant plus que l'offre est en perpétuelle évolution (Figure 4).

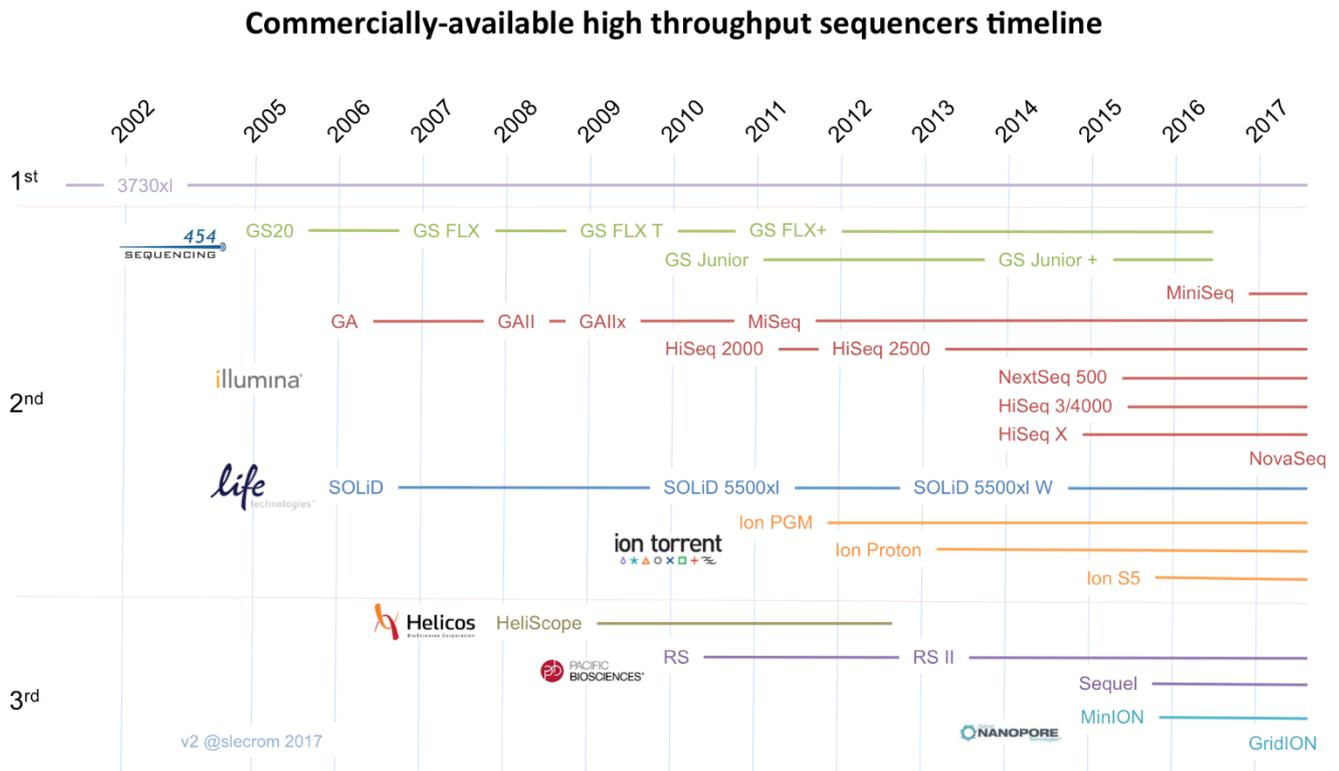


Figure 4 Aperçu de l'offre en séquenceurs NGS au fil du temps.

Les fabricants Roche et Thermo Fisher Scientific ont tour à tour proposé des « séquenceurs de paille » avec, respectivement, le 454 GS Junior et l'Ion PGM. L'idée était de permettre aux scientifiques de s'équiper en machines NGS sans être dépendants de plateformes technologiques et être ainsi autonomes. Le point négatif de cette stratégie est le besoin en maintenance du système qui devra être absorbé par le laboratoire ou être sous-traité par le fabricant à des tarifs prohibitifs, il sera difficile de rentrer dans une logique d'amortissement de l'équipement, à moins que le débit d'analyse soit très important (laboratoire de biologie clinique). Oxford Nanopore propose aujourd'hui un séquenceur nomade de la taille d'un téléphone portable pour moins de 900€, les tarifs de consommables sont distribués entre la puce

de séquençage (à partir de 500€) et le kit de préparation de la banque d'ADN à séquencer (570€ pour le plus basique).

- Le séquenceur NGS génère un signal brut produit par le brin d'ADN en cours de synthèse, ce signal diffère selon les technologies de séquençage et peut être issu :
 - Du signal de lumière blanche produit par une cascade réactionnelle faisant suite à l'incorporation d'un nucléotide (Roche).
 - Du signal d'un fluorochrome couplé au nucléotide incorporé (Illumina, PacBio).
 - De la modification du pH de la réaction suite à l'incorporation d'un nucléotide (Thermo Fisher Scientific).
 - De la modification du potentiel électrique de la membrane à travers laquelle le brin d'ADN à lire est guidé (ONT).

Une fois obtenus, ces différents types de signaux doivent être convertis en séquence ADN, le *base calling*, il est réalisé :

- En associant la temporalité du signal avec le type de nucléotide présent dans le milieu réactionnel (séquençage séquentiel, Roche, Thermo Fisher Scientific).
- En convertissant le signal du fluorochrome en nucléotide, quatre couleurs donc quatre possibilités (Illumina, PacBio).
- En comparant le signal électrique à une base de données comportant les valeurs de modification du potentiel électrique propres à chaque combinaison de nucléotides (*k-mers* de 5 nucléotides) (ONT).

Actuellement, le *base calling* est une étape extrêmement « gourmande » en ressources informatiques et notamment en ce qui concerne les données du MinIon qui nécessitent jusqu'à plusieurs semaines de calculs. Le fabricant promet de son côté un *base calling* réalisable en temps réel, cela ouvre de nouvelles perspectives quant à l'optimisation des *runs* de séquençage.

- La manipulation des données de séquençage demeure la « bête noire » du biologiste puisqu'il s'agit de bio-informatique. Les jeux de données à traiter sont si volumineux que leur traitement ne peut pas encore être réalisé sur un ordinateur de bureau standard à l'aide d'un logiciel « presse-bouton ». Les ressources de calcul nécessaires sont disponibles uniquement *via* un *cluster* de calcul, cela implique une utilisation des outils d'analyse en lignes de commande par l'intermédiaire d'un terminal. L'adaptation à la ligne de commande n'est pas si complexe sous réserve de ne pas être réfractaire à l'informatique. La difficulté réside néanmoins

dans le choix des algorithmes, leur paramétrage et la levée des bogues. Les deux derniers points nécessitent des compétences propres en bio-informatique disponibles au sein des plateformes de bio-informatique voire de génomique.

Le projet Galaxy dont le développement a été initié il y a une dizaine d'années est une interface graphique permettant de manipuler des jeux de données NGS tout en profitant des ressources de calcul d'un *cluster*, c'est l'alternative idéale à la ligne de commandes (Afgan et al., 2016; Giardine et al., 2005). Toutefois, l'accès à un *cluster* de calcul reste nécessaire et tous les algorithmes ne sont pas incrémentés dans Galaxy, il est donc important d'être en lien avec les développeurs Galaxy pour qu'ils ajoutent et adaptent à la demande les outils souhaités.

De la même manière que pour les technologies de séquençage, les outils de manipulation de séquences sont en perpétuelle évolution et peuvent potentiellement devenir rapidement obsolètes. Ces 14 dernières années, l'offre en outils de bio-informatique est passée d'un état quasi-anecdotique à une variété très riche (Figure 5).

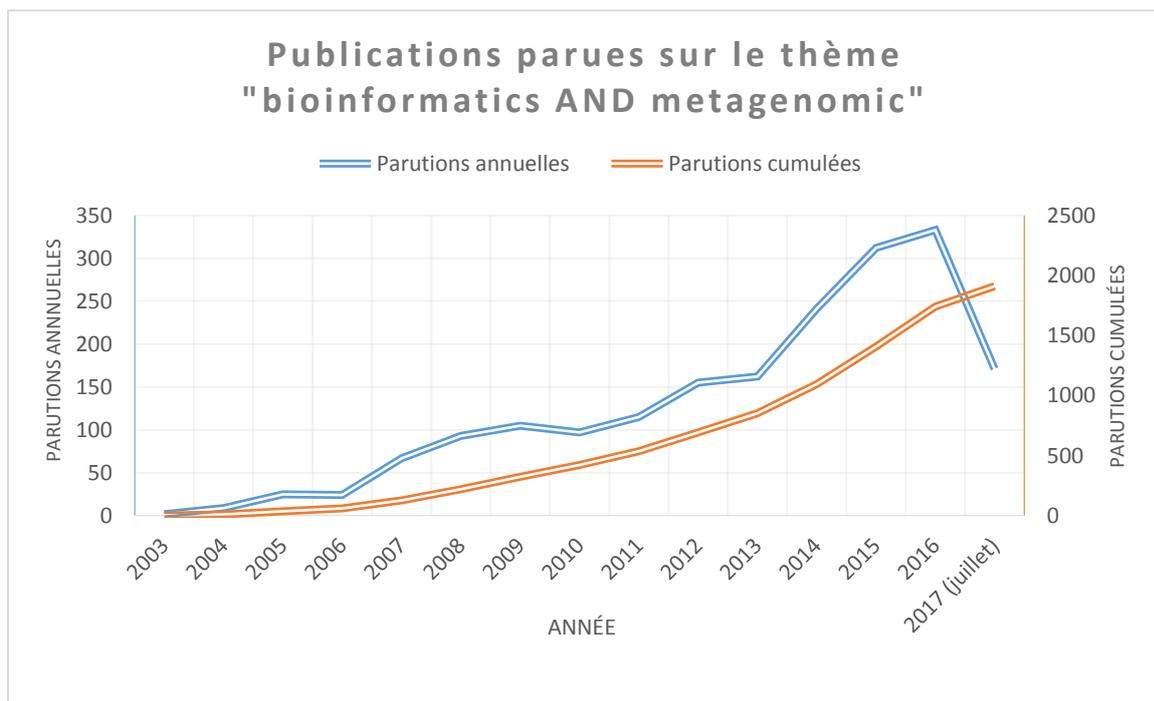
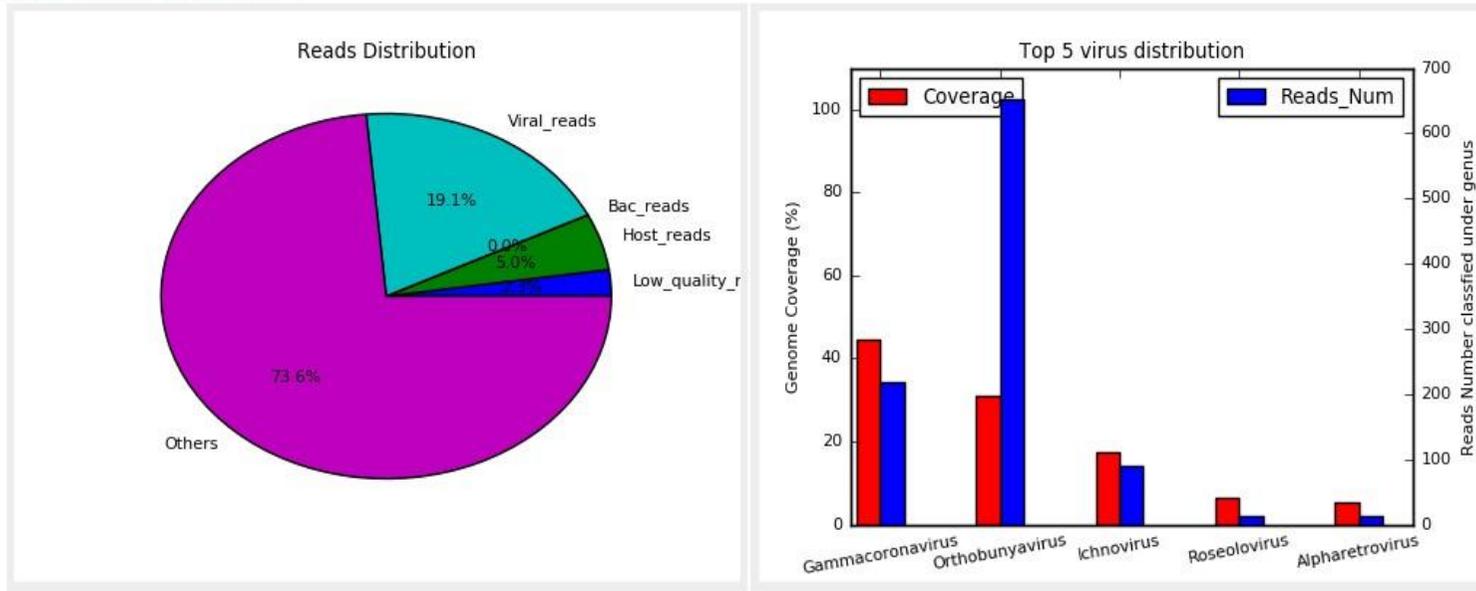


Figure 5 Évolution des parutions de publications traitant de bio-informatique et de métagénomique

Les bio-informaticiens captent également le marché des *pipelines* d'analyse qui sont très recherchés des biologistes, en effet, un outil tel que *Virus Identification Pipeline* (Li et al., 2016b) est idéal car il permet une analyse complète et fiable (Delpuech, 2017) de jeux de données de métagénomique pour la description du virome. L'utilisation de cet outil à l'aide de deux lignes de commandes (dont une générée par l'outil à copier puis coller) en fait un *pipeline* facilement utilisable par les biologistes. De plus, le résultat est généré sous forme de tableaux et de graphiques lisibles et directement exploitables (Figure 6).

Reads Distribution



Summary Report

Species	Genus	GI	%Coverage	Reads_hit
Guinea fowl coronavirus GfCoV/FR/2011 complete genome	<u>Gammacoronavirus</u>	807045770	44.7100	216
Shamonda virus N and NSs genes, segment S, genomic RNA, isolate Ib An 5550	<u>Orthobunyavirus</u>	401829616	31.0600	652
Glypta fumiferanae ichnovirus segment C9, complete sequence	<u>Ichnovirus</u>	124378202	17.3500	13
Human herpesvirus 7, complete genome	<u>Roseolovirus</u>	51874225	6.5800	3

Figure 6 Capture d'écran d'un exemple de résultats généré par VIP (Li et al 2016).

Objectifs de recherche

Introduction à la partie expérimentale

Les travaux détaillés dans cette partie expérimentale répondent à des questions de recherche intégrées dans un projet de thèse CIFRE (Convention Industrielle de Formation par la Recherche en Entreprise) avec le laboratoire pharmaceutique Filavie. Ces questions gravitent autour de deux outils majeurs de la biologie moléculaire : la PCR quantitative en temps réel et le séquençage de nouvelle génération ainsi que de deux applications : le diagnostic des maladies virales animales et le contrôle des médicaments immunologiques vétérinaires.

Cette thèse s'inscrit dans la stratégie du laboratoire d'accueil et du laboratoire Filavie, porteur du projet de thèse CIFRE, de s'approprier des méthodes permettant de répondre à la fois à des questions de recherche cognitive (description des communautés d'agents pathogènes et de leur dynamique chez différents hôtes) et finalisées vers la création d'outils de diagnostic et de contrôle des produits biologiques. Les enjeux sont multiples :

- Détection d'agents pathogènes connus, inscrits sur une liste positive
- Détection sans *a priori* et caractérisation d'agents infectieux (pathogènes ou non) présents dans un milieu complexe (appareil respiratoire, tube digestif)
- Génotypage d'un agent infectieux connu, pathogène ou vaccinal

Le fil conducteur du projet de thèse consiste à développer et tester des procédés innovants de détection et caractérisation génétique des virus. Les perspectives offertes par ces technologies dans le domaine du diagnostic sont prometteuses puisque les coûts et les temps d'analyse mais aussi la sensibilité, la spécificité et les volumes réactionnels peuvent évoluer favorablement.

Ces techniques innovantes ont été appliquées à trois études portant sur :

- La description des co-infections respiratoires chez la dinde en France et au Maroc ; preuve de concept de l'utilisation de la PCR en format microfluidique Biomark de Fluidigm. L'objectif recherché dans ce volet consiste à détecter un panel connu d'agents infectieux (Croville et al. 2017, *In prep*)

- La mise en cause et identification du coronavirus agent de la maladie foudroyante de la pintade ; application de la métagénomique avec l'utilisation du séquenceur MiSeq d'Illumina. Il s'agit dans ce volet de rechercher sans *a priori* le(s) virus potentiellement associé(s) à une entité pathologique. (Liais et al., 2014)
- Le séquençage du génome complet d'un avipoxvirus à partir de lésions cliniques, preuve de concept de l'utilisation du séquenceur nomade MinION d'Oxford Nanopore Technologies. Ce volet vise à réaliser le séquençage complet d'un virus (en l'occurrence un poxvirus, grand virus à ADN) par séquençage sans *a priori*, en mettant à profit la capacité de cette technique à générer de très longues séquences, ce qui facilite l'assemblage d'un génome. (Croville et al., 2017, *In prep*).

Première partie expérimentale

Criblage d'agents infectieux respiratoires aviaires par PCR à haut débit

Le diagnostic syndromique en pathologie respiratoire voire entérique se fait par la recherche d'un panel d'agents infectieux déjà établi. Un diagnostic très ciblé sera concentré sur quelques agents seulement, mais ce panel est constitué de micro-organismes dont la présence est la plus probable sur la base du contexte clinique et épidémiologique. Cependant, dans le cas d'une situation épidémiologique atypique, la recherche d'une liste précise ne sera pas pertinente et il faudra alors avoir recours à une exploration large et sans *a priori* comme pour l'épisode d'intoxication alimentaire à *Escherichia coli* O104 :H4 de 2011 (Loman et al., 2013).

Le contrôle de pureté microbiologique de produits pharmaceutiques et notamment de vaccins est également basé sur la recherche d'un panel de micro-organismes qui est cette fois défini sur des bases réglementaires. Il y a dans ce cas la nécessité de répondre avec certitude sur l'absence ou la présence d'agents infectieux et cette réponse doit être étayée par la définition d'un seuil de détection analytique.

Cette recherche à large spectre présente l'inconvénient de nécessiter un haut débit d'analyse. En effet, le criblage de vingt échantillons par RT-PCRq en duplicatas avec un panel de dix couples d'amorces représente déjà un minimum de 400 réactions. Outre le coût d'analyse (environ 25cts d'€ par réaction pour la recherche de génome ADN et 70cts d'€ pour la recherche de génome ARN), le temps de mise en œuvre et l'utilisation de l'équipement de laboratoire sont conséquents et peu compatibles avec un débit d'analyses élevé.

Des solutions de diagnostic sont déjà proposées sur le marché en santé humaine avec par exemple le NxTAG® *respiratory pathogen panel* ou xTAG® *respiratory viral panel fast* de la société Luminex (Tx, USA) ou encore le FilmArray® *respiratory panel* de Biomérieux (Marcy-l'Étoile, France). Les kits Cypher One™ and VaxArray® ainsi que le FluChip-8g sont quant à eux développés par la société InDevR (Boulder, Co, USA) et sont dédiés au diagnostic d'infection à virus influenza. Ces kits sont développés pour la recherche des virus et bactéries les plus pertinents en pathologie respiratoire mais ce sont des technologies propriétaires non évolutives et non adaptées à la santé animale. Quelques kits de détection de panels existent en santé animale mais sont basés sur de la PCR multiplex Taqman et ne ciblent que quelques agents comme le kit ID Gene™ BVD/BD Triplex d'IDvet genetics (Montpellier, France).

Aujourd'hui, il y a un réel besoin d'outil de détection d'agents pathogène présentant les caractéristiques suivantes : compatible avec une analyse à haut débit, évolutif pour permettre une adaptation en temps réel aux cibles à rechercher, robuste sur les plans de la spécificité et de la sensibilité.

Une des solutions se trouve dans la miniaturisation des procédés pour réduire drastiquement les volumes de réactifs nécessaires. La société Fluidigm (Ca, USA) a développé le système Biomark HD qui est basé sur une automatisation des réactions de PCR en volumes réduits à l'échelle du nanolitre.

La technologie microfluidique permet, grâce à la réduction des volumes, de réaliser une quantité de réactions de RT-qPCR significativement supérieure à celle que l'on connaît avec des plaques de 96 ou 384 puits. Cette technologie innovante place le criblage d'échantillons à un niveau de rendement à haut débit. En effet, le choix se portera sur une puce 48.48 ou 96.96 permettant, respectivement, de cribler jusqu'à 48 échantillons avec 48 couples d'amorces sous la forme de 2304 réactions ou de cribler jusqu'à 96 échantillons avec 96 couples d'amorces soit 9216 réactions de PCR quantitative en temps réel. Chaque réaction est individuelle et correspond à l'analyse par RT-qPCR d'un échantillon par un couple d'amorces. La chimie employée sera basée sur l'intercalant EVA Green ou sur les sondes d'hydrolyse Taqman. Le fait de travailler en réactions individuelles et non en multiplex présente l'avantage indéniable de s'affranchir de réactions croisées non spécifiques au sein d'un même puits. De plus, tout défaut de spécificité est détectable car tous les témoins sont testés avec toutes les amorces.

La PCR à haut débit Biomark a par exemple déjà été appliquée au typage moléculaire de *Streptococcus pneumoniae* (Dhoubhadel et al., 2014) et a permis aux auteurs de s'affranchir de mise en culture pour réaliser une étude épidémiologique à grande échelle des maladies à pneumocoques chez les enfants en bas âge. Les auteurs soulignent notamment la possibilité de détecter plusieurs sérotypes en une seule réaction.

L'analyse de l'expression de gènes est également mesurable par RT-PCR Biomark comme dans l'étude de Sorg. Les auteurs ont suivi l'expression de 22 gènes impliqués dans la réponse immunitaire chez des cellules épithéliales primaires mammaires de bovin après stimulation par *Escherichia coli* ou *Staphylococcus aureus*. Les auteurs soulignent des niveaux de résultats similaires à ceux obtenus avec de la PCR en temps réel conventionnelle (Sorg et al., 2013).

L'étude décrite ici se positionne dans le cadre du diagnostic de maladies respiratoires de la volaille. Le panel d'agents à rechercher en première intention est constitué de 16 virus,

bactéries ou champignons. De ce fait, une puce 48.48 permettra d'analyser près de 48 échantillons avec une vingtaine de couples d'amorces ciblant les micro-organismes les plus pertinents. Il est également possible de réaliser un typage viral ou bactérien et de rechercher des marqueurs de résistance aux antibiotiques. Jusqu'à 24 couples d'amorces, il sera possible de dupliquer chaque analyse. Lorsque la chimie EVA Green est employée, l'analyse de la courbe de fusion apportera la capitale notion de spécificité d'amplification.

En première intention, ce sont des amorces de PCR sélectionnées dans la littérature qui ont été testées, cependant certains couples ne fonctionnaient pas dans nos mains et ont été remplacés par des amorces développées au laboratoire.

L'analyse des résultats semble aisée grâce à la *heat-map* générée par le logiciel, elle est directement reliée aux valeurs de Cq (cycle de quantification) obtenues (Figure 7).

La couleur du puits est directement corrélée à la valeur de Cq sans pondération en fonction du pic de dissociation, cela suggère que le logiciel est davantage adapté à l'interprétation de résultat Taqman que SybrGreen. Il est donc nécessaire qu'une analyse fine et précise des résultats soit réalisée à partir des résultats bruts avec la comparaison des valeurs de Tm de chaque point d'analyse par rapport au témoin positif.

L'utilisation du Biomark a démontré sa capacité à réaliser un criblage à haut débit tout en conservant les qualités de sensibilité et de spécificité de la PCR quantitative en temps réel. L'analyse systématique des contrôles positifs et négatifs permet de valider l'absence de réactions croisées ou bien de les détecter et de résoudre le problème par une analyse complémentaire (PCR ou séquençage).

Il est toutefois nécessaire de garder à l'esprit que l'interprétation des résultats par le clinicien et le biologiste est incontournable puisqu'il leur revient de les valider (ou non) selon la nature des signaux obtenus et selon le contexte épidémiologique.

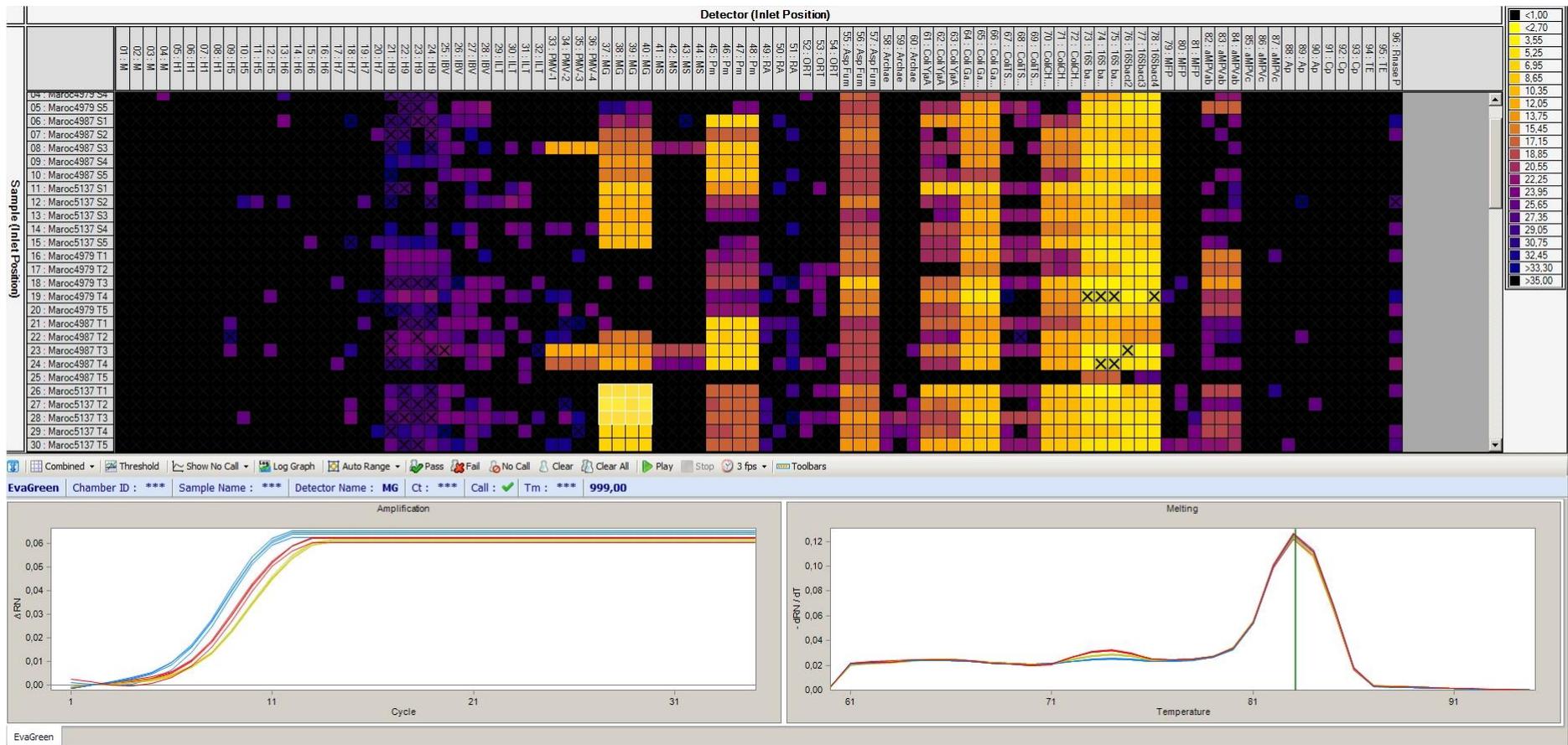


Figure 7 Le haut de la figure correspond aux résultats de RT-PCRq Biomark sous forme de heat-map, avec en haut à droite la légende du code couleur selon la valeur de C_q . La partie basse de la figure correspond aux courbes d'amplification et pics de dissociation correspondant aux puits de réaction sélectionnés.

L'étude décrite dans ce document propose un test de PCR en temps réel destiné au criblage à haut débit de 179 prélèvements respiratoires effectués chez des dindes de chair en France et au Maroc grâce à un panel de 19 couples d'amorces de PCR. Cette investigation a permis, dans un premier temps, d'estimer les charges de génomes d'agents pathogènes présents dans des échantillons cliniques et, dans un second temps, de décrire les co-infections pour apporter un diagnostic rapide du profil microbiologique de chaque lot d'élevage.

Ce criblage de plusieurs cibles en simultané a permis de mettre en avant les profils de co-infections. En France, des co-infections au métapneumovirus aviaire, paramyxovirus et aux mycoplasmes sont marqués. Ce sont *Pasteurella multocida*, *Mycoplasma gallisepticum* et le métapneumovirus aviaire qui dressent le profil de co-infection dans les échantillons du Maroc. Il y a clairement un effet géographique, favorisant l'affection par certains agents pathogènes.

Une étape de pré-amplification a permis d'enrichir le contenu en ADNc des échantillons, grâce à cet enrichissement, la détectabilité finale a été largement augmentée. Cette détectabilité accrue du Biomark est un atout à double tranchant : on peut espérer détecter des cibles de manière très fine mais cette performance peut entraîner des difficultés d'interprétation au niveau clinique.

Cet outil est remarquablement efficace lorsqu'il est associé à des couples d'amorces robustes et validées. Il y aurait un avantage à travailler en PCR Taqman du fait de la spécificité accrue du signal, Michelet et al. présentent l'utilisation d'un panel d'amorces et sondes de PCRq permettant la détection de 37 bactéries et parasites infectant *Ixodes ricinus* en Europe (Michelet et al., 2014). En effet, il n'y a pas de perturbation du signal liée, par exemple, à la production de dimères d'amorces ou de produits de PCR non spécifiques. Cependant, l'expérience en PCRq au laboratoire étant historiquement orientée vers de la PCR SYBR Green et les amorces dont nous disposons au moment de l'étude étant également développées pour du SYBR Green, c'est vers cette dernière chimie que nous nous sommes tournés.

La miniaturisation des équipements de PCR étant accompagnée d'une réduction exceptionnelle des volumes réactionnels (volumes de 6,7 nL, <http://www.fluidigm.com/biomark-hd-system.html>) entraîne automatiquement la raréfaction des génomes recherchés. L'alternative à cet écueil proposée par le fabricant est une pré-amplification des ADN en amont de la PCRq afin de compenser un niveau de détection d'amorces de PCRq potentiellement faible. Cette pré-amplification est en fait une PCR réalisée sur chaque échantillon avec le mélange d'amorces, c'est donc en quelques sortes une PCR multiplex. Cependant, cette étape n'est pas prise en considération alors qu'elle pourrait être à

l'origine d'un biais d'amplification. Afin de résoudre cette potentielle source de biais, Korenková et al. ont étudié différentes conditions de pré-amplification nécessaires à une distribution optimale des valeurs de Cq en partant de quantités d'ADN et ADNc adéquates. Les auteurs concluent que le succès de l'étape de pré-amplification est conditionnée par le nombre de cycles (devant être compris entre 15 et 18) et la concentration de l'échantillon (devant se situer entre 5 et 20 ng) car une trop forte quantité réduirait les quantités de réactifs nécessaires au bon déroulement de la réaction. Enfin, les auteurs démontrent que la variabilité induite par l'étape de pré-amplification est inférieure à celle causée par la réaction de rétro-transcription (Korenková et al., 2015).

A BROAD-RANGE PANEL ASSAY FOR AVIAN RESPIRATORY PATHOGENS ON A NANOFLUIDIC PCR PLATFORM

Guillaume Croville¹, Charlotte Foret¹, Pauline Heuillard¹, Alexis Senet¹, Mattias Delpont¹,
Mohammed Mouahid², Mariette Ducatez¹, Faouzi Kichou² and Jean-Luc Guérin^{1*}

¹IHAP, Université de Toulouse, ENVT, INRA, Toulouse, France

²Institut Agronomique et Vétérinaire Hassan II, Rabat, MOROCCO

*Corresponding author : jl.guerin@envt.fr / +33 608 363 147

ABSTRACT (219 words)

Respiratory syndromes (RS) are among the most significant pathological conditions in food animals and are caused by complex coactions of pathogens and environmental factors. In poultry, low pathogenic avian Influenza A viruses, metapneumoviruses, infectious bronchitis virus, infectious laryngotracheitis virus, *Mycoplasma spp.*, *Escherichia coli* and/or *Ornithobacterium rhinotracheale* (ORT) in turkeys, are considered as key co-infectious agents of respiratory syndromes. *Aspergillus sp.*, *Pasteurella multocida*, *Avibacterium paragallinarum* or *Chlamydia psittaci* may also be involved in respiratory outbreaks.

An innovative quantitative PCR method, based on a nanofluidic technology, has the ability to screen up to 96 samples with 96 pathogen-specific PCR primers, at the same time, in one RTqPCR run. This platform was used for the screening of avian respiratory pathogens: 16 respiratory agents, including viruses, bacteria and fungi potentially associated with respiratory infections of poultry were targeted. Primers were designed and validated for SYBR green real-time quantitative PCR (RTqPCR) and subsequently validated on the Biomark high throughput PCR nanofluidic platform (Fluidigm©). As a clinical assessment, tracheal swabs were sampled on turkeys showing respiratory syndromes and submitted to this panel assay. Beside systematic detection of *E. coli*, avian metapneumovirus, *Mycoplasma gallisepticum* and *M. synoviae* were frequently detected, with distinctive co-infection patterns between French and Moroccan flocks. This proof-of-concept study illustrates the potential of such panel assay for unveiling respiratory co-infections profiles in poultry.

INTRODUCTION

Respiratory syndromes are among the most frequent and significant pathological conditions in farm animals. In poultry species, they are also associated with egg-drop syndromes in layers and breeders, resulting in losses in production of both eggs and day-old-chicks. The control of respiratory syndromes relies on a massive metaphylactic use of antibiotics, most of the time without any sound knowledge on pathogens involved (1, 2). Routine diagnostic procedures are based on classical bacteriology and molecular tests targeting a very limited number of viruses and mycoplasmas. A key challenge in the diagnostic and management of respiratory infections is dissecting coinfections. So far, very few panel assays are available in veterinary medicine (3, 4). In human medicine, the increasing availability of panel assays has dramatically changed the knowledge of respiratory infectious agents, and subsequently may impact their therapeutic management with rational bases of antibiotics prescription.

The inherent limitations of culture-based detection methods have resulted in development of innovating pathogen detection and quantification methods. The objectives of these platforms are to speed-up the detection process and to improve accuracy. The recent devices are nucleic acids based assays allowing amplification and detection of pathogens, which reduces detection times and biological samples requirements.

In field samples, difficulties come from pathogen concentration which can be low, a large number of other interfering species and inhibitor factors complicating pathogen identification and quantification.

The identification of pathogens through genetic analysis techniques is becoming more conventional by use of real-time quantitative PCR (RT-qPCR). To bypass low concentrations, amplification performed during RT-qPCR allows detection of few DNA copies.

Miniaturizing the PCR process induces a reduction in sample and reagent volumes required and allows processing of thousand samples in one PCR run. High-throughput quantitative PCR becomes therefore affordable and conceivable as a routine test.

A commercial microfluidic based PCR device, Biomark™ 48.48 or 96.96 Dynamic Array, has been developed by the Fluidigm Corporation (San Francisco, CA). This platform has the specific capability to screen up to 96 samples with 96 pathogen-specific PCR primers on a 96.96 chip (9216 reactions) or to screen 48 samples with 48 pathogen-specific PCR primers on a 48.48 chip (2304 reactions), at the same time, in one RTqPCR run. The main advantage of this microfluidic chip is the large number of reactions that can be run in parallel and the fact that this is not a multiplex test, which means that PCR primers interferences are not possible.

In this paper, we developed a panel assay for 16 avian respiratory pathogens on the Biomark platform. As a proof-of-concept, we used this panel assay to screen clinical samples from turkey flocks and evidenced unprecedented comprehensible pathogens profiles.

RESULTS

Primers validation on conventional real-time cycler. A set of primers was designed *in silico* and submitted to a first validation on a LC480 thermocycler (Table 1). Each cross-reaction test performed with the primers listed in table 1 (**5-17**) was negative. The plasmids corresponding to in-house designed primers and some corresponding to primers selected in the literature targeting aMPV, AIV and Ap were tested with their dedicated primers on a range of concentrations from 10^1 to 10^6 copies per reaction. PCR primers efficiencies were directly obtained from the LightCycler 480 on the basis of the plasmids standard curves. Efficiency

values were comprised in a 1.75-2.25 interval, corresponding to a 75-125% efficiency; these values and limits of detection (LOD) are listed in table 2.

PCR primers validation on the Biomark platform. After validation on the LightCycler480 platform, the selected PCR primers were subsequently tested on the Biomark platform. Figure 1 provides an overview of the positive controls tested on the Biomark platform. Nineteen PCR primer pairs generated expected amplification curves whose quantification cycle (C_q) values were below 30. The melting peaks temperature of the positive controls were considered as standard values for validation of the specificity of the reaction. Therefore, the field samples T_m were validated if they were in a range of $\pm 1^{\circ}\text{C}$ from the reference values. Figure 2 shows an example of specific amplification curves and melting peaks obtained with avian metapneumovirus primers on field samples. At this stage, a total of 19 PCR primer pairs were validated (Table 1) using standard plasmids and cDNA in order to check that cross-reactions did not occur between PCR primers and non-specific targets (data not shown).

Field samples screening on the Biomark platform. As a proof-of-concept, 178 field samples, corresponding to 17 turkey flocks from France or Morocco facing acute respiratory signs were screened using the Biomark respiratory panel: a synthetic report of screening results is presented in table 3 and the comprehensive set of data is available as supplementary data (*Supplementary Table 2*).

Screening for avian respiratory viruses. All samples, from both countries, tested negative for broad-influenza A assay targeting M gene. Fifty-seven/178 samples tested positive for avian metapneumovirus and 29/178 were found positive for Paramyxovirus-1. Further sequence

analysis showed that these PMV-1 clustered in La Sota Clone 30 genotype, suggesting that they actually correspond to vaccine viruses (data not shown). Surprisingly, 3 Moroccan turkeys tested positive for *avian infectious laryngotracheitis virus* (ILTV).

Screening for avian respiratory bacteria. 128/178 turkey flocks tested positive for *Ornithobacterium rhinotracheale*, 44/178 for *Mycoplasma gallisepticum*, 39 for *Mycoplasma synoviae*, 33 for *Bordetella avium*, and 3 for *Riemerella anatipestifer*.

Birds from Morocco only tested positive for *Pasteurella multocida* (41/88) and *Avibacterium paragallinarum* (13/88), only one bird testing positive for *Chlamydia psittaci*.

A specific screening was performed for *Escherichia coli* so as to define genotypes (9): *E.coli* affiliated to B2 genotype (clustering most avian pathogenic *E.coli*) was detected in as many as 102/178 birds, *i.e.* 16/17 flocks.

Screening for avian respiratory Fungi. All French and Moroccan turkeys tested positive for *Aspergillus fumigatus*.

Comparative analysis of coinfection patterns. Coinfection profiles on French and Moroccan poultry flocks are displayed in Table 3. Coinfections could be identified in all submitted flocks. While avian metapneumovirus, *Ornithobacterium* and *E.coli* were widespread in turkeys from both countries and should be controlled at a global scale, *Mycoplasma synoviae* seemed specifically prevalent in France and *Pasteurella multocida* in Morocco, respectively.

DISCUSSION

We developed a panel assay targeting all the significant poultry respiratory agents in our knowledge (18-20). A set of 19 primers was selected and validated on a nanofluidic high throughput PCR platform. This panel assay was validated on reference strains and then, on a set of field samples as a clinical proof-of-concept. This panel assay was shown at least as sensitive and specific as real-time PCR and was able to provide in a single assay a comprehensive picture of viral and bacterial coinfections. The benchmark between microfluidic and conventional real-time PCR platforms showed similar PCR detection thresholds and cross reactivity. The main specific features of Fluidigm PCR is the Dye (Eva green *versus* SYBR green) and the volume of reaction (9 nL *vs.* 5 to 25 μ L). However, it has been previously shown that the results obtained on the Biomark platform show a very good concordance with conventional real-time PCR (21).

This respiratory panel assay is *not* based on a multiplex approach: a monoplex real-time PCR test is performed in each well and may provide quantitative data based on a full range of standards. The loads of each pathogen were not further analyzed here, but may be very helpful to the clinician to evaluate their respective pathological relevance.

The design of the platform is based on a systematic screening of each sample with all the primers pairs, which led necessarily to unexpected results: we had to exclude some PCR primers designed and validated in published reports, since they cross-reacted with other targets that had probably not been tested so extensively. In the same way, a contamination event during the process would be clearly identified as a whole line of samples (or primers) would generate inconsistent positive signals

The main advantages of the Fluidigm Biomark platform are (i) its microfluidic format, allowing as many as 48 to 96 PCR assays per sample using as few as 5 µl of nucleic acids and (ii) its versatility, making easy any updating or addition of primers in case of emergence of viruses or variants or need for any further typing. Indeed, as a first attempt, the respiratory panel assay presented here is mostly based on broad-range primers, designed to provide a general knowledge on the pattern of pathogens. The next version of this respiratory panel shall detect specific variants, virulence or antibiotics resistance markers.

Respiratory infections are the first indication for antibiotic treatments in food animals, most of the time without any supportive microbiological clinical data. They result from complex and versatile combinations of environmental and infectious factors and can hardly be elucidated only on the basis of clinical or necropsic signs. In human medicine, the recent introduction of panel assays has dramatically improved the knowledge on viruses involved in acute respiratory syndroms and is helpful in decision-making to define better treatment protocols (22, 23); In veterinary medicine, the main limit in routine diagnostic of respiratory diseases is the unavailability of such molecular tools allowing rapid, easy and affordable screening for a panel of pathogens, adapted for bovine, porcine or poultry species. This knowledge on pathogens, including their subtype and load, would be critical to define more accurate vaccine or therapeutic protocols and reduce the use of antibiotics. Molecular assays make also possible to analyze samples deposited on neutralizing transport devices such as Flinders Technology Associates (FTA) cards, allowing completely safe international shipments of nucleic acids for implementation of surveillance programs at a global scale (24).

As a clinical proof-of-concept study, we screened tracheal samples from French and Moroccan poultry facing acute respiratory syndromes. As expected, coinfections were widely detected in most flocks, including viruses and bacteria in most cases (25). *Escherichia coli* was detected in almost all birds, the set of primers (9) allowing the discrimination of phylogroups and the

identification of B2 strains, which are clustering most of avian pathogenic *E.coli* (26). *Ornithobacterium rhinotracheale* is also very much prevalent everywhere, which is in accordance with observations on the field and clinical laboratories.

This overview on coinfection patterns may be a relevant help for changing practices in administration of vaccines or antibiotics, for example toward *Mycoplasma gallisepticum* versus *Pasteurella multocida*.

All the data do not show the same relevance: the systematic detection of *Aspergillus fumigatus* was likely due to environmental contaminations from the air of poultry houses (27), which questions the relevance of this target in this respiratory panel in the future.

The development of new molecular panel assays, adapted to veterinary medicine will be one of the contribution toward a better understanding and control of respiratory coinfections in food animals, in the perspective of an evidence-based use of antibiotics.

MATERIALS AND METHODS

PCR development and PCR primers validation. Vaccines or reference strain of viruses and bacteria were used as positive controls (*supplementary table 1*).

DNA and RNA were extracted using the Nucleospin RNA virus kit (Macherey Nagel), as previously done by Cheval and colleagues (28). RNA and DNA were extracted following manufacturer instructions and were then eluted in 50µL RNase free water before storage at -80°C, as recommended by the manufacturer, simultaneous viral RNA/DNA isolation can be performed by addition of Proteinase K solution.

cDNA synthesis. For RNA viruses, cDNA synthesis was performed by use of RevertAid Reverse Transcriptase kit (Life Technologies, Carlsbad, CA). Ten microliters of RNA were mixed with 0.2 μ g of random hexamer primers and the volume was completed to 11.5 μ l with distilled water before a 5 min incubation step at 65°C. Then, on ice, 4 μ L of 5X reaction buffer was added to the mixture at the same time as 20 units RiboLock Rnase Inhibitor, 2 μ L of 10mM dNTP Mix and 200 units RevertAid H Minus Reverse Transcriptase. The RT reaction was performed at 42°C for 60 min and terminated by heating at 70°C for 10 min. cDNA were stored at -20°C.

qRT-PCR primers design and selection. The sets of primers were based on in-house designed primers or selected primers from literature. Primers design was done following these criteria: (i) 60°C melting temperature, (ii) 15-20 nt long primers, (iii) 100-300 nt amplicon size, (iv) GC% included between 40 and 60%.

Primers selected from literature were chosen if these criteria were respected. The 60°C melting temperature was a necessary condition from the Biomark platform whose melting temperature was set at 60°C.

Primers validation. A first analyze was done *in silico* using blastn in order to check primer sets specificity: a 100% identity towards the target and the expect value as close as possible to zero were expected.

Primers were then validated on a lightCycler 480 system (Roche, Basel, Switzerland), by using iTaq SYBR green kit (Biorad, Hercules, CA) and following the manufacturer instructions. The amplification program consisted of a first 10 minutes step at 95°C followed by 45 cycles including one 95°C step during 15 seconds and one 60°C step during 40 seconds. The quantification cycles (Cq) for the five replicates were expected to be in a short range of 0.2 Cq. Fluorescence curves and melting peaks were first analyzed and the PCR products were loaded

on an agarose gel if necessary. Cross-reaction were intensively checked; each primers pair being tested on all positive controls.

Standard plasmids preparation. Standard plasmids allowing accurate quantification were produced using the commercial kit Strataclone PCR Cloning kit (Agilent, Santa Clara, CA). Each validated PCR test led to the production of a plasmid, to obtain standard controls directly from the PCR products. Colonies harboring plasmids were picked and incubated at 37°C in 5mL LB medium-Ampicillin (100µg/mL) during 12-16h before plasmid extraction by use of Nucleospin plasmid Quickpure kit (Macherey-Nagel, Düren, Germany). Quantification of plasmids was done with the Qubit fluorometer and the dsDNA broad range assay kit (Thermo Fisher Scientific, Waltham, MA). Then, serial dilutions of plasmids were carried out in order to obtain a range of concentrations from 10^7 to 10^1 plasmid copies for 2µL, which is the sample volume used on the LightCycler 480 system. Plasmid dilutions were done with distilled water supplemented with 10ng/µL salmon sperm DNA (Thermo Fisher Scientific).

Standard curves, linearity, primers specificity. For some sets of plasmid dilutions were tested on the LightCycler 480 from 10^1 to 10^6 plasmid copies for 2µL, a linear curve with a 1.75-2.25 efficacy and reproducibility for the five replicates (Cq range <0.2) were expected for validation of the plasmids and the primers. This test has been performed on the newly designed primers targeting the following bacteria *Bordetella avium*, *Chlamydia psittaci* and *Ornithobacterium rhinotracheale*. The primers selected in the literature targeting avian metapneumovirus (aMPV), avian influenza virus (AIV, M gene) and *Avibacterium paragallinarum* (Ap) were also submitted to that efficacy test.

The Biomark platform: reaction details. The nanofluidic chip used in this study, BioMark™ 96.96 Dynamic Array chips, the IFC Controller HX and the BioMark™ Real-Time PCR System are manufactured by Fluidigm Corporation. The IFC Controller HX utilizes pressure to control the valves in the chips and load samples, reagents and primers mix into the reaction chambers. The BioMark system can be used to thermal cycle these nanofluidic chips and image the data in real time. The use of BioMark™ 96.96 Dynamic Array chips allows the real time quantification of 9216 points at the same time.

A pre-amplification was first performed on cDNA using a mix of 66% TaqMan® PreAmp Master Mix (Applied Biosystem) and 34 % of a 0.2 μM mix of all primers used in the study. 3.7 μL of this mix was added to 1.3 μL of each cDNA. PCR was performed with an initial 10 min step at 95°C, followed by 14 cycles of a 2-steps amplification profile consisting of 15 s at 95°C for denaturation and 4 min at 60°C for annealing and extension. These products were then diluted by 5 in TE 1X.

A sample was prepared using 66% of 2 X TaqMan® Gene Expression Master Mix (Applied Biosystem), 7% of 20X DNA Binding Dye Sample Loading Reagent (Fluidigm), 7% of EvaGreen (Biotium) and 20% of TE 1X. 5 μL of this mix was added to 1.7 μL of the dilution of pre-amplified cDNA, and 5 μL of this sample mix was loaded into each sample loading well of the chip. 1.7 μL of the primers (20 μM each, forward and reverse for each gene) were mixed with 5 μL of 2X Assay Loading Reagent (Fluidigm), and 5 μL were loaded into each primer loading well of the chip. The BioMark™ 96.96 Dynamic Array chip was then loaded into the IFC Controller HX to mix primers and samples. PCR was performed with an initial 2-steps erase and hot start phase consisting of 2 min at 50°C and 10 min at 95°C, followed by 35 cycles of a 2-steps amplification profile consisting of 15 s at 95°C for denaturation and 1 min at 60°C for annealing and extension. Images were captured at each cycle with the camera of the BioMark™ Real-Time PCR System. A melting phase was then performed from 60°C to 95°C

with an image taken every 1°C. Expression data were analyzed with the Fluidigm Real-Time PCR Analysis software and Melting curves were analyzed Fluidigm with the BioMark Melting-curve Analysis software.

Biomark runs were dedicated to primers validation; cDNA and control plasmids were used as samples and screened with the 19 primer pairs.

Field samples collection and screening. A total of 178 meat turkeys from 17 distinct farms in France and Morocco were sampled following different criteria: flocks (i) showing respiratory signs (RS) for less than 5 days, (ii) with an intra-flock RS prevalence of at least 10% and (iii) antibiotics-free since the beginning of the RS and for at least one week. Tracheal swabs were sampled on 5 to 20 affected turkeys. Swabs were then kept at 4°C and shipped within 24 hours before storage at -80°C until further processing. Before collection of clinical samples, the attending veterinarian exposed the protocol to and received the informed consent of the owner of the birds.

DNA and RNA extraction, cDNA synthesis and real-time quantitative PCR were performed on the field samples as described in the above section.

178 field samples were screened on the Biomark platform, using two 9216-microwell cell arrays. For this validation, all these samples were tested four times with each of the 19 PCR primer pairs.

In addition to samples, standard plasmids used as positive controls and no template controls (NTC) were loaded on the Biomark plates.

Cq value and Tm determination. The amplification curves and melting peaks were displayed and analyzed using Fluidigm Real-Time PCR Analysis software, version 4.1.2. Concerning

analysis settings, baseline correction was linear and the Cq threshold was automatically defined. Cq values and melting peak temperatures were therefore directly obtained from the software.

Results were exported as .csv files and formatted in a spreadsheet program in order to generate a synthetic and user-friendly table where, for each sample, the reactions results were displayed as positive or negative taking into account the Cq and the melting peak values. If the Cq and the melting peak values of the four replicates were consistent, the mean of the four values was considered.

ACKNOWLEDGMENTS

This project was funded by the French Institut Carnot de Santé Animale (ICSA) and was partly performed in the framework of a Hubert Curien Partnership (PHC) (Grant PRAD/ 11-07 /24193 PC, Campus France, entitled “Emerging diseases of turkeys in Morocco”). We thank the staff of the GenoToul platform of Genomics for excellent technical assistance with the experiments on Biomark. Guillaume Croville was PhD candidate from the EPHE-SIEB doctoral school, France.

REFERENCES

1. Hughes L, Hermans P, Morgan K. 2008. Risk factors for the use of prescription antibiotics on UK broiler farms. *J Antimicrob Chemother* 61(4):947-52.
2. Agunos A, Carson C, Léger D. 2013. Antimicrobial therapy of selected diseases in turkeys, laying hens, and minor poultry species in Canada. *Can Vet J* 54(11):1041-52.
3. Anderson S, Wakeley P, Wibberley G, Webster K, Sawyer J. 2011. Development and evaluation of a Luminex multiplex serology assay to detect antibodies to bovine herpes virus 1, parainfluenza 3 virus, bovine viral diarrhoea virus, and bovine respiratory syncytial virus, with comparison to existing ELISA detection methods. *J Immunol Methods* 7;366(1-2):79-88.
4. Lung O, Ohene-Adjei S, Buchanan C, Joseph T, King R, Erickson A, Detmer S, Ambagala A. 2017 Multiplex PCR and Microarray for Detection of Swine Respiratory Pathogens. *Transbound Emerg Dis* 64(3):834-848.
5. Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. *Infect Immun* 70(12):6871-9.
6. Cecchinato M, Lupini C, Munoz Pogoreltseva OS, Listorti V, Mondin A, Drigo M, Catelli E. 2013. Development of a real-time RT-PCR assay for the simultaneous identification, quantitation and differentiation of avian metapneumovirus subtypes A and B. *Avian Pathol* 42(3):283-9.
7. Chen X, Mifflin JK, Zhang P, Blackall PJ. 1996. Development and application of DNA probes and PCR tests for *Haemophilus paragallinarum*. *Avian Dis* 40(2):398-407.
8. Chen YC, Higgins MJ, Maas NA, Murthy SN. 2006. DNA extraction and *Escherichia coli* quantification of anaerobically digested biosolids using the competitive touchdown PCR method. *Water Res* 40(16):3037-44.

9. Clermont O, Bonacorsi S, Bingen E. 2000. Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. *Appl Environ Microbiol* 66(10): 4555–4558.
10. Farsang A, Ros C, Renström LH, Baule C, Soós T, Belák S. 2002. Molecular epizootiology of infectious bronchitis virus in Sweden indicating the involvement of a vaccine strain. *Avian Pathol* 31(3):229-36.
11. Fouchier RA, Bestebroer TM, Herfst S, Van Der Kemp L, Rimmelzwaan GF, Osterhaus AD. 2000. Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. *J Clin Microbiol* 38(11):4096-101.
12. Jarquin R, Schultz J, Hanning I, Ricke SC. 2009. Development of a real-time polymerase chain reaction assay for the simultaneous detection of *Mycoplasma gallisepticum* and *Mycoplasma synoviae* under industry conditions. *Avian Dis* 53(1):73-7.
13. Johnson GL, Bibby DF, Wong S, Agrawal SG, Bustin SA. 2012. A MIQE-compliant real-time PCR assay for Aspergillus detection. *PLoS One* 7(7):e40022.
14. Kho CL, Mohd-Azmi ML, Arshad SS, Yusoff K. 2000. Performance of an RT-nested PCR ELISA for detection of Newcastle disease virus. *J Virol Methods* 86(1):71-83.
15. Mahmoudian A, Kirkpatrick NC, Coppo M, Lee SW, Devlin JM, Markham PF, Browning GF, Noormohammadi AH. 2011. Development of a SYBR Green quantitative polymerase chain reaction assay for rapid detection and quantification of infectious laryngotracheitis virus. *Avian Pathol* 40(3):237-42. doi: 10.1080/03079457.2011.553582.
16. Register KB, Yersin AG. 2005. Analytical verification of a PCR assay for identification of *Bordetella avium*. *J Clin Microbiol* 43(11):5567-73.

17. Wang XP, Zhu DK, Wang MS, Cheng AC, Jia RY, Chen S, Chen XY, Tang T. 2012. Development and application of specific polymerase chain reaction assay targeting the *gyrB* gene for rapid detection of *Riemerella anatipestifer*. *Poult Sci* 91(10):2450-3.
18. Glisson JR. 1998. Bacterial respiratory disease of poultry. *Poult Sci* 77(8):1139-42.
19. Kleven SH. 1998. Mycoplasmas in the etiology of multifactorial respiratory disease. *Poult Sci* 77(8):1146-9.
20. Jones RC. 2010. Viral respiratory diseases (ILT, aMPV infections, IB): are they ever under control? *Br Poult Sci* 51(1):1-11.
21. Spurgeon SL, Jones RC, Ramakrishnan R. 2008. High Throughput Gene Expression Measurement with Real Time PCR in a Microfluidic Dynamic Array. *PLoS One* 3(2): e1662.
22. Jartti T, Söderlund-Venermo M, Hedman K, Ruuskanen O, Mäkelä MJ. 2013. New molecular virus detection methods and their clinical value in lower respiratory tract infections in children. *Paediatr Respir Rev* 14(1):38-45.
23. McCulloh RJ, Andrea S, Reinert S, Chapin K. 2014. Potential Utility of Multiplex Amplification Respiratory Viral Panel Testing in the Management of Acute Respiratory Infection in Children: A Retrospective Analysis. *J Pediatric Infect Dis Soc* 3(2):146-53.
24. Awad F, Baylis M, Jones RC, Ganapathy K. 2014. Evaluation of Flinders Technology Associates cards for storage and molecular detection of avian metapneumoviruses. *Avian Pathol* 43(2):125-9.
25. Sid H, Benachour K, Rautenschlein S. 2015. Co-infection with Multiple Respiratory Pathogens Contributes to Increased Mortality Rates in Algerian Poultry Flocks. *Avian Dis* 59(3):440-6.

26. Guabiraba R, Schouler C. 2015. Avian colibacillosis: still many black holes. *FEMS Microbiol Lett* 362(15):fzv118.
27. Sowiak M, Bródka K, Kozajda A, Buczyńska A, Szadkowska-Stańczyk I. 2012. Fungal aerosol in the process of poultry breeding--quantitative and qualitative analysis. *Med Pr* 63(1):1-10.
28. Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, Pariente K, Rousseaux C, Dorange F, Berthet N, Brisse S, Moszer I, Bourhy H, Manuguerra CJ, Lecuit M, Burguiere A, Caro V, Eloit M. 2011. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 49(9):3268-75.

FIGURE LEGENDS

Figure 1. BioMark specificity test

Each square represents a single real-time PCR reaction. Rows indicate the pathogen in the positive control and columns represent the specific primers set. CP values for each reaction are indicated by color; the corresponding color scale is in the legend on the right.

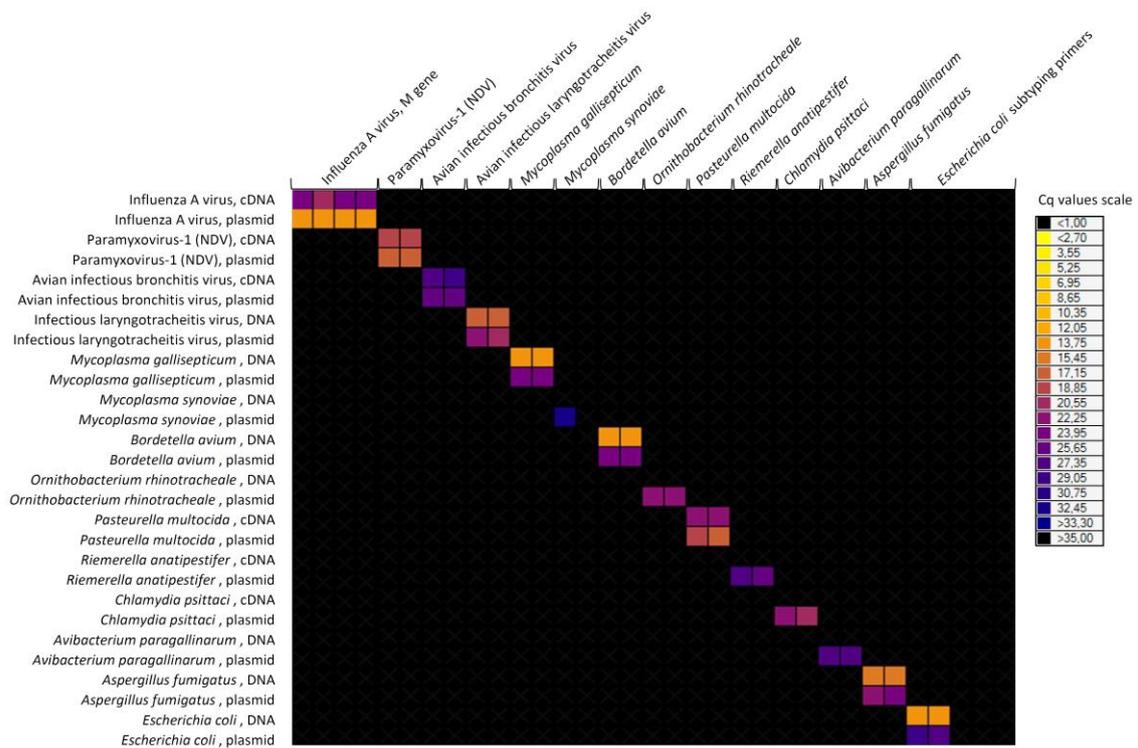


Figure 2. Biomark amplification and Tm curves

The curves on the left correspond to Eva green fluorescence generated by the amplification of DNA during PCR reaction. On the right, the melting peaks derived from the dissociation curves which are generated at the end of the reaction for the checking of the specificity of the reaction.

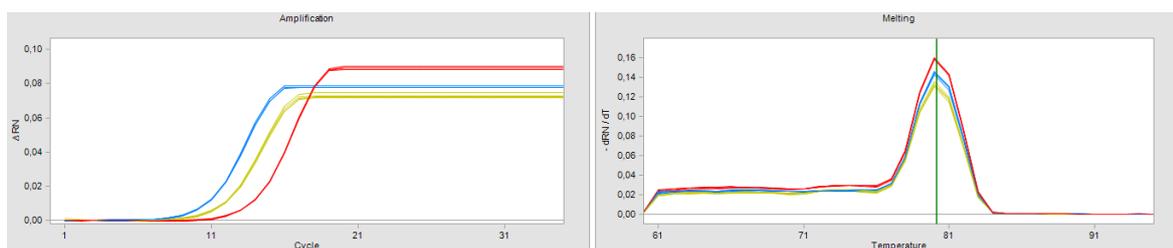


Table 1. List of primers selected from literature and designed.

Pathogen	Targeted gene	Forward primer name	Forward primer (5'-3')	Amplicon length (bp)	Reference
		Reverse primer name	Reverse primer (5'-3')		
Influenza A virus	M	M52C M253R	CTTCTAACCGAGGTCGAAACG AGGGCATTTTGGACAAAKCGTCTA	250	Fouchier <i>et al.</i> 2000
Metapneumovirus subtypes A & B	SH	SH-f SH-r	TAGTTTTGATCTTCCTTGTTGC GTAGTTGTGCTCAGCTCTGATA	200	Cecchinato <i>et al.</i> 2013
Paramyxovirus-1 (NDV)	Fusion (cleavage site)	FIP1 FIP2	TACTTTGCTCACCCCCCTT CATCTTCCCAACTGCCACT	280	Kho <i>et al.</i> 2000
Avian infectious bronchitis virus	N	N791 N1129	GTGATGACAAGATGAATGAGGA CAGCTGAGGTCAATGCTTTATC	380	Farsang <i>et al.</i> 2002
Avian infectious laryngotracheitis virus	UL15a	UL15aF UL15aR	TTGCTGTGCTATTTTCGCGTG GTAAATCGTTTAGTGCGGCAT	113	Mahmoudian <i>et al.</i> 2011

Pathogen	Targeted gene	Forward primer name Reverse primer name	Forward primer (5'-3') Reverse primer (5'-3')	Amplicon length (bp)	Reference
<i>Mycoplasma gallisepticum</i>	16S	MG14F MG13R	GAGCTAATCTGTAAAGTTGGTC GCTTCCTTGCGGTTAGCAAC	186	Jarquin <i>et al.</i> 2009
<i>Mycoplasma synoviae</i>	16S	MSLF MSLR	GAGAAGCAAAATAGTGATATCA CAGTCGTCTCCGAAGTTAACAA	214	Jarquin <i>et al.</i> 2009
<i>Bordetella avium</i>	<i>recA</i>	BaREC2f BaREC2r	CGGAATCGTCGGGTAAAACG TGGAAGCGTACTGGACATCG	200	In-house primers
<i>Ornithobacterium rhinotracheale</i>	<i>gyrA</i>	ORT101F ORT101R	TGGGCAAGGGAACCTTTGGTT TGTCGGCAAGCATTTCCTCA	101	In-house primers
<i>Pasteurella multocida</i>	<i>gyrB</i>	gyrBPMF gyrBPMR	GCCCTTTCCGATAAATTGCAA ATCGCGGCTAATGGTGCTT	100	Boyce <i>et al.</i> 2002
<i>Riemerella anatipestifer</i>	<i>gyrB</i>	gyrBP1 gyrBP2	AGAGCGAGAAGAAAAACCT CTCCATAAGCATAGAGAAGA	194	Wang <i>et al.</i> 2012
<i>Chlamydia psittaci</i>	<i>ompA</i>	Cp2fbis Cp2r	CTCGCCCTGTCTTACAGATTG GCATCAAAAGTTGCTCGTGACC	74	In-house primers
<i>Avibacterium paragallinarum</i>	<i>hp1</i>	HP1F HP1R	TGAGGGTAGTCTTGACGCGAAT CAAGGTATCGATCGTCTCTACT	500	Chen <i>et al.</i> 1996
Archaea	16S rDNA	A751F A976R	CCGACGGTGAGRGRYGAA YCCGGCGTTGAMTCCAATT	225	Just <i>et al.</i> 2013
<i>Escherichia coli</i> (4 subtypes)	<i>gadA</i>	GadA.F GadA.R	GCGTTGCGTAAATATGGTTTGCCGA CGTCACAGGCTTCAATCATGCGTT	305	Chen <i>et al.</i> 2006
<i>Escherichia coli</i> (4 subtypes)	<i>chuA</i>	ChuA.1 ChuA.2	GACGAACCAACGGTCAGGAT TGCCGCCAGTACCAAAGACA	279	Clermont <i>et al.</i> 2000
<i>Escherichia coli</i> (4 subtypes)	<i>yjaA</i>	YjaA.1 YjaA.2	TGAAGTGTCAGGAGACGCTG ATGGAGATGCGTTCCTCAAC	211	Clermont <i>et al.</i> 2000
<i>Escherichia coli</i> (4 sous-types)	TspE4C2	TspE4C2.1 TspE4C2.2	GAGTAATGTGCGGGGCATTCA CGCGCCAACAAAGTATTACG	152	Clermont <i>et al.</i> 2000
<i>Aspergillus fumigatus</i>	18S rDNA	Asp-fw Asp-rv	CTTGATTGCTGAAGACTAAC CTAACTTTCGTTCCCTGATTAATG	76	Johnson <i>et al.</i> 2012

Table 2. Limit of quantification (LOQ), Limit of detection (LOD), efficacy and r^2 values for 5 PCR primers pairs.

<i>Pathogen</i>	<i>LOQ</i> (copies/2 μ L)	<i>LOD</i> (copies/2 μ L)	<i>Efficacy</i>	<i>r</i> ²
<i>Avian metapneumovirus (A, B)</i>	10 ⁶ -10 ²	50	1,92	0,997
<i>Influenza virus (M)</i>	10 ⁶ -10 ²	50	1,96	0,999
<i>Avibacterium paragallinarum</i>	10 ⁵ -10 ²	50	1,86	0,996
<i>Chlamydia psittaci</i>	10 ⁶ -10 ²	50	1,96	0,999
<i>Bordetella avium</i>	10 ⁵ -10 ¹	5	2,02	0,995

Table 3. Synthetic report of screening performed on swabs sampled on French and Moroccan turkeys.

	<i>Flock</i>	<i>M</i>	<i>aMPV</i>	<i>PMV-1</i>	<i>IBV</i>	<i>ILT</i>	<i>MG</i>	<i>MS</i>	<i>ORT</i>	<i>BA</i>	<i>PM</i>	<i>CP</i>	<i>RA</i>	<i>AP</i>	<i>Asp fum</i>	<i>E. coli B2 genotype</i>
<i>French samples</i>	FR1	-	3/12	8/12	1/12	-	-	-	12/12	3/12	-	-	-	-	12/12	9/12
	FR2	-	7/8	-	-	1/8	2/8	3/8	8/8	-	-	-	-	-	8/8	7/8
	FR3	-	-	-	-	-	-	9/9	8/9	-	-	-	-	-	9/9	1/9
	FR4	-	1/12	-	-	-	-	5/12	9/12	4/12	-	-	-	-	12/12	6/12
	FR5	-	8/11	5/11	-	-	5/11	11/11	11/11	-	-	-	-	-	11/11	9/11
	FR6	-	-	2/15	-	-	3/15	1/15	4/15	-	-	-	2/15	-	15/15	7/15
	FR7	-	10/10	-	-	-	4/10	5/10	6/10	-	-	-	-	-	10/10	6/10
	FR8	-	1/11	-	-	11/11	-	-	-	-	-	-	-	-	11/11	11/11
	FR9	-	-	-	-	-	-	-	-	-	-	-	-	-	2/2	-
<i>Moroccan samples</i>	MO1	-	11/15	2/15	-	1/15	1/15	-	14/15	-	10/15	-	1/15	-	15/15	9/15
	MO2	-	1/15	5/15	-	-	13/15	5/15	12/15	1/15	14/15	-	-	-	13/15	9/15
	MO3	-	6/15	-	-	2/15	13/15	-	14/15	-	11/15	1/15	-	-	15/15	12/15
	MO4	-	1/9	-	-	-	-	-	3/9	5/9	1/9	-	-	-	8/9	1/9
	MO5	-	3/10	2/10	-	-	3/10	-	10/10	7/10	4/10	-	-	-	9/10	2/10
	MO6	-	4/18	5/18	-	-	-	-	17/18	-	1/18	-	-	-	18/18	11/18
	MO7	-	1/5	-	-	-	-	-	-	-	-	-	-	-	5/5	1/5
	MO8	-	-	-	-	-	-	-	-	-	-	-	-	-	1/1	1/1

M: Influenza A virus, M gene ; *aMPV*: avian metapneumovirus ; *PMV-1*: Paramyxovirus-1 (NDV) ; *IBV* : Avian infectious bronchitis virus ; *ILT*: Avian infectious laryngotracheitis virus; *MG*: Mycoplasma gallisepticum; *MS*: Mycoplasma synoviae; *ORT*: Ornithobacterium rhinotracheale; *BA*: Bordetella avium; *PM*: Pasteurella multocida; *CP*: Chlamydia psittaci; *RA*: Riemerella anatipestifer; *AP*: Avibacterium paragallinarum; *Asp fum*: Aspergillus fumigatus

Deuxième partie expérimentale

Identification de l'agent responsable de la maladie foudroyante de la pintade par métagénomique

La recherche d'agent étiologique par une procédure sans *a priori* peut s'avérer indispensable lorsque toutes les procédures de virologie « classiques » ont été mises en œuvre sans produire de résultats convaincants. Le cas de la maladie foudroyante de la pintade (MFP) en est une illustration concrète. Cette entérite aiguë, à l'origine d'un très fort taux de mortalité, a été décrite pour la première fois dans les années 1980 dans le sud de la France. Les recherches menées dans les années 1980 ont proposé diverses origines virales comme un *toga-like virus* (Brahem et al., 1992), un réovirus (Andral et al., 1987), un herpesvirus ou encore un astrovirus (Toffan et al., 2012) mais sans réellement justifier leurs hypothèses car elles étaient basées sur de la microscopie électronique (ME). De plus, l'observation d'artéfacts par ME peut sérieusement entraver les résultats comme Ackermann et Tiekotter le décrivent en 2012 dans le cadre de l'observation de bactériophages par ME (Ackermann and Tiekotter, 2012).

L'agent responsable de la MFP étant très difficilement cultivable et par conséquent complexe à caractériser sur le plan génétique, le séquençage à très haut-débit sans *a priori* était une option raisonnable, rendue possible par les développements technologiques récents. Au vu du débit de données et de la taille des séquences produites, la technologie Illumina semblait la plus appropriée, avec précisément l'appareil Miseq, mieux adapté aux projets d'infectiologie qui ne nécessitent pas les volumes de lectures générées par le Hiseq.

La technologie Illumina doit être privilégiée pour deux grand types d'applications :

- La recherche de mutations rares. En effet, la fidélité élevée du séquençage Illumina (taux d'erreur de 10^{-3} (Fox et al., 2014)) est nécessaire pour ce type de projets car des SNP présents en de très faibles proportions doivent pouvoir être interprétés avec confiance.
- La métagénomique ou la recherche d'agents infectieux sans *a priori*. La masse de données produites (30 millions de séquences sur Miseq) et la petite taille des séquences (environ 300 paires de bases) sont deux critères nécessaires à ce genre de travaux. Une théorie intéressante énoncée dans Quick et al, propose que la proportion de longues séquences diminue dans le cas de faibles quantités de virus et que, par conséquent, il est préférable de cibler de courts brins d'ADN (Quick et al.,

2017). Les auteurs citent notamment les travaux de Worobey et al qui ont appliqué cette approche, baptisée *jackhammering*, pour l'amplification de génomes de HIV-1 à partir d'échantillons dégradés datant des années 1970 (Worobey et al., 2016).

D'un point de vue technique, la chimie Illumina appartient à la famille du séquençage par élongation. Les ADN à séquencer sont couplés à des séquences adaptatrices à leurs extrémités afin de les hybrider sur une lame. Les brins d'ADN sont alors répartis sur toute la surface de la lame avant leur amplification clonale. Le signal fluorescent émis par la synthèse d'un seul brin d'ADN n'étant pas détectable (comme pour le 454 de Roche), une amplification clonale du brin à séquencer est nécessaire. Cette amplification clonale est originale dans le sens où elle se fait en « pont ». Une fois l'ADN hybridé à la lame, il y a synthèse du brin d'ADN complémentaire et élimination du brin d'origine de sorte à ce que l'ADN à séquencer soit directement lié à la lame (Figure 8). Le brin néo-synthétisé est « plié » pour former un pont et le complémentaire de l'adaptateur de sa seconde extrémité va s'hybrider à la lame. La synthèse du brin complémentaire de cet ADN courbé et la dénaturation du double brin permettront d'avoir un brin sens et un brin anti-sens côte à côte. Cette amplification en pont a lieu ainsi de suite de sorte à obtenir des milliers de molécules d'ADN. Avant le séquençage, le brin anti-sens est éliminé et la synthèse du brin complémentaire a lieu avec des nucléotides fluorescents. Au fur et à mesure de la synthèse, la fluorescence est acquise et permettra de déterminer la composition en bases azotées du brin d'ADN (Figure 9). Une fois le brin sens lu, le brin anti-sens est synthétisé puis le brin sens est dégradé. Enfin, la lecture du brin anti-sens a lieu selon le même principe que pour le brin sens.

L'intérêt de cette lecture dans les deux sens (*paired-end*) varie selon le contexte. Une double lecture sur un petit fragment de 150 pb permet d'accroître la confiance en la séquence obtenue. Dans un autre type d'application, l'intérêt sera de pouvoir lire un brin de 300 pb avec 150 pb dans un sens puis 150 pb dans l'autre.

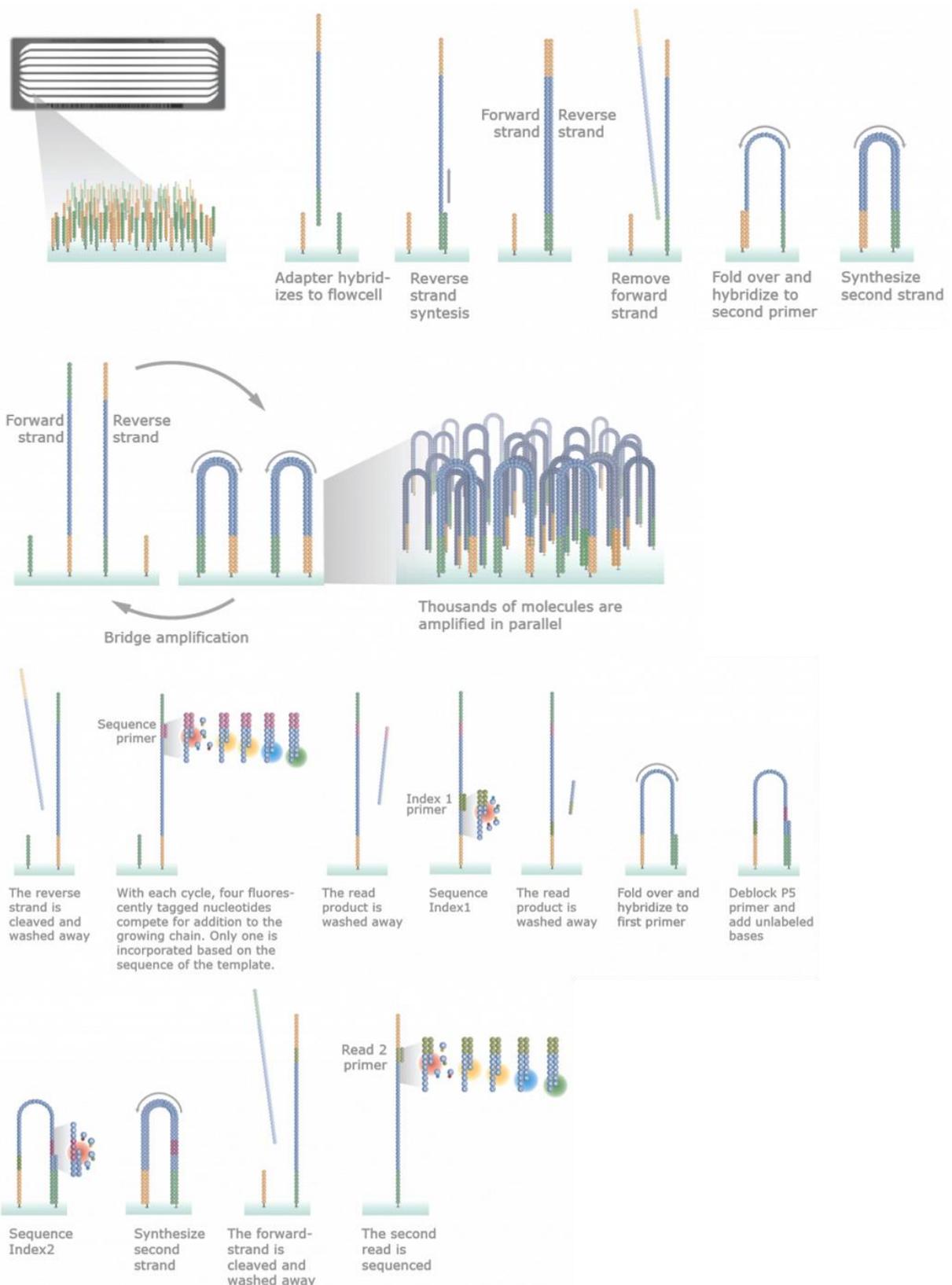


Figure 8 Principe de fonctionnement de la technologie Illumina (<http://www.biofidal-lab.com/>)

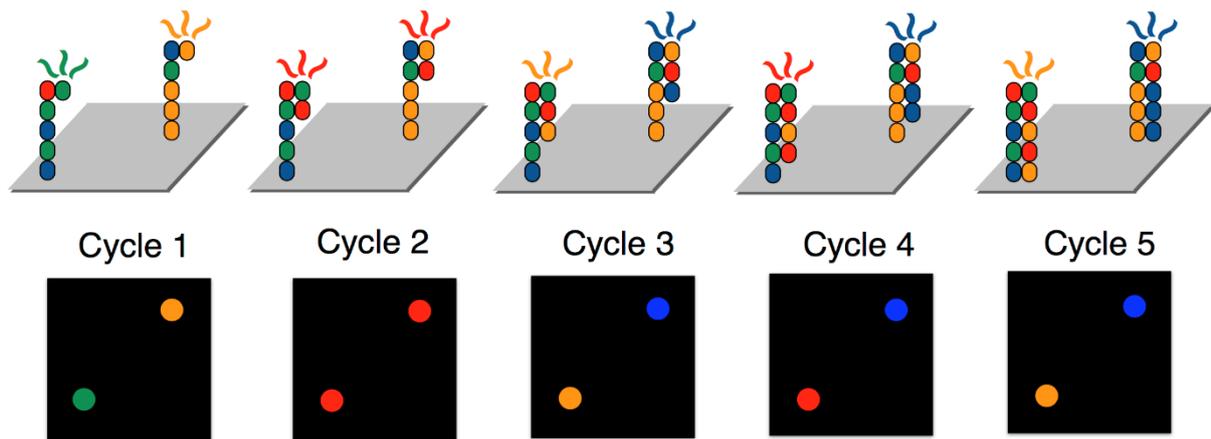


Figure 9 Représentation graphique du signal fluorescent généré par l'intégration d'un nucléotide couplé à un fluorochrome. La partie supérieure de l'illustration représente l'incorporation, au cours des cycles de PCR, des nucléotides complémentaires au brin à séquencer fixé sur la lame. La partie inférieure représente la vue de dessus avec le suivi des couleurs émises par les fluorochromes au cours des cycles de PCR. (<http://data-science-sequencing.github.io>)

Cette étude décrit une reproduction expérimentale de la maladie suite à l'inoculation, à des pintades saines, de contenu digestif (clarifié et filtré) de pintades prélevé dans le cadre d'un cas clinique de terrain. La maladie ayant été reproduite avec succès, du contenu digestif des animaux d'expérimentation a été prélevé puis soumis à une ultracentrifugation. Les culots ont été traités par des nucléases de sorte à éliminer les acides nucléiques libres non encapsidés puis l'extraction des ARN par le Trizol a été réalisée.

Le protocole d'amplification des ARN a été inspiré de travaux publiés (Victoria et al., 2009). La quantité annoncée de séquences générées par le Miseq étant de 30 millions, il était alors envisageable de multiplier (ou « multiplexer ») la quantité d'échantillons à séquencer en une seule réaction. En supposant qu'il était raisonnable d'attribuer 1 million de séquences par échantillon pour obtenir une vision détaillée des agents infectieux présents, nous avons réparti 30 échantillons sur une lame de Miseq. L'étude récente de Quick et al sur le séquençage du virus Zika démontre la nécessité d'attribuer plusieurs millions de séquences puisque, par exemple, sur près de 2 millions de séquences obtenues pour un échantillon, seulement 28 s'alignaient sur le génome de référence du virus Zika (Quick et al., 2017).

Pour le *multiplexage*, il est nécessaire de réaliser une librairie de séquençage par échantillon. La préparation des librairies comprend une étape de ligation d'un « code-barres » nucléotidique afin de pouvoir, après le séquençage, réattribuer les séquences à leur échantillon propre. Cette étape d'étiquetage étant longue et coûteuse, nous avons décidé de réaliser trois librairies contenant dix échantillons chacune, les dix échantillons mélangés devant donc être étiquetés de façon artisanale et selon la méthode de Victoria et al.

Brièvement, la méthode consiste en une rétro-transcription des ARN à l'aide d'hexamères aléatoires contenant une étiquette en 5'. Le brin complémentaire de l'ADNc alors étiqueté est synthétisé puis une PCR est réalisée avec des amorces correspondant à l'étiquette et sa séquence complémentaire. Étant donné que la rétro-transcription a été effectuée à l'aide d'amorces aléatoires, les tailles de produit de PCR sont très variables ; il s'agit donc ensuite de sélectionner sur gel d'agarose une « fenêtre » autour de 300 pb et de purifier les ADN avant de les soumettre au séquençage.

Les résultats du séquençage à haut débit ont permis de mettre en évidence le génome d'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 ainsi une comparaison phylogénique. Afin de confirmer l'association du virus à cette entérite, des lots de pintades ont été testés par PCR conventionnelle pour confirmer les liens de causalité. De plus, des cas spontanés (déclarés avant et après le début de cette étude) ont été criblés et ont été testés positifs pour la présence de gfCoV. 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 la présence du virus sur l'épithélium intestinal, et au passage, la proximité antigénique de ce coronavirus avec le TCoV. Le caractère fulminant de cette entérite reste toutefois atypique par rapport aux autres entérites aviaires à coronavirus puisque le TCoV (entérique également) cause le plus souvent très peu de signes cliniques chez la dinde.

Une fois extraites du fichier de séquençage, les séquences de coronavirus aviaire ont été alignées face à un génome de référence de coronavirus de la dinde TCoV, 70% du génome de référence étaient couverts par les données NGS. De sorte à réaliser une étude phylogénétique, le gène de spicule a été séquençé par la technique de Sanger, ce qui a permis de qualifier ce virus comme étant très proche (sur la base génétique) du TCoV nord-américain.

L'analyse par bioinformatique des séquences produites par le séquenceur Illumina peut s'avérer très lourde au vu du volume de données générées. La première étape consiste en un tri des séquences selon leur étiquette puis en l'appariement des deux sens de lecture pour ne former qu'une seule séquence. Ensuite, le fichier Fastq (un fichier fastq est un fichier fasta comprenant l'information de qualité pour chaque base) obtenu est confronté à une base de données virale, étape longue et coûteuse. L'échantillon correspondant au contenu intestinal de pintade a généré

près de 480 000 séquences dont environ 50 000 ont été affiliées à des séquences virales soit plus de 10%, ce qui est extrêmement satisfaisant. Parmi les séquences virales, 1441 ont été identifiées comme appartenant à la famille des *Coronaviridae*. Ces résultats ont directement été obtenus avec le programme GAAS (*Genome relative Abundance and Average Size*) dont le fonctionnement repose sur l’algorithme BLAST (*Basic Local Alignment Search Tool*). Les résultats sont rendus sous forme de « krona », un diagramme circulaire interactif, il suffit de cliquer sur l’espèce, le genre ou la famille pour voir les affiliations proposées par l’outil GAAS (Figure 10).

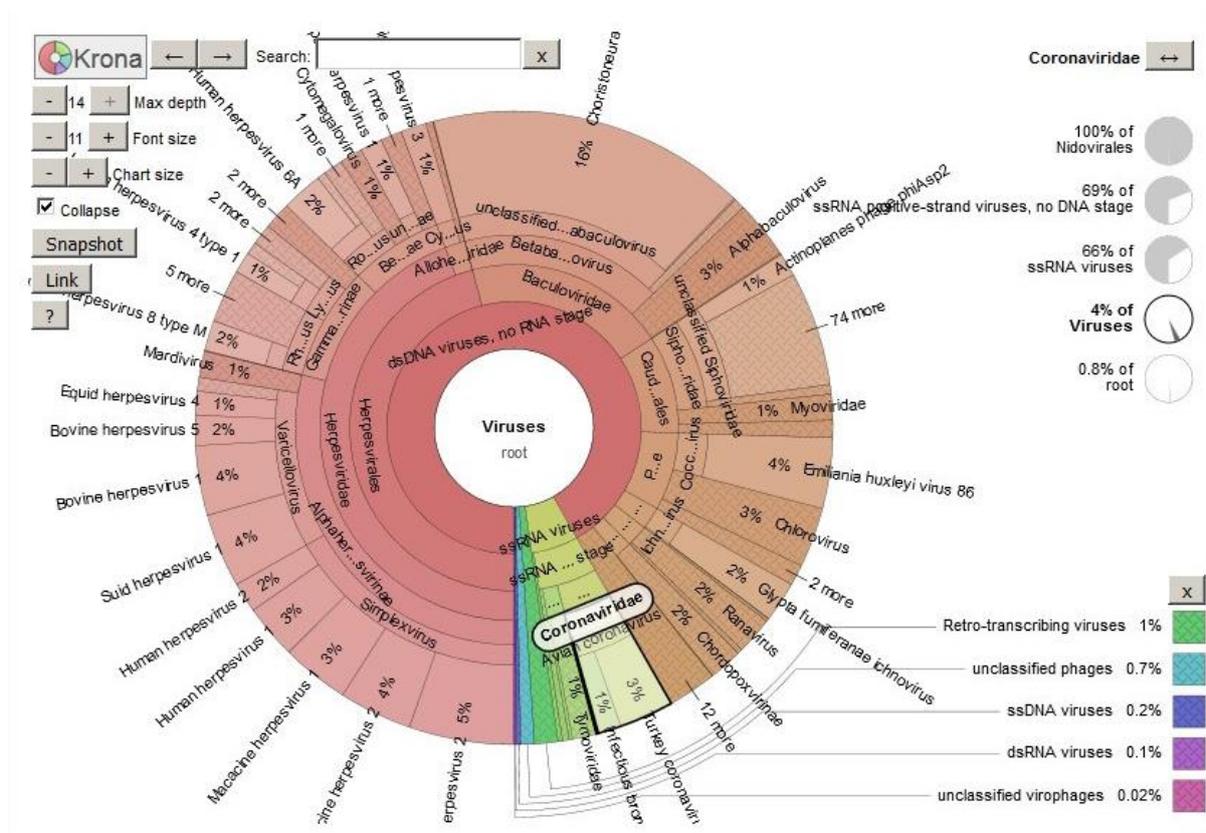


Figure 10 Krona généré par l’outil d’analyse bioinformatique GAAS. Tous les résultats de l’affiliation taxonomique sont regroupés dans ce diagramme.

Ces travaux sont une illustration de l’apport des NGS dans l’identification d’agents étiologiques jusqu’alors non caractérisés. La quantité de séquences attribuées à un échantillon ne doit pas être négligée car dans ces travaux ce ne sont que 0.3% des séquences qui finalement étaient pertinentes.

Des résultats d'affiliation taxonomique restitués sous forme de krona sont à interpréter avec précautions car une seule information est finalement proposée, la quantité de séquences appartenant à telle ou telle espèce virale. L'inconvénient de résultats sous cette forme est l'absence d'informations sur la couverture du génome présent. Ainsi, 1000 séquences affiliée à une espèce virale mais ne s'alignant que sur 0.5% du génome ne seront pas suspectées comme étant des séquences endogènes.

Les éléments viraux endogènes (EVE) sont des reliques génomiques qui résultent d'intégrations accidentelles, dans les cellules germinales, de génomes viraux. Une fois « fossilisés » dans le génome de l'hôte, ces séquences sont transmises verticalement. Les rétrovirus nécessitant l'intégration de leur propre génome dans celui de l'hôte pour se répliquer, ils représentent donc la majorité des EVE. Lors de nos travaux de métagénomique, des séquences d'herpesvirus ou de poxvirus étaient systématiquement retrouvées. Après un alignement desdites séquences, nous apercevions que, dans le cas des herpesvirus, elles s'alignaient sur seulement quelques centaines des bases du génome de référence au niveau des gènes de télomérases. Nous pouvions alors en déduire que ces séquences étaient finalement issues du génome de l'hôte.

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 Donnadiou, 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 poults (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 poults 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

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. Donnadiou); and North Carolina State University, Raleigh, North Carolina, USA (J. Guy)

¹ These authors contributed equally to this article.

ultracentrifugation (100,000 × g, 2 h). The pellets were treated with RNase (Ambion, 20 µg/mL) and DNase (Invitrogen, 10 U/µL) (both from Life Technologies, Grand Island, NY, USA) to concentrate the viral material.

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 $10^{0.3}$ (7). †Percent within eukaryote virus reads.

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

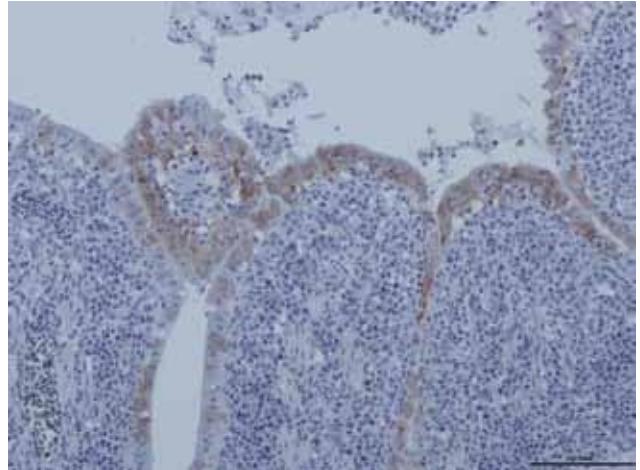


Figure 1. Replication of guinea fowl coronavirus in intestinal epithelium cells of experimentally infected birds as evidenced by immunohistochemical testing with a turkey coronavirus-specific monoclonal antibody, France, 2010–2011 (9). Scale bar indicates 0.1 mm.

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).

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%).

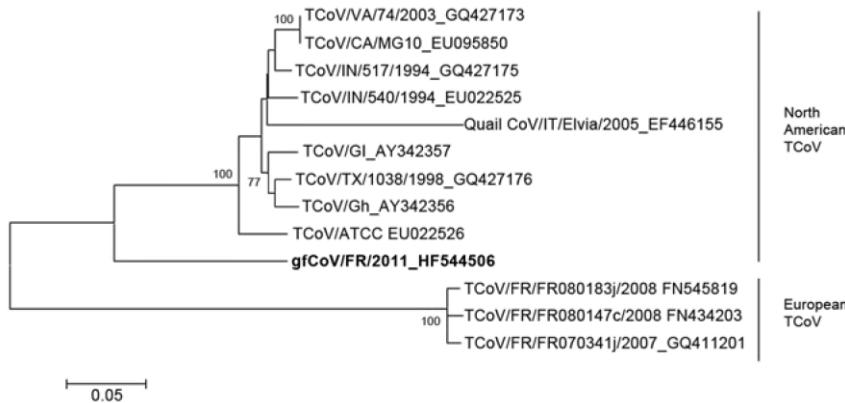


Figure 2. Phylogenetic analysis of the complete spike (S) gene of the fulminating disease of guinea fowl coronavirus (gfCoV, in **boldface**) in relation to turkey coronaviruses (TCoVs) at the nucleotide level. The tree was generated by using MEGA 5.05 (<http://megasoftware.net/mega.php>) by the neighbor-joining 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.

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, 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, 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. <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:pii:e00473-12.
5. Gouilh MA, Puechmaille SJ, Gonzalez J-P, Teeling E, Kittayapong P, Manuguerra J-C. SARS-coronavirus ancestor's foot-prints in SouthEast 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

Troisième partie expérimentale

Surveillance et typage d'avipoxvirus par séquençage de génomes complets avec la technologie MinIon d'Oxford Nanopore Technologies.

Le séquençage à très haut débit comme outil de diagnostic ou de suivi d'évolution génétique en temps réel n'est pas encore à la portée de tous les laboratoires. En effet, compte tenu de l'investissement nécessaire, tant financier que méthodologique, et de la lourdeur du séquençage puis de l'analyse bioinformatique, les NGS sont pour l'heure réservées à un usage dédié à des utilisateurs spécialisés.

Des exemples d'application de la métagénomique au diagnostic existent déjà, notamment en médecine humaine. Par exemple, Murkey *et al* décrivent le cas d'une patiente de 58 ans transplantée des deux poumons, hospitalisée après huit jours de maux de tête, nausée, vomissements, rigidité de la nuque et photophobie. Suite à l'échec des examens cliniques classiques, une analyse métagénomique du liquide céphalo-rachidien de la patiente a révélé la présence du génome complet du virus de l'hépatite E (HEV) (Murkey et al., 2017). Cet exemple de diagnostic par NGS est très élégant, néanmoins, au vu du volume de séquences générées (plus de 23 millions) et de la quantité de séquences affiliées au virus HEV (14 665 séquences) mais aussi du type de prélèvement, il est difficile d'imaginer pouvoir transposer cette méthodologie à du diagnostic rapide en médecine vétérinaire. Dans le même esprit, Mongkolrattanothai et al décrivent le diagnostic d'une neurobrucellose chez un enfant de 11 ans grâce à la métagénomique. Cette fois, la quantité de séquences générées (par séquençage Illumina) pour la fraction ADN est d'environ 23 millions dont seulement 277 ont été affectées au genre *Brucella* (Mongkolrattanothai et al., 2017). Une fois encore, la démarche employée ne correspond pas aux moyens accessibles en médecine vétérinaire.

Les deux études citées ci-dessus ont fait appel à du séquençage de technologie Illumina, qui génère un débit de séquences extrêmement élevé et une fidélité de séquençage remarquable. Cependant, la taille des séquences produites n'excède pas 600 pb (<https://www.illumina.com>).

Les méthodes NGS produisant des séquences courtes sont également parfaitement adaptées à la recherche de motifs génétiques particuliers comme par exemple la détection de mutations ponctuelles liée à la résistance aux antiviraux chez des patients infectés par le VIH-1 (Simen et al., 2014). Dans ce type d'application, le séquençage est réalisé sur une partie déterminée du

génomique comme, par exemple, le gène codant pour la transcriptase inverse puisqu'un des antiviraux utilisé dans la lutte du VIH est un inhibiteur de la transcriptase inverse.

L'étude décrite dans cette troisième partie expérimentale traite du séquençage du génome complet de poxvirus aviaire par la technologie d'Oxford Nanopore sans enrichissement ni amplification de génome au préalable.

L'apparition sur le marché du séquenceur MinION d'Oxford Nanopore Technologies et notamment de la nouvelle version de puce R9.4 a bouleversé l'offre des NGS. D'une part, la miniaturisation de l'équipement de séquençage est révolutionnaire (Figure 11) puisque le séquenceur MinION tient dans la main : aussi, son utilisation est possible n'importe où dès lors que l'on dispose d'un ordinateur portable performant.



Figure 11 Séquenceur MinION (Oxford Nanopore Technologies)

Le principe de fonctionnement de la technologie ONT consiste en un suivi du champ électrique qui traverse un pore protéique et qui est modifié par le passage d'un brin d'ADN au travers de celui-ci (Figure 12). Cette technologie est innovante car il s'agit d'une lecture « directe » du brin d'ADN, par opposition aux autres chimies qui sont basées sur la détection des nucléotides incorporés lors de la synthèse du brin d'ADN complémentaire. Le signal brut généré est donc un historique de variation du signal électrique générée par des *k-mers* de cinq nucléotides.

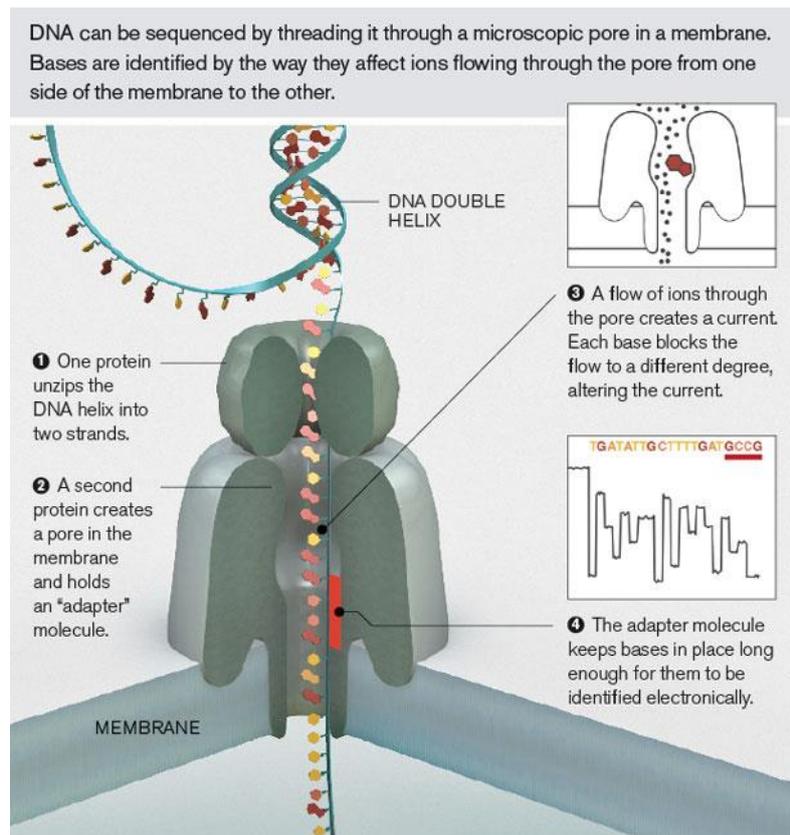


Figure 12 Illustration du passage d'un brin d'ADN au travers d'un pore protéique lors du séquençage MinION (John MacNeil, Tech Review).

Cet équipement est accessible « à tout le monde » et, sous réserve d'un accompagnement appuyé par une plateforme de séquençage, il est tout à fait envisageable de pouvoir l'utiliser en autonomie. La perspective de pratiquer les NGS au même titre que la PCR devient donc, *en théorie*, à portée de main. En ce qui concerne la chimie de séquençage, les performances peuvent atteindre un débit de 10 Gb avec des séquences dont la longueur peut atteindre plusieurs dizaines de kb. Le taux d'erreur de séquençage s'est nettement amélioré après être passé de 38.2% (Laver et al., 2015) à 7.5% avec la version R9.0 (Jain et al., 2017), ce qui fait du MinION un sérieux concurrent du PacBio RSII (séquenceur de troisième génération produisant de longs fragments). Malgré une fidélité de séquençage très élevée, le PacBio RSII nécessite un investissement en équipement très onéreux et présente le défaut de nécessiter une pièce climatisée dédiée. Là encore, l'évolution technologique est cependant très rapide, avec la mise en service très récente d'un équipement PacBio de nouvelle génération, le Sequel.

C'est presque naturellement que nous nous sommes tournés vers la technologie ONT avec comme ambition d'obtenir le génome complet de l'agent infectieux directement à partir de matériel pathologique.

Les poxvirus aviaires sont responsables de la variole aviaire, qui se présente sous trois formes : cutanée (ou sèche), diphtérique (ou humide), et septicémique. La forme cutanée est la plus couramment rencontrée. Elle se manifeste par l'apparition de lésions prolifératives sur les régions cutanées aptères (bec, cire, paupières, crête, barbillons, pattes) (Tripathy and Reed, 2013; Van Riper and Forrester, 2007). Ces lésions sont dues à une hyperplasie des cellules épithéliales et se présentent d'abord comme des papules qui s'enflamment et grossissent. Des vésicules apparaissent à leur surface et grossissent jusqu'à devenir confluentes puis percer, donnant à la lésion un aspect de masse ulcérée qui va progressivement évoluer vers une masse très vascularisée et couverte d'une croûte plus ou moins volumineuse. En cas d'évolution favorable, la lésion se dessèche progressivement et la croûte tombe, laissant place à du tissu cicatriciel sain (Tripathy and Reed, 2013). La forme diphtérique est la seconde forme la plus courante. Elle peut être observée seule ou en association avec la forme cutanée et se manifeste par l'apparition de lésions diphtériques, généralement sur les muqueuses orales, respiratoires supérieures, cloacales et oculaires. Les lésions se présentent comme des papules enflammées qui grossissent, deviennent confluentes et se couvrent d'une pseudomembrane nécrotique jaunâtre. Le retrait de cette pseudomembrane fait apparaître des ulcères hémorragiques. Les lésions sont le plus souvent observées sur la trachée, mais elles peuvent s'étendre aux sinus et les obstruer (Tripathy and Reed, 2013).

Avant de travailler à partir des lésions cliniques, la mise au point de la méthode s'est faite à partir d'un isolat de variole aviaire cultivé sur membrane chorio-allantoïdienne d'œuf de poule embryonné. Lorsque l'objectif est de générer de longues séquences nucléotidiques, un grand soin doit être apporté lors de l'extraction de l'ADN. Pour cette raison, le protocole d'extraction employé a fait appel à de « vieilles » méthodes utilisées avant la diffusion des kits commerciaux. En l'occurrence, nous nous sommes référé au « grimoire » de 2001 : *Sambrook J and Russell DW -Molecular cloning – Third edition – Volume 1. Protocol 5 Preparation of genomic DNA from mouse tails and other small samples*. Une fois extraits, les ADN ont été soumis au protocole de séquençage MinION et les séquences produites ont été alignées sur un génome complet Fowlpox virus de référence.

La première expérience test réalisée à partir de la culture virale ayant permis d'obtenir un génome complet d'avipoxvirus, nous avons considéré que le protocole était suffisamment robuste pour l'appliquer à l'analyse directe de lésions cliniques. Sept tissus (issus de trachées, de poumon ou de lésions cutanées) ont été soumis au séquençage MinION dans les mêmes conditions que pour le test.

La préparation des bibliothèques d'ADN à séquencer a été réalisée selon un protocole sans amplification par PCR des acides nucléiques. Au vu des quantités d'ADN de départ dont nous disposons (entre 550 µg et 900 µg), il n'était pas nécessaire d'amplifier les ADN de quelque façon que ce soit, en amont du séquençage. On trouve plusieurs avantages à cela : les biais d'amplification sont éliminés, la taille des fragments n'est pas limitée par les performances de la polymérase de PCR et du temps est économisé. Deux approches de préparation de bibliothèques d'ADN sans PCR sont disponibles : l'une est basée sur un processus de ligation qui permet de contrôler la taille des fragments d'ADN et l'autre est basée sur la transposase qui permet une préparation extrêmement rapide des bibliothèques en environ 10 minutes (Figure 13).

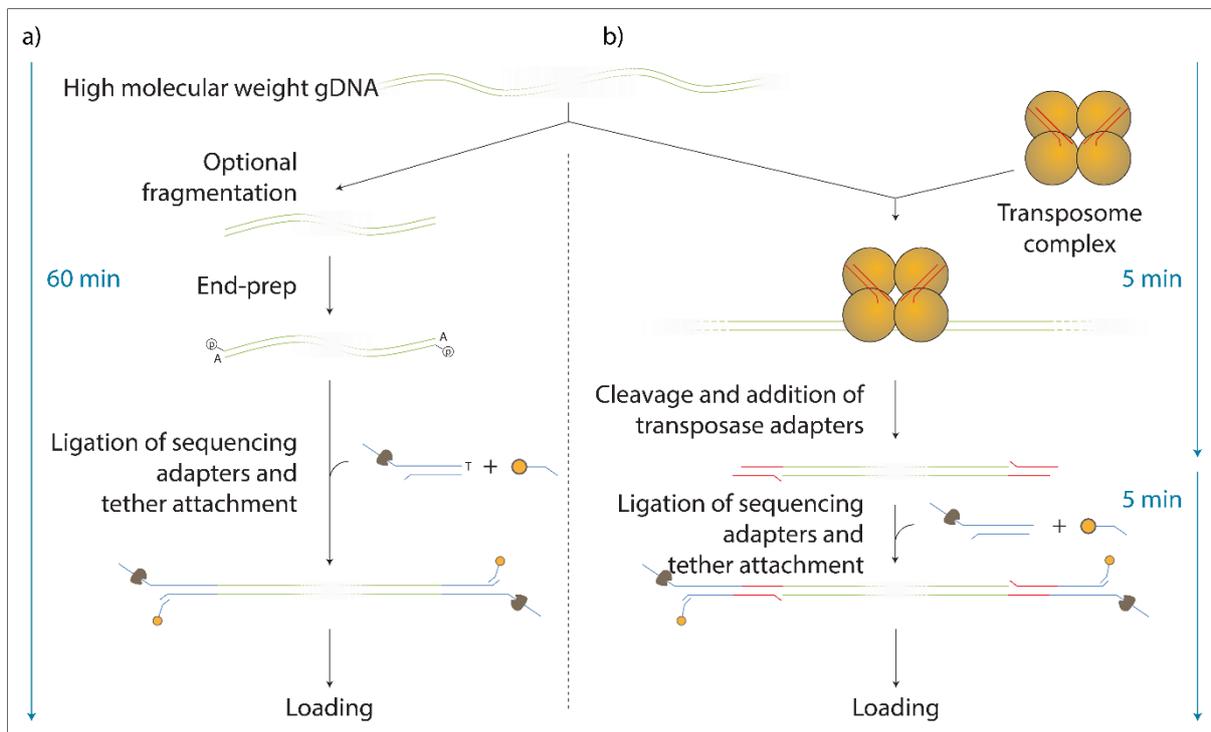


Figure 13 Représentation schématique de la préparation des bibliothèques d'ADN à séquencer sans amplification préalable des acides nucléiques. a) protocole avec ligation b) protocole "rapide"

Les données brutes issues du séquençage se présentaient sous format fast5 et ont dans un premier temps été analysées avec Albacore (v1.1.0) pour la conversion du signal brut en séquences nucléotidiques et pour le démultiplexage. Ensuite, Poretools (Loman and Quinlan, 2014) a été utilisé pour convertir les fichiers au format fastq. Chaque fichier fastq (un fichier par échantillon) a été aligné contre le génome d'avipoxvirus de référence (NC_002188.1) avec l'outil bwa (Li and Durbin, 2009) et les alignements ont été travaillés avec la suite d'outils SAMtools software package (Li et al., 2009). Ensuite, les alignements de séquences contre la référence étaient visualisés *via* le programme IGV (Robinson et al., 2011) (Figure 14).

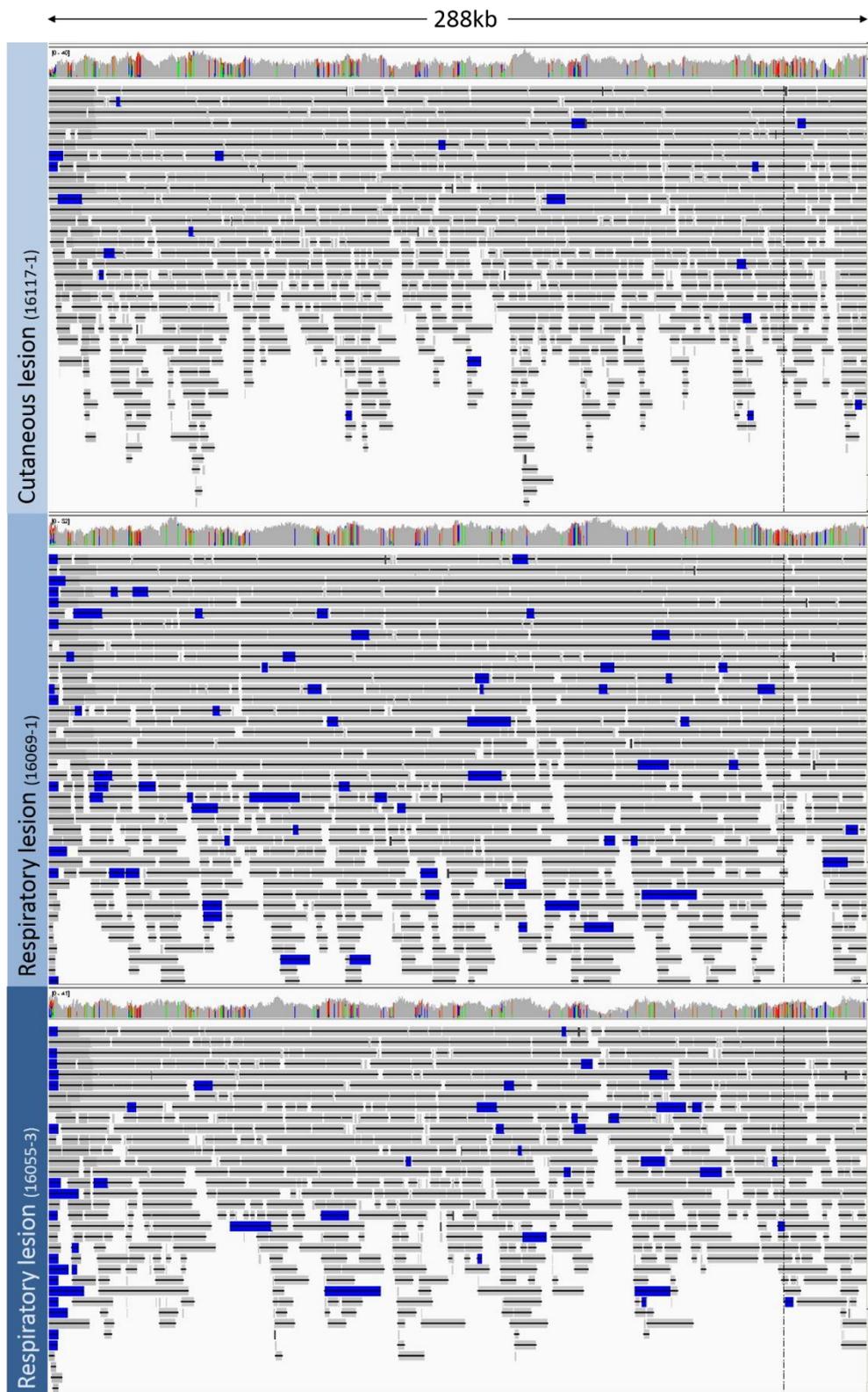


Figure 14 Affichage via le programme IGV de l'alignement de séquences des trois échantillons analysés contre le génome de référence. Les séquences sont représentées par les lignes grises horizontales.

La quantité de séquences s'alignant sur le génome d'avipoxvirus et la profondeur de lecture moyenne étaient calculées avec les options *flagstat* et *depth* des SAMtools. Enfin, la séquence consensus était créée en utilisant le programme IGV.

Ces travaux exploratoires ont permis de répondre positivement aux questions posées initialement, à savoir : a) est-il possible d'identifier l'agent infectieux responsable d'une maladie par séquençage direct des acides nucléiques du tissu infecté ? Et b) la technologie MinION permet-elle de réaliser un séquençage et un assemblage de grands génomes contenant des régions répétées comme les génomes de poxvirus ?

Les tissus analysés ont été inclus dans cette étude car ils présentaient des lésions cliniques spécifiques d'une infection à poxvirus. Leur analyse directe, sans étape d'amplification *in vivo* ou *in vitro* a mis en évidence la présence de génome d'avipoxvirus dans des quantités suffisantes pour obtenir leurs génomes complets. De plus, la grande taille des séquences obtenues (jusqu'à 61 kb) a facilité leur assemblage face au génome de référence. Les génomes de poxvirus aviaires de cette étude mesurent 288 kb et contiennent de nombreuses régions répétées, ce qui rend leur analyse complexe si la chimie de séquençage NGS utilisée produit de courtes séquences de seulement quelques centaines de pb. Ici, ce problème a été contourné grâce à une taille moyenne de séquences de 1 300 à 3 200 pb selon les échantillons.

Une condition nécessaire au succès d'une telle analyse est la quantité de matériel génétique de l'agent infectieux vis-à-vis du matériel génétique cellulaire. Un agent pathogène abondant dans un environnement riche en cellules ne sera pas pour autant facilement détectable, de plus, la quantité de séquences allouées à l'échantillon définit également la capacité à détecter des génomes faiblement présents. Dans notre cas, dans trois des sept échantillons analysés, en dépit d'une présence importante de matériel cellulaire, nous avons obtenu suffisamment de séquences virales pour assembler des génomes complets.

Il est important de préciser que nous avons connaissance de l'agent étiologique responsable de la maladie donc du fait qu'il s'agissait d'un génome à ADN monocaténaire. Cela nous a orienté sur la méthode de préparation des ADN à séquencer. De futurs projets de métagénomique clinique nécessiteront des travaux de mise au point de façon à pouvoir cibler les génomes à ADN et ARN. L'analyse bioinformatique fut quant à elle également facilitée car il s'agissait de reséquençage de génomes connus donc l'analyse consistait à aligner les séquences obtenues face à un génome de référence. Lors de métagénomique clinique, qu'elle soit virale ou bactérienne, le traitement des séquences sera plus lourd car chaque séquence produite devra

être comparée aux bases de données et l'interprétation du résultat devra tenir compte de la quantité de séquences pertinentes et de la proportion du génome candidat qu'elles couvrent.

L'analyse des données produites par la technologie MinION d'Oxford Nanopore a, malgré son arrivée récente sur le marché, déjà nécessité des améliorations en matière de conversion du signal brut en séquence ADN (*base calling*). Un protocole 1D générant 1 million de séquences soit deux à trois milliards de bases entraîne un *base calling* d'une durée de deux jours sur un serveur local.

Rapid Whole-Genome Based Typing and Surveillance of Avipoxviruses Using Nanopore Sequencing

Guillaume Croville¹, Guillaume Le Loc'h¹, Catherine Zanchetta², Maxime Manno², Christelle Camus-Bouclainville¹, Christophe Klopp³, Maxence Delverdier¹, Marie-Noëlle Lucas¹, Cécile Donnadiou², Mattias Delpont¹ and Jean-Luc Guérin¹

Affiliations :

¹Université de Toulouse, ENVT, INRA, UMR 1225, 31076 Toulouse, France

²GeT-PlaGe, INRA, 31326, Castanet-Tolosan Cedex, France.

³Plateforme Bioinformatique Genotoul, UR875, Biométrie et Intelligence Artificielle, INRA, Castanet-Tolosan, France.

Running title: real-time poxvirus whole-genome sequencing

Keywords: fowlpox, poxvirus, next-generation sequencing, Oxford Nanopore MinION

Abstract: (47 words)

Fowlpox lesions were sampled on hens, total DNA was extracted and processed for sequencing on a MinION, Oxford Nanopore. Reads up to 61 kbp were generated and readily assembled into fowlpox virus complete genomes, suggesting that surveillance of emerging poxviruses could benefit from real-time whole genome sequencing.

(1042 words)

Avian pox is an infectious disease caused by avipoxviruses (APV), resulting in cutaneous and/or tracheal lesions. This condition may have a major economic impact in gallinaceous poultry (1) and is an emerging concern for wildlife, as evidenced for British great tits (2). Fowlpox is caused by different clades of avipoxviruses, the most important being clade A (fowlpox viruses) and clade B (canarypox viruses), both showing much more intra-clade genetic variability than previously assumed (3, 4, 5). Poxviruses share large genome sizes (from 130 to 360 kb) (6), featuring repetitions, deletions or insertions as a result of a long-term recombination history.

So far, avian pox diagnosis and surveillance are based on histopathological identification of typical cytoplasmic inclusions, isolation on chorioallantoic membrane (CAM) of chicken embryos and PCR targeting one or multiple viral genes (7). Very few APV whole genomes are available so far in public databases (i.e. 7 complete genomes of in NCBI GenBank as of the 30th of August, 2017). The increasing performances of next-generation sequencing (NGS) opened new opportunities for surveillance of poxviruses, based on timely and affordable workflows. One major drawback of 1st and 2nd generation NGS technologies is their limited read length, making the assembly of complex and large viral genomes very difficult. We investigated the application of the 3rd generation Oxford Nanopore Minion technology to achieve real-time whole-genome sequencing directly from lesions, without any enrichment or isolation step.

The study

During fall 2015, avian pox was diagnosed in a commercial layer farm in western France. This farm showed very good management and biosecurity. Mortality increased suddenly and reached 2% daily. Diseased birds developed severe dyspnea and suffocation before death. The

main macroscopic lesion was a very severe hyperplastic and necrotizing tracheitis (Figure 1A). Almost no cutaneous lesion could be observed in any bird.

In spring 2016, two independent broilers breeders farms were diagnosed for cutaneous fowlpox in Southwest France: affected birds showed conjunctivitis and nodular hyperplastic lesions on combs and wattles (Figure 1B), moderate egg drop and mild to moderate mortality.

All flocks were submitted to a comprehensive clinical investigation and tissue sampling.

Histopathology performed on tracheal and cutaneous lesions showed severe epithelial hyperplasia and squamous metaplasia, with superficial necrosis, and a myriad of cytoplasmic eosinophilic inclusions in epithelial cells, characteristic of poxvirus infection (Figure 1C and 1D). Tracheal swabs (cases 1 and 2) or cutaneous tissues (cases 3 and 4) sampled on diseased birds, tested positive for p4b locus shared by all avipoxviruses as well as for *fpv140* gene, distinguishing fowlpox virus and canarypox virus (data not shown).

A first test run was performed to validate DNA extraction process, library preparation and sequencing. The sample was obtained from a field isolate propagated on a CAM. 39 625 out of the 502 171 generated sequences were fowlpox-specific and allowed a whole genome alignment having 638X mean depth. Sequencing of non-enriched pox lesions with the MinIon is therefore feasible.

Then, seven other samples: tracheal tissue (n=2), lung tissue (n=1), central nodular cutaneous tissue (n=2), peripheral nodular cutaneous tissue (n=2) were sequenced in a unique run. Briefly, total DNA was checked for quality, libraries were prepared using the ONT PCR Barcoding kit and loaded on a R 9.4 flowcell. Sequence data were analyzed using Albacore, Poretools, bwa and the Samtools package. (See Technical Appendix for details).

All samples contained avipoxvirus sequences and complete whole genome alignments could be obtained from three out of the seven samples, corresponding to two tracheal and one

central nodular cutaneous sample (Table 1; Figure 2). The three complete genomes were (annotated using GATU (8) and deposited in GenBank under accession numbers MF766432, MF766431 and MF766430, for cutaneous lesion 16069, respiratory lesion 16069 and respiratory lesion 16055, respectively.

Due to poor pox DNA abundance in the lung sample or in one of the peripheral nodular cutaneous sample and the simultaneous excess of cellular DNA, we could not obtain a whole genome alignment from these tissues. Sequence data after quality filtering and sequence alignment is summarized in Table 1. Briefly, the sequencing of the cutaneous sample produced 3 905 fowlpox virus sequences out of 163 824; the sequencing of the tracheas generated 3 265 and 2 234 fowlpox virus sequences out of 124 534 and 141 670. The longest mapped read was obtained with the cutaneous sample and sized 61 kb.

Conclusions

We readily generated whole APV genomes from cutaneous or tracheal lesions, without any isolation or PCR-based enrichment. These genomes are highly similar to the prototype fowlpox reference strain NC_002188.1, including its REV insertion. The significance of REV remains unclear: it could putatively induce an increased virulence (1) or poor fowlpox vaccine efficacy (9,10) but it seems actually much common in fowlpoxviruses (11). Here again, a genome-based surveillance will be of great help to assess genetic diversity and significance of these retroviral insertions.

Our hypothesis was that active poxviral lesions should show significant viral-to-cellular ratios, which makes possible direct sequencing without need of enrichment. Indeed, fowlpox virus reads loads ranged from 0.75% to 2.62 %. The long read size eases the assembly step and lowers the bioinformatics capacity requirements and processing time compared to huge sets of 2nd generation short reads. We had tentatively included peripheral cutaneous lesions or

lung tissue in respiratory forms, but viral loads were much lower in both tissues, which is in accordance with the pathological pictures.

ONT MinION is an emerging NGS platform, able to generate up to 10 Gigabases by flowcell within a few hours, with no need of highly specialized laboratory environment. The poor sequencing fidelity compared to other NGS platforms is still a well-known drawback of Oxford Nanopore technology, which still makes difficult to generate a publication-ready sequence quality. Anyway, MinION reads improve in quantity and quality, making new applications accessible for veterinary surveillance. Furthermore, regarding applications in the field of diagnostic and epidemiological surveillance of pathogens, a remaining rate of nucleotides mismatches can be accepted since it does not impact typing and decision-making.

Based on our findings, ONT could be readily applied to whole-genome sequencing of other poxviruses, such as Myxoma virus in rabbits, Lumpy skin disease virus in ruminants, and potentially zoonotic poxviruses, such as Monkeypoxvirus. Global surveillance of emerging poviruses and comprehension of their evolution will dramatically benefit from this increasing availability of complete genomes.

Acknowledgments

This work was performed in the framework of the LabCom VIRAL, associating ENVT, INRA and FILAVIE and funded by the French National Research Agency (ANR).

References

1. Zhao K, He W, Xie S, Song D, Lu H, Pan W, Zhou P, Liu W, Lu R, Zhou J, Gao F. Highly pathogenic fowlpox virus in cutaneously infected chickens, China. *Emerg Infect Dis.* 2014 Jul;20(7):1208-10
2. Lawson B, Lachish S, Colvile KM, Durrant C, Peck KM, Toms MP, Sheldon BC, Cunningham AA. Emergence of a novel avian pox disease in British tit species. *PLoS One.* 2012;7(11):e40176
3. Le Loc'h G, Ducatez MF, Camus-Bouclainville C, Guérin JL, Bertagnoli S. Diversity of avipoxviruses in captive-bred Houbara bustard. *Vet Res.* 2014 Oct 1;45:98.
4. Bányai K, Palya V, Dénes B, Glávits R, Ivanics É, Horváth B, Farkas SL, Marton S, Bálint Á, Gyuranecz M, Erdélyi K, Dán Á. Unique genomic organization of a novel Avipoxvirus detected in turkey (*Meleagris gallopavo*). *Infect Genet Evol.* 2015 Oct;35:221-9.
5. Gyuranecz M, Foster JT, Dán Á, Ip HS, Egstad KF, Parker PG, Higashiguchi JM, Skinner MA, Höfle U, Kreizinger Z, Dorrestein GM, Solt S, Sós E, Kim YJ, Uhart M, Pereda A, González-Hein G, Hidalgo H, Blanco JM, Erdélyi K. Worldwide phylogenetic relationship of avian poxviruses. *J Virol.* 2013 May;87(9):4938-51.
6. Lefkowitz EJ, Wang C, Upton C. Poxviruses: past, present and future. *Virus Res.* 2006; 117(1):105-18.
7. Lecis R, Secci F, Antuofermo E, Nuvoli S, Scagliarini A, Pittau M, Alberti A. Multiple gene typing and phylogeny of avipoxvirus associated with cutaneous lesions in a stone curlew. *Vet Res Commun.* 2017 Jun;41(2):77-83.

8. Tcherepanov, T., Ehlers, A. and Upton, C. Genome Annotation Transfer Utility (GATU): Rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics* 2006 7:150.
9. Singh P, Kim JH, Tripathy DN. Re-emerging fowlpox: Evaluation of isolates from vaccinated flocks. *Avian Pathol.* 2000;29(5):449–55.
10. Koo BS, Lee HR, Jeon EO, Jang HS, Han MS, Min KC, Lee SB, Kim JJ, Mo IP. An outbreak of lymphomas in a layer chicken flock previously infected with fowlpox virus containing integrated reticuloendotheliosis virus. *Avian Dis.* 2013 Dec;57(4):812-7.
11. Tadese T, Fitzgerald S, Reed C. Detection and differentiation of re-emerging fowlpox virus (FWPV) strains carrying integrated reticuloendotheliosis virus (FWPV-REV) by real-time PCR. *Vet Microbiol.* 2008, 127:29-49

Tables and figures

Table 1 Sequence data after quality filtering and sequence alignment

	<i>Filtered reads</i>	<i>Reads mean length (pb)</i>	<i>Reads mapping fowlpox genome</i>	<i>Identity % vs ref. genome (NC_002188.1)</i>	<i>Mean depth</i>	<i>Longest mapped read (bp)</i>
<i>Amplified isolate on CAM</i>	502 171	3 630	39 625	100%	638X	77 063
<i>Cutaneous lesion</i>	163 824	1 352	3 905	99,7	22X	61 912
<i>Respiratory lesion 16069</i>	124 534	2 568	3 265	99,8	30X	49 439
<i>Respiratory lesion 16055</i>	141670	3 253	2234	99,8	19X	37 780

Figure legends

Figure 1. Pathological appearance of fowlpox cases. Gross lesions of respiratory (1A) and cutaneous fowlpox (1B); note hyperplastic nodules coated by a fibrino-necrotic exudate on the trachea (1A), and small and coalescent proliferative lesions on the comb and eyelids (1B). Microscopic lesions of respiratory pox: severe proliferative, hyperplastic and metaplastic, incompletely obstructive lesion of the tracheal mucosa (x20 magnification, 1C) with numerous typical cytoplasmic eosinophilic inclusions (arrows) (x200 magnification, 1D).

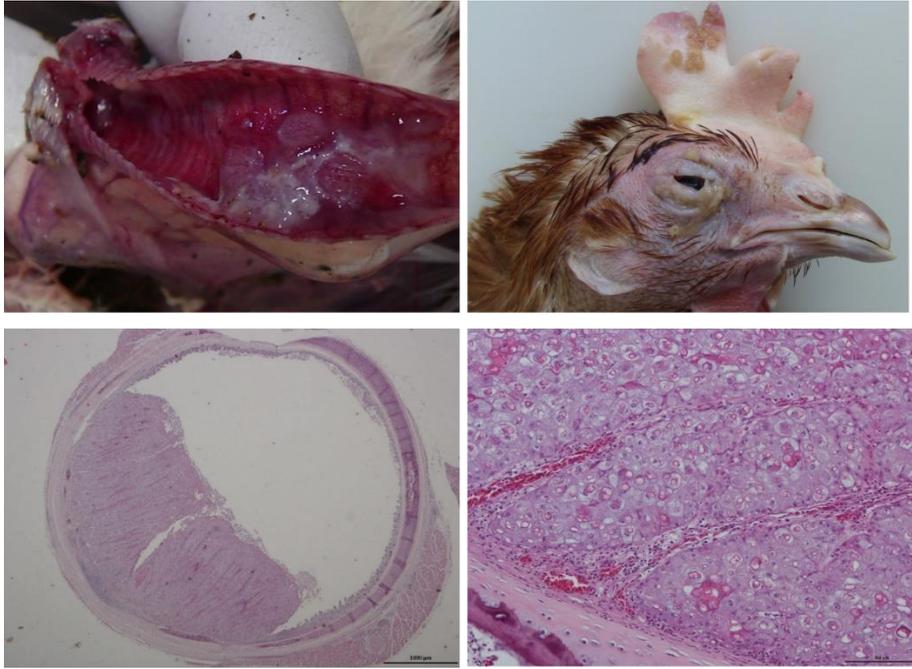
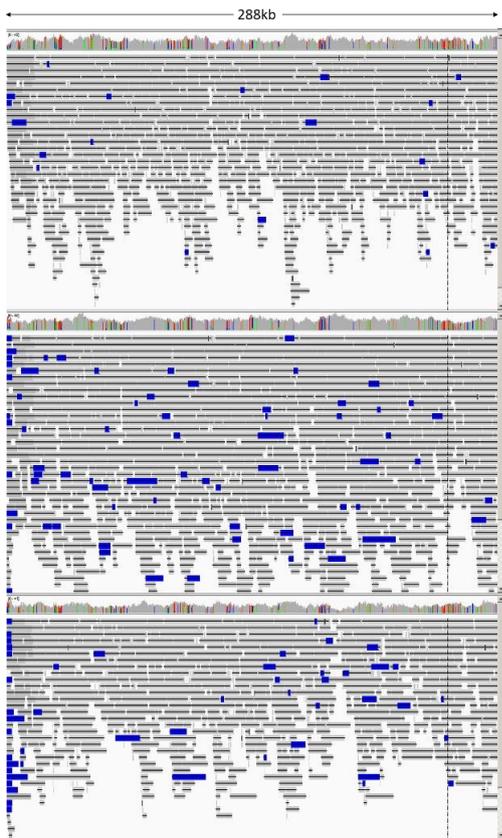


Figure 2. Sequence alignment display on the reference fowlpox genome with IGV



Technical appendix: protocols

DNA extraction

Total DNA was extracted from tissue culture or from the collected lesions after RNase A treatment following manufacturer's instructions (ThermoFisher scientific). The extraction protocol was derived from Sambrook J and Russell DW (2001). Briefly, tissue pieces were incubated in SNET lysis buffer (made of Tris-Cl, EDTA, NaCl, SDS and TE) supplemented with proteinase K at a 0,4 mg/mL final concentration. Then, the preparation was incubated overnight at 55°C with agitation in a shaking incubator. The DNA extraction was performed using phenol:chloroform:isoamyl alcohol, then the extracted DNA was precipitated by adding isopropanol before pelleting by centrifugation. DNA recovery was made after rinse with 70% ethanol and dissolving in TE. DNA quantitation for each sample was done by OD measurement and the purity of the solution was estimated considering the 260/280 ratio which is supposed to average 1.8 for a pure DNA solution. After that, DNA sample were directly submitted to MinION library preparation without any enrichment step.

MinION library preparation

Samples were prepared for sequencing following the 1D Native barcoding genomic DNA protocol (EXP-NBD103 and SQK-LSK108, ONT). 1.7µG DNA was sheared in a Covaris g-TUBE centrifuged at 7200 rpm for 2 minutes, inverting the tube between centrifugation steps. Then, a 0.4X AMPure beads purification has been performed. DNA repair, end repair and A-tailing (M6630 and E7546, NEB) were performed on DNA (average size of 8kb) and each step followed by a 0.5X AMPure beads purification. Native barcodes have been ligated using the Blunt/TA ligase master mix (M0367, NEB). Equimolar amounts of each barcoded sample have been pooled to obtain a 700ng final quantity (~ 0.13pmols of DNA). Adapters have been ligated

using the NEBNext Quick ligation Module (E7546, NEB). A 1X AMPure beads purification followed both of the ligation steps.

MinION sequencing

MinION sequencing was performed as per manufacturer's guidelines using R9.4 flowcells (FLO-MIN106, ONT) and ran for up to 48 h. The MUX scan reported 1350 and 1434 active pores for each flowcell respectively.

Data analysis

The raw sequence data under fast5 format was loaded to Albacore (v1.1.0) for base calling and demultiplexing. After being extracted from the fast5 files by Poretools (Loman and Quinlan 2014), each fastq file obtained was aligned against the avipox genome NC_002188.1 employed as a reference using bwa (Li and Durbin 2009) and the alignments were manipulated with the SAMtools software package (Li 2009). The resulting sorted files were then displayed within the IGV browser (Robinson 2011). Basic statistics concerning the aligned sequences were obtained by analysis of the bam files using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads mapping fowlpox virus genome and the mean sequencing depth were counted with the flagstat option and the depth option of the SAMtools, respectively. Consensus sequence was generated by the IGV browser whose method is derived from Cavener, 1987.

Each consensus sequence was annotated with GATU (Tcherepanov 2006), using Fowlpox virus genome NC_002188.1 as a reference, and deposited in GenBank.

Discussion générale et perspectives

La détection des agents pathogènes est un sujet d'intérêt majeur, directement en lien avec la santé animale et humaine. Les émergences et réémergences de maladies infectieuses sont fréquemment au cœur de l'actualité, y compris les maladies vectorisées par les arthropodes piqueurs dont la dispersion géographique ne cesse d'augmenter en raison des évolutions climatiques et de la globalisation des transports commerciaux (Gould et al., 2017).

La capacité à détecter les micro-organismes fait l'objet d'une « course à l'armement » qui s'appuie sur des innovations technologiques permanentes. La surveillance des maladies infectieuses à des fins de prévision d'émergence peut également reposer sur le suivi des « mégadonnées » ou « *big data* » avec, par exemple, une veille des épisodes infectieux réalisée sur les réseaux sociaux ou sur les requêtes soumises au moteur de recherches *Google* (Simonsen et al., 2016). Les avancées sur les plans de la miniaturisation des procédés, des nanotechnologies et des volumes réactionnels ultra-réduits s'ajoutant aux systèmes connectés à distance (*cloud*) sont à l'origine d'une diminution des coûts d'utilisation et de systèmes intuitifs et performants (Zarei, 2017).

Les animaux d'élevage sont évidemment sensibles aux maladies infectieuses et, sur le principe, les moyens mis en œuvre pour les identifier sont similaires à ceux déployés pour la santé humaine. Toutefois, le secteur de la santé animale doit obéir à des contraintes économiques, ce qui limite l'offre de solutions de détection de micro-organismes efficaces.

L'objectif de ces travaux était de développer et évaluer l'application de trois méthodologies de détection d'agents infectieux, dans trois contextes distincts. Les techniques développées dans ce manuscrit peuvent dans certains cas être complémentaires pour une même application, mais leur sélection ne pourra se faire en l'absence d'une expérience significative permettant un choix rationnel et justifié.

- Le criblage d'échantillons respiratoires de volailles par PCR quantitative à haut-débit Biomark a révélé des profils de co-infections bactériennes et/ou virales, voire fongiques. La description et l'analyse de ces coinfections respiratoires a été introduite en pneumologie humaine il y a moins de 10 ans, grâce à la mise à disposition de kits de criblage « multiplex ». De fait, l'introduction de ces tests a profondément amélioré la connaissance des associations d'agents pathogènes dans les voies respiratoires hautes ou basses, même si leur impact sur la prise en charge effective des patients est discuté. En médecine vétérinaire, les contraintes économiques sont évidemment plus fortes et rendent obligatoires la recherche de solutions les

plus économiques, tout en respectant les critères de qualité des analyses de diagnostic microbiologique.

Ces travaux sont l'exemple même que la gestion des maladies respiratoires ne peut pas être abordée sous l'angle « une maladie = un agent infectieux ». Il est donc capital que des outils de criblage à large spectre et robustes soient développés et mis au service du diagnostic vétérinaire.

Le point critique, puisqu'il s'agit de PCR quantitative en temps réel, réside dans la validation et la mise à jour des amorces de PCR employées. Les travaux présentés dans ce manuscrit rentrent dans le cadre d'une preuve de concept à des fins de recherche clinique et non de diagnostic. La validation des amorces n'a pas été soumise à un processus strict et systématique de détermination de limite de détection et d'efficacité. Dans la continuité de ce projet, il sera nécessaire que les tests de validation énoncés précédemment soient effectués pour les couples d'amorces déjà employés et organisés automatiquement pour l'inclusion de nouveaux couples d'amorces. En ce sens, les recommandations MIQE (Bustin et al., 2009) sont pertinentes et doivent être appliquées pour améliorer les pratiques expérimentales en PCR quantitative et, ainsi, obtenir des résultats robustes, fiables et publiables. Dès lors, il sera envisageable de proposer en routine l'analyse d'échantillons respiratoires de volailles par PCRq biomark comme outil d'étude de cohortes pour la description de co-infections dans des contextes épidémiologiques particulier.

- Par opposition à la recherche d'une liste de micro-organismes déjà établie, la recherche de l'agent étiologique responsable de l'entérite foudroyante de la pintade s'est faite par métagénomique. Le séquençage sans *a priori* des acides nucléiques viraux à partir de prélèvements issus de sujets malades a mis en évidence la présence d'un coronavirus aviaire. Suite à des analyses complémentaires par séquençage Sanger, son génome a présenté une proximité nucléotidique avec le coronavirus TCoV responsable d'une entérite chez la dinde. Cette étude est une illustration typique de l'apport de la métagénomique dans l'identification et la caractérisation d'agents infectieux « inconnus », quand les informations cliniques, pathologiques et les méthodes traditionnelles de virologie ne permettent pas d'orienter vers un ou quelques virus candidat(s).

Suite au succès de cette recherche, la même méthodologie a été appliquée à des prélèvements respiratoires effectués sur des canards présentant des signes cliniques respiratoires aigus. Les données générées ont ensuite été soumises au même traitement bioinformatique que celles concernant l'entérite de la pintade, avec l'algorithme GAAS (Angly et al., 2009). Cette étude

descriptive est encourageante au vu des difficultés rencontrées en métagénomique respiratoire, qu'il s'agisse du fort bruit de fond lié aux séquences de l'hôte ou des faibles charges virales (Li et al., 2016a). Le criblage des grandes familles virales par bioinformatique a révélé moins de 10 séquences de métapneumovirus aviaire de type C (aMPVc) dans 3 échantillons sur 10. L'un des échantillons présentait cependant 389 séquences d'aMPVc sur un total de 301 524 séquences.

La présentation des résultats sous forme de krona (diagramme circulaire interactif illustrant les familles virales identifiées et permettant de préciser l'identification jusqu'à l'espèce virale) est très ergonomique mais ne présente que l'affiliation taxonomique ainsi que le nombre de séquences y correspondant. Il n'est pas rationnel de présenter si peu de données car les séquences identifiées comme appartenant à une espèce virale donnée peuvent potentiellement correspondre à des séquences virales endogènes intégrées dans le génome de l'hôte. L'apport significatif de l'outil VIP (Li et al., 2016b) concerne l'information de couverture de génome qu'il propose. En effet, dans le cas de l'aMPVc évoqué précédemment, VIP annonce un pourcentage de couverture de 15%, ce qui oriente vers une identification « valide ». Cependant, VIP n'affilie que 56 séquences à l'aMPVc contre 400 avec GAAS. Il s'avère alors nécessaire de confronter directement les fichiers de séquençage au génome de référence suspecté par une méthode d'alignement avec l'outil BWA (Li and Durbin, 2009) et de visualiser les résultats sous forme graphique à l'aide d'IGV (Robinson et al., 2011).

- L'industrie du séquençage compte aujourd'hui trois générations, la première est le séquençage Sanger (Sanger et al., 1992), la seconde correspond aux nouvelles générations de séquençage nécessitant une amplification clonale (de type Illumina) et la troisième génération est la technique NGS qui permet le séquençage de molécules uniques sans amplification clonale préalable (de type Oxford Nanopore Technologies, ONT ou Pacific Bio).

Dans le cadre d'un projet pilote d'évaluation du séquenceur MinION (ONT) auprès de la plateforme de séquençage de l'INRA de Castanet Tolosan, nous avons souhaité répondre à une problématique depuis longtemps présente dans nos esprits, à savoir : est-il possible de caractériser sur le plan génétique un agent infectieux en soumettant directement la lésion clinique qu'il a induite au séquençage NGS ? En guise de mise au point méthodologique, nous avons dans un premier temps soumis au séquençage une culture avec succès. Ce rapide retour d'expérience nous a permis de procéder de la même manière avec les lésions cutanées et d'obtenir des génomes complets d'avipoxvirus depuis une extraction des ADN totaux.

En plus de l'innovation technologique proposée par ONT, la miniaturisation et les faibles coûts de l'équipement (le séquenceur est accessible pour moins de 900€) modifient le paradigme des NGS en les rendant nomades. À ce tarif, nulle nécessité d'amortissement, le rêve de tout biologiste « amateur de NGS » est à portée de main : on peut faire du séquençage NGS dans son laboratoire, à des tarifs abordables. De plus, la politique de communication compulsive d'Oxford Nanopore sur le réseau social à l'oiseau bleu a été terriblement efficace. Le message de la simplicité d'utilisation est très bien passé, au point de voir bon nombre de scientifiques s'équiper sans expérience préalable avant de se tourner vers les plateformes de génomique, faute de succès en autonomie (retour d'expérience de la plateforme de séquençage de de Castanet Tolosan).

Notre expérience acquise sur les NGS et la biologie moléculaire appliquées à la virologie nous incite à insister sur la nécessité de ne pas négliger la phase de mise au point préalable des techniques. Aussi tentant que cela puisse être, il est finalement déraisonnable de s'engager dans l'utilisation d'une technologie nouvelle sans en avoir levé les potentiels verrous méthodologique en amont. Lorsqu'un projet consiste à décrire les populations de micro-organismes d'un milieu complexe par métagénomique, le premier questionnement doit concerner la quantité de séquences à consacrer à un échantillon. Pour ce faire, une technique consiste à contaminer (ou « spiker ») un échantillon témoin avec des quantités connues d'ADN suivant la méthode de dilutions limites et de soumettre cet échantillon test à une réaction de séquençage NGS. Il est alors possible de déterminer la quantité de séquences nécessaires pour atteindre les limites de détection de techniques de référence, comme la PCR.

L'analyse des séquences par bioinformatique est un enjeu majeur, d'autant plus que la manipulation de jeux de données si massifs relève plus de l'informatique que de la biologie. Jusqu'à très récemment, l'analyse de données NGS était réservée aux bioinformaticiens de métier car les biologistes n'ont, pour la plupart, pas les compétences nécessaires pour sélectionner les algorithmes pertinents, les comparer et valoriser les résultats produits. Quelques années se sont écoulées pendant lesquelles tout biologiste s'essayant à la bioinformatique peinait sérieusement à exploiter ses données de façon autonome. Fort heureusement, des bioinformaticiens conscients de l'état de frustration des biologistes impuissants face à leurs données de NGS se sont décidés à développer des outils d'analyses automatisés capables de réaliser une analyse intégrale sans que l'utilisateur ait à composer avec chaque étape de calcul individuelle. En 2016, 36 articles scientifiques comportant les termes « metagenomic »,

« bioinformatics » et « pipeline » ont été référencés sur la base de données Pubmed alors qu'il y en avait que deux en 2010.

En définitive, au vu de l'évolution permanente des outils d'analyse bioinformatique, l'accompagnement par des bioinformaticiens apparaît indispensable afin de disposer en permanence des outils les plus adaptés à ses analyses. Aussi, une fois l'équilibre trouvé entre les méthodologies de séquençage et d'analyse bioinformatique, il convient de l'exploiter au maximum en dépit de l'évolution constante et de l'offre en outils d'analyses.

Les travaux détaillés dans cette partie expérimentale répondent à des questions de recherche intégrées dans un projet de thèse CIFRE avec le laboratoire pharmaceutique Filavie. Le projet initial était de développer des outils de biologie moléculaire destinés à la validation de la pureté microbiologique de préparations vaccinales afin de garantir la présence unique de l'agent vaccinal. La vérification de l'identité nucléotidique de la souche vaccinale par biologie moléculaire entre également dans ce projet puisqu'une dérive génétique ou l'émergence d'une sous-population variante peuvent être à l'origine d'accidents de vaccination.

Aujourd'hui, le développement de couples d'amorces de PCRq dédiés à la recherche des micro-organismes listés par la Pharmacopée Européenne est toujours en cours et la chimie employée s'oriente vers la technologie Taqman afin de répondre davantage aux exigences de spécificité de l'entreprise partenaire. L'outil Biomark initialement envisagé pour effectuer ces analyses ne se révèle finalement pas adapté à un contrôle « au fil de l'eau » des lots de semences virales de l'entreprise car le débit d'échantillons à analyser n'est pas suffisamment élevé pour rentabiliser l'utilisation d'une plaque de 2 304 réactions. Le criblage d'un échantillon vis-à-vis des 21 agents à rechercher peut en fait être réalisé sur une plaque de qPCR conventionnelle.

Au vu de l'évolution des techniques de NGS, il est raisonnable d'envisager que la validation de la pureté microbiologique ainsi que la vérification de l'identité nucléotidique des souches vaccinales soient, à court terme, entreprises par NGS. Le séquenceur MinION paraît adapté à ce type d'application car (i) la longueur des séquences générées permet d'obtenir des génomes complets à partir de quelques séquences seulement et (ii) la quantité de séquences générées (jusqu'à 10 millions) est compatible avec une application de métagénomique pour la recherche des agents listés par la Pharmacopée Européenne et pour la recherche de sous-populations virales minoritaires contaminantes. Des risques liés à la vaccination ont déjà été décrits comme l'association du syndrome de Guillain-Barré avec certains vaccin Influenza mais aussi des vaccins à poliovirus contaminés par le virus simien SV40 et sa responsabilité dans le

développement de cancers (Committee et al., 2002). Il est donc nécessaire d'être en mesure de détecter tout micro-organisme présent dans un échantillon avec des performances de détectabilité au moins égales à celles des techniques de référence. La méthodologie à mettre en place consiste, comme décrit précédemment, à inoculer une solution test avec des agents infectieux de diverses natures (bactéries, virus enveloppés ou non, génome à ADN, génome à ARN mono ou bicaténaire) en dilutions limites et de déterminer quelle est la quantité minimale détectable par séquençage MinION. Ainsi, lors d'analyses de routine, il conviendra de contaminer systématiquement les échantillons avec un ou des agent(s) infectieux dont la présence dans l'échantillon est improbable, cela fera office de témoin. De plus, l'une des innovations à venir étant le « *base-calling* » en temps réel, il sera possible d'arrêter l'analyse et de fournir un résultat dès lors que les séquences correspondant à la limite de détection basse auront été obtenues.

Le diagnostic en médecine vétérinaire est à l'aube d'un tournant méthodologique grâce à la démocratisation des NGS et leur adaptation au domaine du diagnostic médical, qui requiert des délais d'analyse courts et des coûts réduits.

L'association de l'enrichissement en acides nucléiques et du séquençage NGS semble être une stratégie pertinente pour résoudre les difficultés rencontrées en métagénomique tant sur le plan de la préparation des échantillons (pour prendre en compte le faible pourcentage de séquences virales dans un fluide corporel, un tissu, une préparation vaccinale ou un prélèvement environnemental) que sur le plan de l'analyse des données par bioinformatique (pour, là encore, être capable de détecter de rares séquences virales). L'analyse bioinformatique sera en effet plus aisée car la recherche de séquences que l'on a cherché à sélectionner ou amplifier est plus aisée qu'une recherche sans *a priori*.

À titre d'exemple, la société Agilent propose la solution SureSelect dont le principe d'enrichissement en amont du séquençage repose sur une capture des acides nucléiques *via* des sondes biotinylées. ThermoFisher Scientific a, de son côté, développé le système Ion AmpliSeq dont l'enrichissement est basé sur de la PCR grâce à l'association de couples d'amorces compatibles avec une amplification en mélange en amont du séquençage. L'atout majeur du système Ion AmpliSeq est son couplage au séquenceur du même constructeur (PGM ou Proton de ThermoFisher Scientific) ainsi que l'existence d'un outil d'analyse dédié. Des panels d'amorces sont disponibles pour de l'analyse d'expression de gènes, de la recherche de

mutations et modifications de gènes impliquées le développement de certains cancers ou encore de la recherche de modifications génétiques impliquées dans les maladies héréditaires (ampliseq panel).

Enfin, la technique Cas9-Assisted Targeting of CHromosome segments (CATCH) (Jiang et al., 2015) associée à du séquençage de long fragments est une solution prometteuse dans la démarche d'enrichissement de génomes sans limite de taille, ni introduction de biais, ni besoin de connaissance précise du gène recherché. Il s'agit de l'excision d'une région génétique grâce à la nucléase Cas9 guidée par un ARN. Le produit de digestion est ensuite déposé sur gel à champ pulsé puis la fenêtre correspondant à la taille de la séquence nucléotidique est purifiée et les acides nucléiques obtenus peuvent être soumis au séquençage MinION (Gabrieli et al., 2017).

Le séquençage et la PCR à haut débit sont des techniques applicables à la détection et à la caractérisation d'agents pathogènes et au contrôle de pureté microbiologique des vaccins vétérinaires. Ces travaux de thèse apportent une contribution à l'évaluation des performances analytiques de ces nouveaux outils, actuelles et en perspective.

Il n'y a aucun doute sur le fait que ces analyses diagnostiques doivent être développées sous couvert d'une validation méthodologique robuste. Au-delà des performances au sens strict, les technologies innovantes décrites dans ce travail offrent également la perspective d'une ergonomie et d'un confort de travail considérables, tout en permettant, quand c'est pertinent, la délocalisation de l'analyse du laboratoire vers le terrain.

Il est difficile de prévoir quel sera le prochain « bon » technologique mais il s'inscrira fort probablement dans cette dynamique d'accessibilité, de rapidité et de performance toujours accrues.

Il restera à veiller à ce que ces outils soient utilisés avec discernement, au bénéfice de la dynamique « *One Health* » visant à gérer les risques sanitaires aux interfaces animal-homme-écosystèmes.

Références Bibliographiques

Ackermann, H.-W., and Tiekotter, K.L. (2012). Murphy's law-if anything can go wrong, it will: Problems in phage electron microscopy. *Bacteriophage* 2, 122–129.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44, 3–10.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Andral, B., Lagadic, M., Louzis, C., Guillou, J., and Gourreau, J. (1987). Fulminating disease in guinea fowl: etiological investigation. *Point Vétérinaire* 19, 515–520.

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., et al. (2009). The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput. Biol.* 5, e1000593.

Araujo, S., Goulart, L.R., Truman, R.W., Goulart, I.M.B., Vissa, V., Li, W., Matsuoka, M., Suffys, P., Fontes, A.B., Rosa, P.S., et al. (2017). qPCR-High resolution melt analysis for drug susceptibility testing of *Mycobacterium leprae* directly from clinical specimens of leprosy patients. *PLoS Negl. Trop. Dis.* 11, e0005506.

Assoumou, L., Charpentier, C., Recordon-Pinson, P., Grudé, M., Pallier, C., Morand-Joubert, L., Fafi-Kremer, S., Krivine, A., Montes, B., Ferré, V., et al. (2017). Prevalence of HIV-1 drug resistance in treated patients with viral load >50 copies/mL: a 2014 French nationwide study. *J. Antimicrob. Chemother.* 72, 1769–1773.

Bagust, T. (2008). Poultry health and disease control in developing countries. *Poultry development review* 1, 1-5.

Beer, M., Reimann, I., Hoffmann, B., and Depner, K. (2007). Novel marker vaccines against classical swine fever. *Vaccine* 25, 5665–5670.

Bergallo, M., Costa, C., Gribaudo, G., Tarallo, S., Baro, S., Negro Ponzi, A., and Cavallo, R. (2006). Evaluation of six methods for extraction and purification of viral DNA from urine and serum samples. *New Microbiol.* 29, 111–119.

Brahem, A., Demarquez, N., Beyrie, M., Vuillaume, A., and Fleury, H.J. (1992). A highly virulent togavirus-like agent associated with the fulminating disease of guinea fowl. *Avian Dis.* 36, 143–148.

Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517, 205–208.

- Bustin, S.A., Benes, V., Garson, J.A., Hellems, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* *55*, 611–622.
- Byrd, A.L., and Segre, J.A. (2016). Infectious disease. Adapting Koch's postulates. *Science* *351*, 224–226.
- Cai, L., Yang, Y., Jiao, N., and Zhang, R. (2015). Evaluation of Tangential Flow Filtration for the Concentration and Separation of Bacteria and Viruses in Contrasting Marine Environments. *PloS One* *10*, e0136741.
- Capua, I., Terregino, C., Cattoli, G., Mutinelli, F., and Rodriguez, J.F. (2003). Development of a DIVA (Differentiating Infected from Vaccinated Animals) strategy using a vaccine containing a heterologous neuraminidase for the control of avian influenza. *Avian Pathol. J. WVPA* *32*, 47–55.
- Chen, E.C., Miller, S.A., DeRisi, J.L., and Chiu, C.Y. (2011). Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens. *J. Vis. Exp.* *50*, e2536.
- Coelho, B., Veigas, B., Fortunato, E., Martins, R., Águas, H., Igreja, R., and Baptista, P.V. (2017). Digital Microfluidics for Nucleic Acid Amplification. *Sensors* *17*, 1495.
- Committee, I. of M. (US) I.S.R., Stratton, K., Almario, D.A., and McCormick, M.C. (2002). Immunization Safety Review: SV40 Contamination of Polio Vaccine and Cancer (National Academies Press (US)) *1*, 1-102.
- Day, M.J. (2015). Introduction to Antigen and Antibody Assays. *Top. Companion Anim. Med.* *30*, 128–131.
- Delpuech, É. Analyse et co-développement d'outils de bioinformatique destinés au traitement de données NGS issues de métagénomique virale. Master de bioinformatique et de biologie des systèmes de Toulouse. Toulouse : Université Paul Sabatier, Toulouse III, 2017, 49 p.
- Dhoubhadel, B.G., Yasunami, M., Yoshida, L.-M., Thi, H.A.N., Thi, T.H.V., Thi, T.A.N., Watanabe, K., Suzuki, M., Morimoto, K., Dang, D.A., et al. (2014). A novel high-throughput method for molecular serotyping and serotype-specific quantification of *Streptococcus pneumoniae* using a nanofluidic real-time PCR system. *J. Med. Microbiol.* *63*, 528–539.
- Fox, E.J., Reid-Bayliss, K.S., Emond, M.J., and Loeb, L.A. (2014). Accuracy of Next Generation Sequencing Platforms. *Gener. Seq. Appl.* *1*, 1-9.
- Gabrieli, T., Sharim, H., Michaeli, Y., and Ebenstein, Y. (2017). Cas9-Assisted Targeting of CHromosome segments (CATCH) for targeted nanopore sequencing and optical genome mapping. *BioRxiv* e110163.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* *15*, 1451–1455.
- Gould, E., Pettersson, J., Higgs, S., Charrel, R., and de Lamballerie, X. (2017). Emerging arboviruses: Why today? *One Health Amst. Neth.* *4*, 1–13.

- Jain, M., Tyson, J.R., Loose, M., Ip, C.L.C., Eccles, D.A., O'Grady, J., Malla, S., Leggett, R.M., Wallerman, O., Jansen, H.J., et al. (2017). MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research* 6, 760.
- Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y., and Zhu, T.F. (2015). Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* 6, 8101.
- Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993.
- Karlsen, F., Steen, H.B., and Nesland, J.M. (1995). SYBR green I DNA staining increases the detection sensitivity of viruses by polymerase chain reaction. *J. Virol. Methods* 55, 153–156.
- Keller, I., Korner-Nievergelt, F., and Jenni, L. (2009). Within-winter movements: a common phenomenon in the Common Pochard *Aythya ferina*. *J. Ornithol.* 150, 483–494.
- Ko, D.H., Kim, H.S., Hyun, J., Kim, H.S., Kim, J.S., Park, K.U., and Song, W. (2017). Comparison of the Luminex xTAG Respiratory Viral Panel Fast v2 Assay With Anyplex II RV16 Detection Kit and AdvanSure RV Real-Time RT-PCR Assay for the Detection of Respiratory Viruses. *Ann. Lab. Med.* 37, 408–414.
- Korenková, V., Scott, J., Novosadová, V., Jindřichová, M., Langerová, L., Švec, D., Šídová, M., and Sjöback, R. (2015). Pre-amplification in the context of high-throughput qPCR gene expression experiment. *BMC Mol. Biol.* 16, 5.
- Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., and Studholme, D.J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760.
- Li, D., Li, Z., Zhou, Z., Li, Z., Qu, X., Xu, P., Zhou, P., Bo, X., and Ni, M. (2016a). Direct next-generation sequencing of virus-human mixed samples without pretreatment is favorable to recover virus genome. *Biol. Direct* 11.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079.
- Li, Y., Wang, H., Nie, K., Zhang, C., Zhang, Y., Wang, J., Niu, P., and Ma, X. (2016b). VIP: an integrated pipeline for metagenomics of virus identification and discovery. *Sci. Rep.* 6, 23774.
- Liais, E., Croville, G., Mariette, J., Delverdier, M., Lucas, M.-N., Klopp, C., Lluch, J., Donnadiou, C., Guy, J.S., Corrand, L., et al. (2014). Novel avian coronavirus and fulminating disease in guinea fowl, France. *Emerg. Infect. Dis.* 20, 105–108.
- Lipkin, W.I. (2013). The changing face of pathogen discovery and surveillance. *Nat. Rev. Microbiol.* 11, 133–141.

Loman, N.J., and Quinlan, A.R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinforma. Oxf. Engl.* 30, 3399–3401.

Loman, N.J., Constantinidou, C., Christner, M., Rohde, H., Chan, J.Z.-M., Quick, J., Weir, J.C., Quince, C., Smith, G.P., Betley, J.R., et al. (2013). A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxicogenic *Escherichia coli* O104:H4. *JAMA* 309, 1502–1510.

Maxam, A.M., and Gilbert, W. (1992). A new method for sequencing DNA. 1977. *Biotechnol. Read. Mass* 24, 99–103.

Mengelle, C., Mansuy, J.-M., Da Silva, I., Guerin, J.-L., and Izopet, J. (2013). Evaluation of a polymerase chain reaction-electrospray ionization time-of-flight mass spectrometry for the detection and subtyping of influenza viruses in respiratory specimens. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* 57, 222–226.

Michelet, L., Delannoy, S., Devillers, E., Umhang, G., Aspan, A., Juremalm, M., Chirico, J., van der Wal, F.J., Sprong, H., Boye Pihl, T.P., et al. (2014). High-throughput screening of tick-borne pathogens in Europe. *Front. Cell. Infect. Microbiol.* 4, 103.

Mongkolrattanothai, K., Naccache, S.N., Bender, J.M., Samayoa, E., Pham, E., Yu, G., Dien Bard, J., Miller, S., Aldrovandi, G., and Chiu, C.Y. (2017). Neurobrucellosis: Unexpected Answer From Metagenomic Next-Generation Sequencing. *J. Pediatr. Infect. Dis. Soc.* 6, 393–398.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1992). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. *Biotechnol. Read. Mass* 24, 17–27.

Murkey, J.A., Chew, K.W., Carlson, M., Shannon, C.L., Sirohi, D., Sample, H.A., Wilson, M.R., Vespa, P., Humphries, R.M., Miller, S., et al. (2017). Hepatitis E Virus-Associated Meningoencephalitis in a Lung Transplant Recipient Diagnosed by Clinical Metagenomic Sequencing. *Open Forum Infect. Dis.* 4, ofx121.

Pubmeddev Home - PubMed - NCBI. Disponible sur <https://www.ncbi.nlm.nih.gov/pubmed/> (consulté le 30.08.2017)

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* 12, 1261–1276.

Rampini, S.K., Bloemberg, G.V., Keller, P.M., Büchler, A.C., Dollenmaier, G., Speck, R.F., and Böttger, E.C. (2011). Broad-range 16S rRNA gene polymerase chain reaction for diagnosis of culture-negative bacterial infections. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 53, 1245–1251.

Reperant, L.A., Fuckar, N.S., Osterhaus, A.D.M.E., Dobson, A.P., and Kuiken, T. (2010). Spatial and temporal association of outbreaks of H5N1 influenza virus infection in wild birds with the 0 degrees C isotherm. *PLoS Pathog.* 6, e1000854.

- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Roingard, P. (2008). Viral detection by electron microscopy: past, present and future. *Biol. Cell* 100, 491–501.
- Russell, W.M.S., and Burch, R.L. (1959). *The principles of humane experimental technique* (London: Methuen). 1, 238
- Saah, A.J., and Hoover, D.R. (1998). [Sensitivity and specificity revisited: significance of the terms in analytic and diagnostic language]. *Ann. Dermatol. Venereol.* 125, 291–294.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1992). DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnol. Read. Mass* 24, 104–108.
- Simen, B.B., Braverman, M.S., Abbate, I., Aerssens, J., Bidet, Y., Bouchez, O., Gabriel, C., Izopet, J., Kessler, H.H., Stelzl, E., et al. (2014). An international multicenter study on HIV-1 drug resistance testing by 454 ultra-deep pyrosequencing. *J. Virol. Methods* 204, 31–37.
- Simonsen, L., Gog, J.R., Olson, D., and Viboud, C. (2016). Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J. Infect. Dis.* 214, 380–385.
- Smuts, H., Kew, M., Khan, A., and Korsman, S. (2014). Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants in silica column-based nucleic acid extraction kits. *J. Virol.* 88, 1398.
- Sorg, D., Danowski, K., Korenkova, V., Rusnakova, V., Küffner, R., Zimmer, R., Meyer, H.H.D., and Kliem, H. (2013). Microfluidic high-throughput RT-qPCR measurements of the immune response of primary bovine mammary epithelial cells cultured from milk to mastitis pathogens. *Anim. Int. J. Anim. Biosci.* 7, 799–805.
- Stoffan, M.A. (2004). Chasing Moore’s Law: Information Technology Policy in the United States. *Coll. Amp Res. Libr.* 66, 384–386.
- Toffan, A., Catania, S., Salviato, A., De Battisti, C., Vascellari, M., Toson, M., Capua, I., and Cattoli, G. (2012). Experimental infection of poult and guinea fowl with genetically distinct avian astroviruses. *Avian Pathol. J. WVPA* 41, 429–435.
- Tripathy, D., and Reed, W. (2013). *Pox. Diseases of poultry.* (John Wiley & sons), pp. 333–350.
- Van Riper, C., and Forrester, D. (2007). Avian pox. *Infectious diseases of wild birds.* pp. 131–176.
- Vandegrift, K.J., Sokolow, S.H., Daszak, P., and Kilpatrick, A.M. (2010). Ecology of avian influenza viruses in a changing world. *Ann. N. Y. Acad. Sci.* 1195, 113–128.
- Veits, J., Römer-Oberdörfer, A., Helferich, D., Durban, M., Suezer, Y., Sutter, G., and Mettenleiter, T.C. (2008). Protective efficacy of several vaccines against highly pathogenic H5N1 avian influenza virus under experimental conditions. *Vaccine* 26, 1688–1696.

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

Visser, M., Burger, J.T., and Maree, H.J. (2016). Targeted virus detection in next-generation sequencing data using an automated e-probe based approach. *Virology* 495, 122–128.

Worobey, M., Watts, T.D., McKay, R.A., Suchard, M.A., Granade, T., Teuwen, D.E., Koblin, B.A., Heneine, W., Lemey, P., and Jaffe, H.W. (2016). 1970s and “Patient 0” HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539, 98–101.

Zarei, M. (2017). Advances in point-of-care technologies for molecular diagnostics. *Biosens. Bioelectron.* 98, 494–506.

Zhang, X., Xu, G., Wang, C., Jiang, M., Gao, W., Wang, M., Sun, H., Sun, Y., Chang, K.-C., Liu, J., et al. (2017). Enhanced pathogenicity and neurotropism of mouse-adapted H10N7 influenza virus are mediated by novel PB2 and NA mutations. *J. Gen. Virol.* 98, 1185–1195.

Zou, N., Ditty, S., Li, B., and Lo, S.-C. (2003). Random priming PCR strategy to amplify and clone trace amounts of DNA. *BioTechniques* 35, 758–760, 762–765.

ANNEXE 1

Contrôles à mettre en œuvre pour la validation de pureté microbiologique de préparations vaccinales

Contrôles à mettre en œuvre pour la validation de pureté microbiologique de préparations vaccinales			Matière 1 ^{ère} origine animale				Lots de semence		Produit fini				
	Monographie/Texte pharmacopée	Principe	poulets SPF	culture cellulaire			autres substance (sérum , ...)			vaccin vivant		vaccin inactivé	
				cellules PR	banque cellules	Lignée		mmf	oiseaux	mmf	oiseaux	mmf	oiseaux
PURETE VIRALE	5.2.2 ELEVAGES DE POULETS SPF (PRODUCTION ET CONTRÔLE)		X	X									
	5.2.4 CULTURES CELLULAIRES (PREPARATION VACCIN)	- Abs de conta virale (ECP, hemadsorbtion) - virus spécifiques (espèce cible + source) - essais sur autres cultures cellulaires (cellules P de l'espèce source +cellules sensibles aux virus pathogènes espèce cible + cellules sensibles aux pestivirus)			X	X							
	5.2.5 SUBSTANCE MA (PREPARATION VACCIN)	Absence de virus étrangers vivants : - essai général (ECP , hémadsorbtion) - essai spécifique (selon analyse de risque) -pestivirus si espèce source réceptive					X						
	2262 SERUM BOVIN	Recherche de contaminants viraux : essai général et spécifique (selon pays d'origine)					(x)						
	2.6.24 LS VACCINS VIRAUX AVIAIRES	cf.TABLEAU 2							X				
	7blm10a (guideline)	liste des agents spécifiques / espèce						X					
	062 VACCINS VETERINAIRES	- Abs de conta virale (ECP, hemadsorbtion) cf 5.2.4 - essais sur autres cultures cellulaires (cellules P de l'espèce source +cellules sensibles aux virus pathogènes espèce cible + cellules sensibles aux pestivirus)						X					
	062 VACCINS VETERINAIRES	analyse de risque globale								X		X	X
	Monographie spé. vaccin							X	X	X	X	X	X
2.6.25 VACCINS VIVANTS AVIAIRES	cf.TABLEAU 3 (+ tab 4 : vaccin aviaire inactivé)										X		
STERILITE BACTERIENNE ET FONGIQUE	2.6.1				X	X	X	x		X		X	

STERILITE MYCOPLASMIQUE	2.6.7				X	X	X	X (culture +fluo)	X (culture)	
--------------------------------	-------	--	--	--	---	---	---	----------------------	----------------	--

Vaccins viraux aviaires : recherche des agents étrangers dans les lots de semence

		AGENTS	TECHNIQUE
RECHERCHE DIRECTE NON SPECIFIQUE	recherche des agents étrangers sur œufs de poules embryonnés	/	inoculation dans cavité allantoïdienne, mbre chorio-allantoïdienne, sac vitellin 2 passages ⇒ Abs de mortalité, abs d'anomalie macroscopique des embryons, examen normal de la membrane chorio-allantoïdienne et abs d'agents hémagglutinants dans liquide allantoïdien
	recherche des agents étrangers sur cultures de cellules rénales de poulet	/	inoculation à des cultures ⇒ Après coloration : abs d'effet cytopathogène, etc ⇒ Après ajout d'érythrocytes de poulet : abs d'agents hémadsorbants ou hémagglutinants
RECHERCHE DIRECTE SPECIFIQUE	recherche des virus de la leucose aviaire	virus de la leucose aviaire	culture sur DF-1/cellules PR ou SR FEP ⇒ Abs d'antigène spécifique de groupe de la leucose par COFAL (fixation du complément) ou ELISA
	recherche du virus de la réticulo-endothéliose aviaire	virus de la reticulo-endothéliose aviaire	culture sur cellules PR ou SR FEP ⇒ Recherche par immunomarquage
	recherche du virus de l'anémie du poulet	virus de l'anémie du poulet	culture sur lignée MDCC-MSBI ⇒ Recherche par immunomarquage

RECHERCHE INDIRECTE SPECIFIQUE	recherche des agents étrangers sur poussins	AGP précipitation sur gelose EIA (dosage immunoenz ex ELISA) HI inhibition de l'hémagglutination Aggl agglutination SN séroneutralisation IS immunomarquage (ex Ac fluo)	inoculation 10 poussins SPF de 2 semaines, suivi 5 s ⇒ Abs de signes cliniques ⇒ Abs d'anticorps contre les agents suivants :				
		SN	EIA	AGP	HI	Aggl	IS
toutes espèces aviaires	Adénovirus aviaires, grpe1	x	x	x			
	virus de l'encéphalomyélite aviaire		x	x			
	virus de la bronchite infectieuse aviaire		x		x		
	virus de la laryngotrachéite infectieuse aviaire	x	x				x
	virus de la leucose aviaire	x	x				
	virus de la néphrite aviaire						x
	orthoréovirus aviaire		x				x
	virus de la réticuloendothéliose aviaire		x	x			x
	virus de l'anémie du poulet	x	x				x
	virus du syndrome de la chute de ponte		x		x		
	virus de la bursite infectieuse aviaire (type) (Gumboro)	x(I,II)	x(I)	x(I)			
	virus A de la grippe		x	x	x		
	virus de la maladie de Marek			x			
	virus de la pseudopeste aviaire (NC)		x		x		
	virus de la rhinotrachéite du dindon		x				
salmonella pullorum					x		
+ dindon	Chlamydia spp		x				
	virus de l'entérite hémorragique infectieuse aviaire			x			
	paramyxovirus 3 aviaire				x		
	virus de la bursite infectieuse aviaire de type 2	x					
+ canard	chlamydia spp		x				
	parvovirus du canard et de l'oie	x	x				
	virus de l'entérite du canard	x					
	virus de l'entérite virale du canard de type 1	x					
+ oie	parvovirus du canard et de l'oie	x	x				
	virus de l'entérite du canard	x					

		polyomavirus de l'oie	inoculation sur oison
--	--	-----------------------	-----------------------

Vaccins viraux vivants aviaires : recherche des agents étrangers dans les lots de produit final

		AGENTS	TECHNIQUE
RECHERCHE DIRECTE NON SPECIFIQUE	recherche des agents étrangers sur œufs de poules embryonnés	/	inoculation dans cavité allantoïdienne, mbre chorio-allantoïdienne, sac vitellin 2 passages ⇒ Abs de mortalité, abs d'anomalie macroscopique des embryons, examen normal de la membrane chorio-allantoïdienne et abs d'agents hémagglutinants dans liquide allantoïdien
	recherche des agents étrangers sur cultures de cellules FEP	/	inoculation à des cultures ⇒ Après coloration : abs d'effet cytopathogène, etc ⇒ Après ajout d'érythrocytes de poulet : abs d'agents hémadsorbants ou hémagglutinants
RECHERCHE DIRECTE SPECIFIQUE	recherche des virus de la chute de ponte	virus de la chute de ponte	culture d'hépatocyte EP ⇒ Après ajout d'érythrocytes de poulet : abs d'agents hémagglutinants
	recherche du virus de la maladie de marek	virus de la maladie de marek	culture sur cellules PR ou SR FEP ⇒ Recherche par immunomarquage
	recherche du virus de la rhinotrachéite du dindon	virus de la rhinotrachéite du dindon	culture sur FEP ⇒ Recherche par immunomarquage culture sur cellules Vero ⇒ Recherche par immunomarquage
	recherche du virus de l'anémie du poulet	virus de l'anémie du poulet	culture sur lignée MDCC-MSBI ⇒ Recherche par immunomarquage
	recherche du virus de l'entérite du canard (si MOA canard ou oie)	virus de l'entérite du canard	culture sur cellules PR ou SR hépatiques d'ECB ⇒ Recherche par immunomarquage
	recherche des parvovirus du canard et de l'oie (si MOA canard ou oie)	parvovirus	culture sur cellules PR ou SR fibroblaste d'ECB ⇒ Recherche par immunomarquage

Vaccins viraux inactivés aviaires : recherche des agents étrangers dans les lots de produit final

RECHERCHE INDIRECTE SPECIFIQUE	recherche des agents étrangers sur poussins	AGP précipitation sur gelose EIA (dosage immunoenz ex ELISA) HI inhibition de l'hémagglutination Aggl agglutination SN séroneutralisation IS immunomarquage (ex Ac fluo)	inoculation 10 poussins SPF de 2 semaines, suivi 5s ⇒ Abs de signes cliniques ⇒ Abs d'anticorps contre les agents suivants :				
		SN	EIA	AGP	HI	Aggl	IS
toutes espèces aviaires	virus de l'encéphalomyélite aviaire		x	x			
	virus de la bronchite infectieuse aviaire		x		x		
	virus de la laryngotrachéite infectieuse aviaire	x	x				x
	virus de la leucose aviaire	x	x				
	virus de la maladie des œufs hardés (adénovirus type 3)						
	virus de la bursite infectieuse aviaire (type) (Gumboro)	x(I,II)	x(I)	x(I)			
	virus A de la grippe		x	x	x		
	virus de la maladie de Marek			x			
	virus de la pseudopeste aviaire (NC)		x		x		

Résumé

La détection des agents pathogènes est un enjeu croissant tant les maladies infectieuses représentent un risque pour la santé animale et humaine. La globalisation des échanges commerciaux et des voyages, l'évolution des pratiques agricoles ou les changements climatiques sont autant de facteurs bouleversant la biologie des micro-organismes et leurs capacités d'émergence. Ce manuscrit décrit trois approches basées sur trois techniques innovantes de biologie moléculaire pour la détection d'agents pathogènes et appliquées à trois contextes différents : (i) la recherche d'une liste précise de micro-organismes par PCR quantitative en temps réel en format microfluidique, (ii) la détection sans a priori d'agents infectieux par métagénomique et séquençage Illumina (Miseq) et (iii) le géotypage d'un agent infectieux sans amplification préalable des génomes par nouvelle génération de séquençage de troisième génération, le MinION d'Oxford Nanopore Technologies. Ces trois études ont permis de montrer l'apport de ces techniques, qui présentent toutes des caractéristiques distinctes, adaptées à différentes applications. Au-delà de l'application de ces techniques au diagnostic microbiologique, leur utilisation dédiée au contrôle de vaccins vétérinaires est une perspective prioritaire de ce travail. En effet, les préparations vaccinales sont soumises à l'obligation de recherche d'une liste d'agents pathogènes à exclure mais aussi à la vérification de l'identité génétique des souches vaccinales. Les performances exponentielles des nouvelles technologies de PCR et de séquençage ouvrent ainsi des perspectives révolutionnaires dans le domaine du diagnostic et du contrôle microbiologique.

Mots Clés

Séquençage de nouvelle génération, PCR microfluidique, diagnostic, métagénomique, virus respiratoires aviaires, bioinformatique.

Abstract

Detection of pathogens becomes an increasing challenge, since infectious diseases represent major risks for both human and animal health. Globalization of trade and travels, evolution of farming practices and global climatic changes are impacting the biology of pathogens and their emerging potential. This manuscript describes three approaches, based on three innovative technologies of molecular biology applied to the detection of pathogens in three different settings: (i) detection of a list of pathogens using real-time quantitative PCR on a microfluidic platform, (ii) unbiased detection of pathogens in complex matrix, using metagenomics and Illumina (Miseq) sequencing and (iii) genotyping of pathogens without pathogen isolation or DNA enrichment using a 3rd generation NGS platform MinION from Oxford Nanopore Technologies. The 3 studies shown the contribution of these techniques, each representing distinctive features, suitable for the respective applications. Beyond application of these techniques to the field of microbial diagnostics, their use for the control of veterinary immunological drugs is a priority of this project. Veterinary vaccines are not only submitted to mandatory detection of listed pathogens to be excluded, but also to validation of the genetic identity of vaccine strains. The exponential availability and performances of new PCR or sequencing technologies open cutting-edge perspectives in the field of microbial diagnostic and control.

Keywords

Next generation sequencing, microfluidic PCR, diagnostics, metagenomics, avian respiratory viruses, bioinformatics.