

OUTILS DE GÉNOMIQUE POUR LA SURVEILLANCE « EN TEMPS RÉEL » DES VIROSES AVIAIRES ÉMERGENTES ET RÉÉMERGENTES

Guillaume Croville¹, Mathilda Walch¹, Debapriyo Chakraborty¹, Adam Jbenyeni¹,
Laetitia Lèbre¹, Sonia Silva¹, Fabien Filaire¹, Timothée Vergne¹ & Jean-Luc Guérin¹

¹Université de Toulouse, ENVT, INRAE, UMR IHAP, 23 chemin des capelles, 31076
Toulouse, France
guillaume.croville@envt.fr

RÉSUMÉ

Afin d'ouvrir la voie du diagnostic de proximité depuis la prise en charge de l'échantillon jusqu'à la caractérisation génétique des agents pathogènes impliqués et leur analyse phylogénétique, des protocoles de métagénomique permettant le séquençage de génomes entiers d'avipoxvirus ou d'aviadénovirus ont déjà été développés. Ceux-ci faisaient appel à l'utilisation du séquençage de troisième génération Oxford Nanopore. Dans le but d'adapter ces méthodes à d'autres virus aviaires (coronavirus aviaires, virus de l'influenza aviaire), nous avons développé un protocole combinant la PCR multiplex et le séquençage Oxford Nanopore, à l'aide du séquenceur Mk1C, pour une détection couplée à un typage rapide des virus impliqués dans un épisode clinique. La méthode proposée est basée sur une approche intégrée « du terrain à la bioinformatique ». Une variété d'échantillons biologiques, y compris des écouvillons trachéaux, des tissus, ou des poussières, a été testée, ce qui a permis de proposer une première version de cette méthode innovante de surveillance. Grâce au séquençage nanopore des produits de PCR ciblant le génome des virus influenza aviaires, des génomes complets ont été obtenus en moins de 48 heures et sans avoir à réaliser de culture virale. Aussi, des génomes complets de coronavirus de la bronchite infectieuse aviaire issus d'écouvillons prélevés lors d'épisodes cliniques ont été obtenus par métagénomique. L'analyse bioinformatique de ce travail a été réalisée avec les outils Minimap2 pour les alignements, iVar pour la production de séquences consensus et NanoSPC pour la métagénomique. Les résultats obtenus suggèrent des temps d'analyses très courts pour l'obtention de génomes complets directement à partir d'échantillons primaires de terrain. L'application de telles méthodes pour la surveillance de viroses aviaires pourrait permettre un suivi en « temps réel » des épizooties.

ABSTRACT

Genomics tools for "real-time" monitoring of emerging and re-emerging avian viruses

In order to develop diagnostic methods from the handling of the sample to the genetic characterization of the pathogens involved and their phylogenetic analysis, we previously developed metagenomics protocols allowing the whole genome sequencing of avipoxvirus or aviadenovirus through the use of Oxford Nanopore third generation sequencing technology. In order to adapt these methods to other avian viruses (avian coronaviruses, avian influenza viruses), we have developed a protocol combining multiplex PCR and Oxford Nanopore sequencing, with the Mk1C sequencing device, for the pathogen detection coupled with a rapid typing. The proposed method is based on an integrated approach "from the field to bioinformatics", allowing rapid detection and typing of avian viruses. A variety of clinical samples, including tracheal swabs, tissues, or dust, has been tested and allowed us to introduce a first version of this innovative method of surveillance. Thanks to nanopore sequencing of PCR products targeting the avian influenza viruses' genomes, the whole genomes were sequenced in less than 48 hours without viral culture. Complete genomes of avian infectious bronchitis coronaviruses from swabs collected during clinical outbreaks were also obtained using metagenomics. The bioinformatics analysis of this work was carried out with Minimap2 for alignments, iVar for the production of consensus sequences and NanoSPC for metagenomics. The results obtained in this study suggest very short analysis times to obtain complete genomes directly from field samples. The application of such methods for the surveillance of avian viruses could allow "real time" monitoring of epizootics.

INTRODUCTION

Les élevages de volailles sont en permanence confrontés à des émergences virales pouvant être à l'origine d'infections systémiques et de sévères affections respiratoires. Les épizooties d'influenza aviaire hautement pathogène survenues ces dernières années ou encore les émergences de divers coronavirus aviaires soulignent la nécessité de disposer d'outils de séquençage et d'analyse bio-informatique fiables, rapides et proches du terrain.

La détection précoce des maladies infectieuses est une clef de la biosécurité, car elle permet d'adapter les mesures de contrôle avant la diffusion au sein d'une unité (élevage) ou à l'échelle d'un territoire. L'application rapide des mesures limite ainsi à la fois l'expansion et la dérive génétique des agents infectieux.

Le diagnostic virologique fait classiquement appel aux techniques directes et indirectes, et en particulier la PCR et le séquençage des acides nucléiques permettant la détection, la détermination de la charge et le typage (plus ou moins précis), voire l'étude de chaînes de transmissions pour certaines applications.

Le séquençage de troisième génération Oxford Nanopore Technologies (ONT) a été largement déployé ces dernières années grâce à la portabilité des séquenceurs MK1b et MK1c et aux volumes de données qu'ils génèrent (le débit théorique d'une *flow cell* MinION étant de plusieurs dizaines de gigabases) qui sont adaptés à des projets de diagnostic rapide.

Cependant, la sensibilité globale des approches de séquençage direct ONT sans a priori dépend grandement de l'efficacité de la purification des acides nucléiques (AN) et des abondances absolue et relative des gènes cibles dans l'échantillon collecté. En effet, les agents pathogènes responsables d'infections systémiques peuvent exister à très faible charge dans les liquides biologiques, sous forme de microorganismes intègres ou sous forme d'AN libres. Par conséquent, un échantillon à faible concentration relative en agents pathogènes (par rapport aux AN totaux) doit être analysé par séquençage à très haut débit (ou séquençage de nouvelle génération, NGS) pour obtenir une quantité suffisante de données.

Le traitement des échantillons et la préparation des banques d'AN à séquencer sont les verrous majeurs pour gagner en séquences d'intérêt dans les jeux de données. Cette étude, centrée sur l'analyse des virus à ARN, ceux de l'influenza aviaire (VIA) et de la bronchite infectieuse aviaire (BI), a porté sur le développement et l'application d'outils de biologie moléculaire et de bio-informatique.

1. MATERIELS ET METHODES

1.1. Extraction des ARN viraux

Les échantillons ont été traités dans le strict respect de la procédure de biosécurité (échantillons infectés par

des virus influenza aviaires hautement pathogènes traités en laboratoire BSL3). Les écouvillons trachéaux, les fragments d'organes et les chiffonnettes (pour la récupération des poussières présentes sur les murs, mangeoires et abreuvoirs) ont été placés dans du PBS 1X avant extraction de l'ARN viral à l'aide du kit ID Gene® Mag Fast Extraction Kit. L'ARN viral a ensuite été stocké à -80°C.

1.2. Amplification des acides nucléiques

1.2.1. Démarche SISPA

Les ARN viraux de BI ont été aléatoirement rétrotranscrits selon la méthode SISPA (*Sequence-Independent, Single-Primer Amplification*) qui consiste à générer des ADN complémentaires double-brin. Le protocole appliqué a été adapté de Greninger et al., 2015.

1.2.2. Démarche PCR multiplex (mPCR)

Les ARN viraux influenza ont été préalablement détectés à l'aide du kit ID Gene Influenza A (IDvet) puis amplifiés selon une méthode combinant deux approches : (i) l'amplification des régions encadrant le site de clivage de l'hémagglutinine (HA) à l'aide d'un jeu de 20 couples d'amorces de PCR et (ii) l'amplification du génome complet à l'aide d'amorces universelles spécifiques des extrémités 3' et 5' des ARN viraux (MBT-uni12 [5' - ACGCGTGATCAGCAAAAGCAGG] / MBT-uni13 [5' - ACGCGTGATCAGTAGAAACAAGG], (Zhou et al., 2012)).

Les ARNv ont été rétrotranscrits à l'aide du kit RevertAid (ThermoFisher) et 1,5 µl d'amorce Uni-12 à 10 µM, puis amplifiés en présence des deux jeux d'amorces en utilisant l'ADN polymérase Phusion™ High-Fidelity (ThermoFisher Scientific), en suivant les instructions des fournisseurs.

Le programme de PCR a été réalisé comme suit : 3mn à 94°C, 5 cycles [45s à 94°C, 45s à 45°C et 2mn à 72°C], 30 cycles [45s à 94°C, 45s à 55°C et 2mn 30s à 72°C], 10mn à 72°C. Les produits de PCR ont été contrôlés sur gel d'agarose et quantifiés au Qubit (ThermoFisher Scientific).

1.3 Préparation des banques d'ADN à séquencer

Les banques d'ADN ont été préparées à l'aide d'un kit de séquençage SQK-LSK109 fourni par ONT associé au kit de multiplexage EXPNBD104 (ONT). Les banques ont été déposées sur *flow cell* MinION ou *flongle* selon la quantité d'échantillons à analyser. Les *runs* de séquençage ont été réalisés sur MinION Mk1C (ONT) et le *basecalling* (conversion du signal brut en séquence nucléotidique) a été effectué en temps réel avec Guppy (v3.5).

1.4. Analyse bio-informatique

Les alignements des séquences obtenues au format Fastq ont été réalisés avec l'outil Minimap2 (Li, 2018) ainsi que la suite Samtools (Li et al., 2009) et la visualisation graphique des alignements avec IGV (Robinson et al., 2011).

Les séquences consensus ont été obtenues avec iVar (Grubaugh et al., 2019) et inspectées en utilisant Bioedit (v7.2.6).

L'analyse métagénomique sans *a priori* a été réalisée à l'aide de NanoSPC (Xu et al., 2020) mais également par un alignement Minimap2 sur une base de données affinée, constituée de séquences de référence pertinentes, et en analysant les résultats d'alignements avec l'option idxstats de Samtools.

2. RESULTATS

2.1 Influenza aviaire

La méthodologie de mPCR décrite ci-dessus a été appliquée à une série de 11 échantillons (écouvillons trachéaux ou chiffonnettes) prélevés dans le sud-ouest et l'ouest de la France durant l'épizootie 2020-2021 d'influenza aviaire, les élevages prélevés ayant été au préalable déclarés comme foyers d'influenza aviaire hautement pathogène (IAHP) par les autorités sanitaires. Les génomes complets des 11 échantillons ont été obtenus après 4 heures de séquençage ONT, avec une profondeur de lecture du site de clivage de la HA supérieure à 1000X, soit plus de 1000 séquences couvrant cette région du génome. Les données de séquences dédiées à la HA s'étendaient de 0.36 Mb à 45.34 Mb (Tableau 1).

L'analyse des séquences de HA a permis de typer les prélèvements comme appartenant au sous-type H5N8 et de déterminer leur profil hautement pathogène sur la base du profil génétique du site clivage de la HA, c'est-à-dire une suite de résidus d'acides aminés polybasiques. L'analyse phylogénétique des séquences de la HA a quant à elle permis d'identifier trois génotypes distincts corrélés aux zones géographiques de prélèvement ou bien à la temporalité du prélèvement, c'est-à-dire une discrimination des échantillons sur la base de la séquence complète de chaque HA entre : (i) six prélèvements des Landes de décembre 2020 (groupe A, figure 1), (ii) deux prélèvements de Vendée et Deux-Sèvres de décembre 2020 (groupe C, figure 1) et (iii) quatre prélèvements de Hautes-Pyrénées et du Gers effectués en janvier et février 2021 (groupe B, Figure 1).

2.2 Bronchite infectieuse aviaire

Un panel de 10 prélèvements sous forme d'écouvillons trachéaux ou de carte FTA ont été soumis à une analyse de type métagénomique grâce à laquelle des génomes complets ou partiels ont été générés. Ces résultats ont permis d'effectuer un typage rapide des virus de BI circulants afin de

déterminer s'il s'agissait de la souche vaccinale ou d'une souche sauvage, et pour certains cas identifier des événements de recombinaison génétique (données non montrées).

3. DISCUSSION ET CONCLUSION

Le couplage de la métagénomique et de la PCR multiplex présenté dans ce travail est justifié par le fait qu'il est techniquement plus complexe d'obtenir des séquences virales sans réaliser d'enrichissement en séquences d'intérêt. Aussi, concernant les VIA, il est primordial de se focaliser sur le site de clivage de la HA afin de pouvoir obtenir à minima l'information de type et de pathotype, même dans le cas d'un échantillon pauvre en matériel viral ou dans le cas où celui-ci serait dégradé.

Malgré un taux d'erreur proche de 10% en séquençage nanopore, la profondeur de lecture supérieure à 10X était suffisante pour obtenir des séquences consensus de qualité située entre 99% et 100% (Xu et al., 2021).

Le processus complet depuis l'extraction des acides nucléiques jusqu'à l'analyse des séquences est réalisable en moins de 48 heures. Cette méthodologie ayant permis de grouper les échantillons d'influenza sur la base de la variabilité génétique du gène HA, nous la considérons comme un moyen efficace de surveillance de la circulation des virus influenza aviaires. Des analyses de phylodynamique sont à prévoir pour établir des liens entre les différentes introductions de ce virus et définir les chaînes de transmission.

La technique de mPCR reste incontournable pour générer dans des délais très courts et de façon spécifique des génomes complets à partir d'échantillons cliniques, cette méthode a été privilégiée pour l'amplification des VIA afin de pouvoir séquencer des génomes présents en faibles quantités dans les échantillons (exemple : CT gène M = 28,5 ; cf. Tableau 1). La métagénomique a été appliquée aux prélèvements de bronchite infectieuse car les charges virales étaient élevées, ce qui ne nécessitait pas d'amplification préalable (données non montrées). C'est une méthode qui, bien qu'elle soit complexe à mettre en œuvre, permet le cas échéant d'informer sur les coinfections que ce soit dans le cadre de surveillance virale ou d'analyse syndromique.

Enfin, les outils de bioinformatique utilisés dans ce travail ont permis de générer des alignements et des séquences consensus de génomes complets en quelques minutes, permettant par exemple de typer les souches virales détectées ou de réaliser des analyses phylogénétiques.

REFERENCES BIBLIOGRAPHIQUES

Greninger A, Naccache S, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen J, Dodd R, Mulembakani P, Schneider B, Muyembe-Tamfum JJ, Stramer S, Chiu C, 2015. In: *Genome Med* 7, 99.

Grubaugh N, Gangavarapu K, Quick J, Matteson N, Goes De Jesus J, Main B, Tan A, Paul L, Brackney D, Grewal S, Gurfield N, Van Rompay K, Isern S, Michael S, Coffey L, Loman N, Andersen K, 2019. In: *Genome Biol* 20, 8.

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. In: *Bioinformatics* 25(16) 2078-9.

Li H, 2018. In: *Bioinformatics*, Volume 34, Issue 18, Pages 3094–3100.

Robinson J, Thorvaldsdóttir H, Winckler W, Guttman M, Lander E, Getz G, Mesirov J, 2011. In: *Nature Biotechnology* 29, 24–26.

Xu Y, Yang-Turner F, Volk D, Crook D, 2020. In: *Nucleic Acids Research*, Volume 48, Pages W366–W371.

Xu Y, Lewandowski K, O Downs L, Kavanagh J, Hender J, Lumley S, Jeffery K, Foster D, Sanderson N, Vaughan A, Morgan M, Vipond R, Carroll M, Peto T, Crook D, Walker S, Matthews P, Pullan S, 2021. In: *Eurosurveillance* 26.

Zhou B, Wentworth DE, 2012. In: *Methods Mol Biol.* 2012;865:175-92.

Tableau 1. Récapitulatif des volumes de données de séquençage ONT générés pour les onze échantillons positifs en virus influenza aviaire. La valeur x de N50 indique que 50% des bases générées lors du séquençage appartiennent à des séquences dont la taille est supérieure à x .

Référence	Type d'échantillon	Date prélèvement (AAAA-MM)	Département	Ct RT-qPCR (gène M)	Volume de données de séquençage (Mb)	N50 (pb)	HA yield (Mb)	N50 HA (pb)
LAB20323	Ecouvillon trachéal	2020-12	40	28,5	230,59	640	30,53	661
LAB20335	Ecouvillon trachéal	2020-12	40	24,7	139,52	891	6,05	636
LAB20338	Ecouvillon trachéal	2020-12	85	27,1	96,09	572	0,36	469
LAB20339	Chiffonnette	2020-12	79	20,4	129,52	882	9,10	641
LAB20347	Ecouvillon trachéal	2020-12	40	14	474,13	937	45,34	1 818
LAB20349	Ecouvillon trachéal	2020-121	40	20,4	2,34	896	0,17	656
LAB20352	Ecouvillon trachéal	2020-12	40	18,9	289,83	808	22,99	648
LAB20353	Pulpe de plume	2020-12	40	23,5	301,95	897	28,36	598
LAB21061	Chiffonnette	2021-01	65	25	28,62	443	1,17	681
LAB21064	Ecouvillon trachéal	2021-01	65	20	28,35	872	7,65	652
LAB21084	Ecouvillon trachéal	2021-02	32	22,8	366,39	888	32,68	673

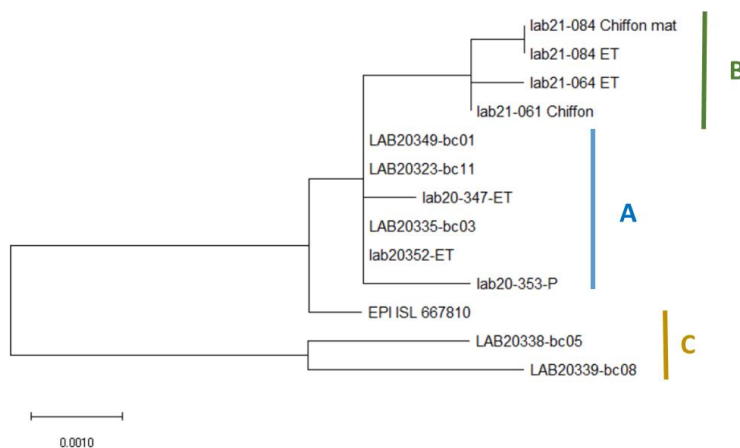


Figure 1. Analyse phylogénétique, par la méthode de Neighbor-Joining, des séquences du gène de l'hémagglutinine des souches de l'épizootie 2020-2021 séquencées sur Minion Mk1C. Les trois groupes phylogénétiques A, B et C regroupent différents génotypes ayant circulé dans le sud-ouest de la France de décembre 2020 à février 2021.