amU
Aix Marseille Université

NNT : 0000AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
Le 24 janvier 2025 par

Lorlane LE TARGA

Suivi de l'évolution du SARS-CoV-2 et des virus respiratoires responsables d'épidémies, dans les eaux usées et chez les patients, par qPCR et séquençage génomique

Discipline
Biologie santé

Spécialité
Bio-informatique et génomique

École doctorale
ED 62-Sciences De La Vie Et De La Santé

Laboratoire/Partenaires de recherche
MEPHI/IHU Méditerranée Infection
Biosellal

Composition du jury

Ilka ENGELMANN Centre Hospitalier Universitaire de Montpelier	Rapporteuse
Patrice MORAND Centre Hospitalier Universitaire de Grenoble	Rapporteur
Florence FENOLLAR Université d'Aix-Marseille	Présidente du jury
Bernard LA SCOLA Université d'Aix-Marseille	Directeur de thèse
Philippe COLSON Université d'Aix-Marseille	Membre invité
Jérôme VENTOSA Biosellal	Membre invité

MEPHI Microbes Evolution Phylogeny and Infections
MÉDITERRANÉE INFECTION
biosellal

Affidavit

Je soussigné, LE TARGA Lorlane, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Pr Bernard LA SCOLA, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille, le 8 novembre 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et/ou brevets et participation aux conférences

1) Liste des publications et/ou brevet réalisées dans le cadre du projet de thèse :

    1. From detection to sequencing of respiratory viruses in wastewater, for optimal monitoring of epidemic
       Lorlane Le Targa, Nathalie Wurst, Philippe Colson, Bernard La Scola, Sarah Aherfi (No published yet)

    2. SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe
       Lorlane Le Targa, Nathalie Wurtz, Alexandre Lacoste, Gwilherm Penant, Priscilla Jardot, Alexandre Annessi, Philippe Colson, Bernard La Scola, Sarah Aherfi (Puplished in Viruses)

    3. Monitoring of Influenza A/H3N2 reassortment in the southern France between 2017-2022
       Lorlane Le Targa, Houmadi Hikmat, Celine Boschi, Laurence Thomas, Raphael Tola, Bernard La Scola, Philippe Colson (No published yet)

    4. Monitoring of Influenza A/H1N1 recombination and reassortment in the southern France between 2017-2022
       Safietou Dialo, **Lorlane Le Targa**, Houmadi Hikmat, Bernard La Scola, Philippe Colson (No published yet)

    5. Matrix-Encoding Gene Diversity of 624 Influenza A/H3N2 Genomes Does Not Show Association with Impaired Viral Detection by Commercialized qPCR Assays
       Lorlane Le Targa, Houmadi Hikmat, Céline Boschi, Bernard La Scola, Philippe Colson (Published in Viruses)

    6. Influenza at the 2021 Grand Magal of Touba and possible spread to rural villages in South Senegal - a genomic epidemiological study
       Ndiaw Goumballa, Fatou Samba Diouf, Mamadou Beye, Masse Sambou, Hubert Bassène, Mamadou Dieng, Adama Aïdara, **Lorlane Le Targa,** Philippe Colson, Philippe Gautret, Cheikh Sokhna (Puplished in International Journal of Infectious Diseases)

7. Sequencing and characterization of human bocavirus genomes from patients diagnosed in Southern France between 2017 and 2022
   Houmadi Hikmat, **Lorlane Le Targa,** Celine Boschi, Justine Py, Marielle Bedotto, Aurélie Morand, Nadim Cassir, Sarah Aherfi, Bernard La Scola, Philippe Colson (Puplished in Journal of Medical Virology)

8. Five-year (2017-2022) evolutionary dynamics of human coronavirus HKU1 in southern France with emergence of viruses harboring spike H512R substitution
   Houmadi Hikmat, **Lorlane Le Targa,** Céline BOSCHI, Justine PY, Aurélie Morand, Jean-Christophe Lagier, Sarah Aherfi, Jacques Fantini, Bernard La Scola, Philippe Colson (Submitted in Journal of Medical Virology)

9. Full-length title: Five-year (2017-2022) evolutionary dynamics of human coronavirus OC43 in southern France based on whole genome next-generation sequencing
   Houmadi Hikmat, **Lorlane Le Targa,** Celine Boschi, Justine Py, Sarah Aherfi, Bernard La Scola, Philippe Colson (No published yet)

10. Characterization of human parainfluenza virus type 2 (HPIV2) circulating in southern France between 2017 and 2022 using whole genome sequencing
    Enzo Bisi, Céline Boschi, Houmadi Hikmat, Jérémy Delerce, **Lorlane Le Targa**, Lucile Lesage, Aurélie Morand, Bernard La Scola , Philippe Colson (Submitted in Journal of Virologica Sinica)

2) Participation aux conférences et écoles d'été au cours de la période de thèse :
   1. Poster: Monitoring the evolution of Influenza A/H3N2 in 843 patients samplefrom 2017 to 2022 to Marseille by whole genomes equencing
      Lorlane Le Targa, Hikmat Houmadi, Philipe Colson, Bernard La Scola
      Journée Infectiopôle Sud, 2023 - IHU Méditerranée Infection, Marseille France

   2. Poster: Monitoring respiratory viruses in wastewater in a one health purpose
      Lorlane Le Targa, Valentin Debelhoir, Amandine Berthet, Floriane Pez and Eric Sellal
      ISWALD, 2023- Lyon, France

Résumé

Les virus respiratoires sont parmi les principaux agents responsables de maladies humaines, ils sont une cause majeure de morbidité et de mortalité, notamment chez les populations vulnérables. La pandémie de SARS-CoV-2 a révélé les limites des systèmes de surveillance actuels, notamment en ce qui concerne le suivi génomique et la circulation de ces virus. La détection des virus respiratoires uniquement chez les personnes symptomatiques ou cas contacts, s'est avérée insuffisante pour suivre efficacement les épidémies.

Dans ce contexte, ce travail s'est intéressé à l'utilisation des eaux usées comme un moyen de surveillance des épidémies de virus respiratoires. Nous avons pu voir grâce à une revue bibliographique que les eaux usées, qu'elles proviennent des réseaux urbains ou des transports longues distance, permettent de savoir quels virus circulent, leurs concentrations et peuvent permettre d'obtenir les séquences de ces virus.

Nous avons aussi pu montrer, par la détection de SARS-CoV-2 dans les eaux usées de deux avions en provenance d'Addis-Abeba et ayant atterri à Marseille en décembre 2021, qu'il était possible de détecter le SARS-CoV-2 par qPCR et de réaliser un séquençage avec la méthode de pré-amplification ARTIC. De plus cet article a montré que les mesures sanitaires mises en place pour les voyageurs n'empêche pas la circulation du virus du SARS-CoV-2 entre les pays.

En parallèle, une attention particulière a été accordée à la surveillance génomique des virus respiratoires chez les patients, en particulier le suivi du réassortiment des virus de la grippe A/H3N2 et A/H1N1 chez les patients de 2017 à 2022. Ce suivi génomique, a pu être réalisé à l'aide du séquençage de nouvelle génération et d'un système d'amplification du génome complet de la grippe A. Cette surveillance génomique a de plus permis de répondre directement à une alerte sur l'apparition de potentiels cas de grippe non détectés par des systèmes de détection commercialisés.

Cette thèse montre l'importance de la surveillance des virus respiratoires à grande échelle à l'aide de nouvelles matrices de surveillance et du séquençage à grande échelle.

Mots clés : SARS-CoV-2, Virus Influenza, Surveillance, Virus respiratoire

Abstract

Respiratory viruses are among the main agents responsible for human diseases; they are a major cause of morbidity and mortality, particularly among vulnerable populations. The SARS-CoV-2 pandemic has highlighted the limitations of current surveillance systems, especially concerning genomic tracking and monitoring of these viruses. Detecting respiratory viruses only in symptomatic individuals or contact cases has proven insufficient for effectively tracking epidemics.

In this context, this work focuses on the use of wastewater as a means of monitoring respiratory virus epidemics. A literature review showed that wastewater, whether from urban networks or long-distance transportation, allows us to determine which viruses are circulating, their concentrations, and to obtain viral sequences.

We also demonstrated through the detection of SARS-CoV-2 in wastewater from two planes arriving from Addis Ababa and landing in Marseille in December 2021 that it is possible to detect SARS-CoV-2 by qPCR and conduct sequencing using the ARTIC pre-amplification method. Additionally, this study showed that the health measures in place for travelers do not prevent the circulation of SARS-CoV-2 between countries.

Simultaneously, special attention was given to the genomic surveillance of respiratory viruses in patients, particularly the tracking of reassortment of influenza A/H3N2 and A/H1N1 viruses in patients from 2017 to 2022. This genomic tracking was achieved using next-generation sequencing and a system for amplifying the full genome of influenza A. Moreover, this genomic surveillance enabled a direct response to an alert concerning potential cases of influenza not detected by commercial detection systems.

This thesis demonstrates the importance of large-scale respiratory virus surveillance using new monitoring matrices and large-scale sequencing.

Keywords: SARS-CoV-2, Influenza virus, Monitoring, Respiratory virus

Remerciements

Je tiens tout d'abord à remercier Pr. Florence Fenollar d'avoir acceptés de présider le jury de cette thèse.

Je remercie ensuite les membres du jury Pr. Ilka Englelmann et Pr Patrice Morand d'avoir accepter d'être mes rapporteurs.

Je tiens à remercier mon directeur de thèse, le Professeur Bernard La Scola. Merci de m'avoir accueillie dans votre équipe. Vous m'avez fait confiance dès le début de ma thèse, même sans m'avoir eue en stage ou en tant qu'étudiante Merci de m'avoir guidée dans mes recherches. Vous vous décrivez vous-même comme peu réceptif à la bio-informatique, mais cela ne vous a jamais empêché de me donner de bons conseils.

Je remercie également tout particulièrement le Professeur Philippe Colson. Merci de m'avoir aidée durant ces trois années. Avec le Professeur La Scola, vous avez formé une équipe formidable pour m'encadrer. Merci pour vos précieux conseils.

Je ne peux que remercier Eric Sellal, directeur de Biosellal, de m'avoir permis de réaliser cette thèse en tant qu'employée de Biosellal. Merci de m'avoir donné cette opportunité.

Un grand merci à Valentin Debelhoir d'avoir été mon référent à Biosellal. J'ai été ta première N-1 en innovation, tu m'as beaucoup appris et tu as toujours eu de bonnes idées lors de nos échanges. Travailler avec toi a toujours été un plaisir. Merci également de m'avoir accompagnée à mon premier congrès (même si je n'ai pas vraiment laissé, ni à toi ni à Jérôme, le choix d'être suivis). Merci de m'avoir aidée à trouver ma « petite » place dans l'équipe innovation.

Merci à tous mes collègues de Biosellal. C'était toujours un plaisir de venir vous voir à Dardilly. Merci à Jérôme de prendre la suite de Valentin, je suis contente de t'avoir comme futur N+1. Nathan et Julien, j'ai été heureuse de partager votre bureau avant le début de ma thèse. Merci à tous de m'avoir toujours accueillie avec le sourire, d'voir toujours su me trouver une place.

Je tiens à remercier tout particulièrement Pierre Pontarotti. Pendant le premier confinement, j'ai tout de même pu réaliser mon stage de M1 grâce à vous. Encore merci de m'avoir accueillie pendant mon stage de M2, vous avez été le premier à m'ouvrir les portes de l'IHU. Vous m'avez brillamment guidée et vous faites partie de ceux qui m'ont donné le courage et l'envie de réaliser cette thèse. Si j'en suis là aujourd'hui, c'est en partie grâce à vous.

Merci à Sarah Aherfi pour ton temps, ta gentillesse et tes corrections. J'ai pris plaisir à écrire avec toi mon Papier sur le SARS-COV-2 et ma revue. Tu ne peu pas savoir à quels points cela m'a fait du bien quand tu m'as dit que ma revue était très bien juste avant le

dépôt de mon manuscrit, c'est en partie grâce à toi. Merci aussi à Nathalie Wurst pour vos corrections pour ma revue.

Je remercie tout particulièrement Hikmat, avec qui je travaille depuis quatre ans. Depuis le M2, nous avons partagé la même pièce, puis la même équipe dès le début de la thèse. C'est toujours un plaisir de travailler avec toi. Je suis heureuse que nous ayons pu nous entraider dans nos recherches. Merci d'avoir toujours eu la patience de corriger mon orthographe et ma grammaire. Merci d'être la personne que tu es, je n'aurais pas pu rêver de meilleure collègue de thèse. D'accord même sur nos désaccords, Collègue à l'IHU, partenaire à la salle de sport, et même rapidement coloc, tout l'IHU sait que nous sommes rarement l'une sans l'autre dans ces couloirs. Merci.

Merci aux autres étudiants actuels de l'équipe : Margaux, Yousra, Safietou, Pulchérie et Anthony. Travailler avec vous est toujours un plaisir (Même Margaux c'est dire !), même si je pense vous faire parfois un peu peur avec ma bioinfo. Bientôt, ce sera votre tour et vous allez tous gérer ! Merci aussi à Céline d'être là depuis le début et de nous soutenir.

Merci à toutes les personnes que je n'ai pas encore citées et qui ont fait partie de la famille de la salle 6 pendant ces quatre ans à l'IHU : Justine, Kadidja, Ornella La Fortune Tchoupou Saha, Rita, Matteo, Mélanie, Marine, Nacera, Léa, Azélie, Christina. Vous avez tous contribué à rendre cette salle (pour ne pas dire cave) un peu plus chaleureuse. Merci à vous pour tous ces moments.

Merci à tous mes amis et collègues de l'IHU de m'avoir accompagnée dans cette aventure, en particulier : Wafaa, Mahmoud, Aïcha, Michelande, Rayan, Fatou et à tous ceux que j'ai pu oublier.

Je remercie aussi l'équipe des ingénieurs : Jeremy, Marielle, Sylvain, Mamadou. Merci de votre aide et de m'avoir partagé vos connaissances.

Merci aux membres du bureau de DynAMU, depuis presque deux ans j'adore rencontrer grâce à vous d'autres doctorants d'Aix-Marseille. Merci de m'avoir permis de faire partie (ou plutôt d'avoir insisté pour que je fasse partie) du bureau. C'est toujours un plaisir d'organiser des événements avec vous. Merci à Clémentine, Célia, PJB, Mouad, Lauralie, Amaury, Valentin et Jonathan. Merci aussi aux adhérents qui viennent toujours avec le sourire partager de bons moments avec nous.

Même si cela fait un peu doublon avec les remerciements à DynAMU, merci au groupe des Rhum Grimpette, avec Célia, Lauralie, Amaury, Julien et Elie. J'ai passé un super été 2024 avec vous. J'aime le fait que nous n'ayons pas besoin d'autre raison qu'un verre et un jeu de cartes pour sortir, même si parfois nous faisons quand même un peu d'escalade et que vous voulez me forcer à faire un semi-marathon !

Merci à mes amies, Siham et Manon, vous faite parti de mon trio de choc et ce n'est pas pour rien ! Partir en vacances avec vous est toujours un plaisir. Merci de m'avoir soutenue, d'avoir écouté mes plaintes sur WhatsApp, mes joies aussi. Vous êtes là depuis le BTS et vous me poussez toujours vers le haut, mes souvenirs avec vous sont précieux.

Merci à ma binôme de TP préférée, Marie. Après les TP de génomique et la quasi-nuit blanche avant le rendu, plus rien ne nous paraît insurmontable. Tu es une personne formidable, on ne se voit pas souvent, mais tu es toujours dans mon cœur. Je sais que je pourrais toujours compter sur toi pour tous.

Merci aussi à tous les collègues de licence.

Merci à Monsieur Mitre de m'avoir juré, les yeux dans les yeux, que je pourrais rejoindre la licence même en changeant de BTS, vous ne m'aviez pas menti. Vous avez été un professeur formidable en BTS et en licence.

Merci à tous les professeurs de BTS BioAC et Biotech qui ont été formidables avec moi : Madame Badou, Madame Grimal, Madame Latour, Madame Bissery, Monsieur Le Texier et tous ceux que j'ai pu oublier.

Merci à Nathalie Janel de m'avoir acceptée pour mon premier stage et de m'avoir donné une première vision de la recherche durant mon stage de BTS. Merci à Nadim Kassis pour les autres stages, pour ta bonne humeur et les blagues inter-labos, et merci de m'avoir encouragée à poursuivre après le BTS.

Merci à Valérie Sterne d'avoir été une tutrice géniale pendant mon alternance. La cristallographie n'était pas ma voie, mais tu m'as parfaitement guidée, merci de m'avoir encouragée. Merci évidemment à toute l'équipe de biologie structurale.

Un grand merci à ma famille de m'avoir soutenue tout au long de mes études. Merci à mes parents, mes frères, ma belle-sœur et ma nièce. Merci à mes oncles et tantes, ainsi qu'à mes cousins et cousines. Yoann et Loïc, je me souviens d'une conversation où j'hésitais à faire une thèse, et vous m'avez dit : « imagine si tu fais une thèse, tu pourras nous dire : moi au moins, je suis docteur ». Si certaines petites sœurs ont besoin d'encouragements infaillibles, pour moi, une raison de taquiner mes frères vaut tout l'or du monde. Merci.

Merci à toutes les personnes qui ont cru en moi plus que je ne l'ai fait moi-même. Sans vous, je n'aurais jamais eu le courage de faire un doctorat.

Merci aussi à toutes les personnes que j'ai pu oublier.

Table des matières

Affidavit **2**
Liste de publications et/ou brevets et participation aux conférences **3**
Résumé **5**
Abstract **6**
Remerciements **7**
Table des matières **10**
Liste des abréviations **11**
Introduction **12**
1. Généralités 12
1.1. Les virus respiratoires humains 12
1.2. Epidémies et pandémies de virus respiratoires 12
1.3. Surveillance des virus respiratoires chez les patients 13
2. Utilisation des eaux usées comme moyen de surveillance des virus respiratoires . 14
3. Suivi génomique de l'évolution des virus respiratoires dans les eaux usées et dans la population 14
3.1. Intérêt du suivi génomique des virus respiratoires 14
3.2. Les systèmes de PCR-multiplexés 15
Objectifs et structure de la thèse **16**
Chapitre I : De la détection au séquençage des virus respiratoires dans les eaux usées **17**
Avant-propos : 17
Chapitre II : Détection du SARS-CoV-2 dans les eaux usées des avions **49**
Avant-propos : 49
Chapitre III : Suivis du réassortiment de la grippe A/H3N2 et A/H1N1 chez les patients de 2017 à 2022 **60**
Avant-propos : 60
Chapitre IV : L'importance du séquençage dans la détection (ou non) de faux positif **113**
Avant-propos : 113
Conclusion **121**
Bibliographie **123**
Annexes **129**

Liste des abréviations

ADN : Acide désoxyribonucléique
ARN : Acide ribonucléique
Ct : Cycle threshold (Seuil de cycle)
ECDC : European Centre for Disease Prevention and Control (Centre européen de prévention et de contrôle des maladies)
HA : Hémagglutinine
HAdVs : Adénovirus humain
HBoV : Bocavirus humain
HCoV : Coronavirus humain
hMPV : metapneumovirus humain
HPIV : Parainfluenza virus humain
NA : Neuraminidase
NGS : Next Generation Sequencing (Séquençage de nouvelle génération)
PCR : Polymerase Chain Reaction (Réaction en chaîne par polymérase)
RhV : Rhinovirus
RSV : Virus Respiratoire Syncitiale
VOC : Variant Of Concerne (Variant préoccupant)
OMS : Organisation Mondiale de la Santé

Introduction

1. Généralités

1.1. Les virus respiratoires humains

Les virus respiratoires humains sont des virus qui se transmettent facilement et rapidement dans la population générale par le biais de microgouttelettes. Ces virus peuvent causer de nombreux symptômes, tels que la toux, les écoulements nasaux, la gorge irritée, la fièvre, etc. Ils peuvent également provoquer des infections aiguës des voies respiratoires supérieures et inférieures.

Il existe de nombreux virus respiratoires circulant chez l'homme. Ils appartiennent à plusieurs familles et genres différents. La plupart sont des virus à ARN Les virus influenzas A et B (FluA et FluB) de la famille des *Orthomyxoviridae* [1][2]Les coronavirus humains (HCoV) de la famille des *Coronaviridae*, séparés en deux genres : *alphacoronavirus* (incluant HCoV-NL63 et HCoV-229E) , et *betacoronavirus* (incluant SARS-CoV-1, SARS-CoV-2, MERS-CoV, HCoV-OC43 et HCoV-HKU1)[2], [3], [4]. Les Human Parainfluenza Virus (HPIV) de la famille des *Paramyxoviridae*, divisés en deux genres : *Respirovirus* (incluant HPIV-1 et HPIV-3) et *Rubulavirus* (incluant HPIV-2 et HPIV-4) [2], [5]. Les rhinovirus A, B et C (RhV) de la famille des *Picornaviridae* [5], [6]. Et enfin le virus respiratoire syncitial (RSV) et le Metapneumovirus (hMPV) tous de la famille des *Pneumoviridae*, appartenant respectivement aux genres *Orthopneumovirus* et *Metapneumovirus* [2].

On retrouve également des virus à ADN tels que : le Bocavirus humain (HBoV), appartenant à la famille des *Parvoviridae* et au genre *Bocaparvovirus* [7]. Et les adénovirus humains (HAdVs), qui font partie de la famille des *Adenoviridae* [8].

1.2. Epidémies et pandémies de virus respiratoires

Les virus respiratoires peuvent se rependre rapidement dans la population et causer des épidémies occasionnelles ou saisonnières ainsi que des pandémies mondiales.

Les épidémies de virus respiratoires ont un impact majeur sur la santé publique à travers le monde à la fois sur la santé de la population, mais aussi sur le prix qu'elles coûtent directement et indirectement. En 2015 aux Etats-Unis, il a été estimé que le virus de la grippe coûtait 11.2 milliards de dollars par an [9]. En 2003aux Etats-Unis, il a été estimé que les virus non-influenza coutaient 40 milliards de dollars par an [10].

En plus des épidémies saisonnières s'ajoutent les pandémies. La dernière pandémie en date est celle de la Covid-19. La Covid-19 est une maladie provoquée par le virus du SARS-CoV-2. Le SARS-CoV-2 est apparu pour la première fois en 2019 à Wuhan (Chine) [4], d'après les données de l'OMS [11] au 1er novembre 2022 et depuis le début de la pandémie de SARS-CoV-2, la COVID-19 a touché plus de 600 millions de personnes (plus de 250 millions en Europe) avec plus de 6,5 millions de morts dans le monde (plus de 2 millions en Europe).

Les virus Influenza restent les virus respiratoires ayant causé le plus de pandémies. Depuis le XXe siècle FluA a été responsable d'au moins 4 pandémies : la grippe « espagnole » (H1N1) en 1918, la grippe « asiatique » (H2N2) en 1957, la grippe de « Hong Kong » (H3N2) en 1968 [12] et la grippe « porcine » (H1N1pdm09) en 2009 [13].

De plus, il est possible que dans le futur, d'autres virus respiratoires connus, endémiques ou épidémiques, ou encore de nouveaux virus respiratoires non découverts provoquent un jour une nouvelle pandémie.

1.3. Surveillance des virus respiratoires chez les patients

Les épidémies et les pandémies de virus respiratoires, ont un impact important sur l'économie mondial et la santé humaine [9], [10]. C'est notamment pour cela qu'il est important de pouvoir réaliser la surveillance de ces virus à travers le monde. Il existe de nombreux programmes de surveillance des épidémies, soit au sein d'un pays, soit au sein de consortiums internationaux regroupant plusieurs pays et/ou continents. Ces programmes de surveillances peuvent être tournés vers un ou plusieurs virus [14], [15], [16].

Les stratégies de suivi épidémique actuelles se basent principalement sur la détection du virus chez les patients se rendant à l'hôpital ou dans des laboratoires d'analyses. Les organismes de surveillance récupèrent les données auprès des laboratoires et hôpitaux faisant partie de leurs réseaux.

Ces réseaux de surveillance permettent d'avoir une vision globale de la prévalence des virus respiratoires chez les patients dans différents pays. Mais la majorité des patients testés sont en général des patients présentant des symptômes respiratoires ou des cas contact. Cela limite notamment la détection des cas chez les personnes asymptomatiques non-cas contact ou la population ne se faisant pas tester. De plus, ces réseaux de sentinelles nécessitent des moyens financiers et humains importants pour tester les patients. Des méthodes de surveillance moins invasives et plus représentatives de la population générale sont donc nécessaires.

2. Utilisation des eaux usées comme moyen de surveillance des virus respiratoires

La surveillance des virus respiratoires dans la population par analyse des prélèvements chez les patients crée une limite dans la surveillance globale de la circulation du virus. Les virus respiratoires, sont excrétés et détectables dans les selles, à la fois chez les personnes symptomatiques, présymptomatiques et asymptomatiques [17], [18], [19], [20]. Il a été montré que les virus respiratoires excrétés par les selles pouvaient par la suite être détectés dans les eaux usées [21], [22]. La surveillance des virus respiratoires par le biais des eaux usées est en conséquence une des propositions pour dépasser cette limite.

Les eaux usées sont des matrices très variables pouvant rendre les analyses difficiles. La variabilité de celles-ci provient de plusieurs facteurs, tels que le nombre d'habitants, le tourisme, les industries, la météo, etc. Mais, il a été montré qu'elles étaient représentatives du nombre d'habitants présents en amont du prélèvement. Suivant le site d'échantillonnage les résultats peuvent concerner un nombre variable d'habitants [21], [22]. Les prélèvements peuvent aussi bien être faits en amont des stations d'épuration, qu'au niveau d'un bâtiment (hôpital, école…), d'un quartier ou encore d'un avion [21], [22], [23], [24], [25].

3. Suivi génomique de l'évolution des virus respiratoires dans les eaux usées et dans la population

Le suivi de la circulation des virus respiratoires chez les patients se faisant tester dans un laboratoire ou dans un hôpital est généralement réalisé par qPCR. Cela permet d'avoir un résultat rapide (quelques heures) et de réaliser la détection de plusieurs virus en simultané [26]. De nombreux systèmes de qPCR commerciaux ou « maison » existent et permettent de tester un large panel de virus d'intérêt, notamment des panels de virus respiratoires [26].

Mais les systèmes qPCR ne permettent pas de détecter de nouvelles mutations ou l'apparition d'un nouveau clade. De plus, lorsque les mutations ont lieu au niveau des sites de fixation, la sensibilité peut baisser et même provoquer une absence de détection [26].

3.1. Intérêt du suivi génomique des virus respiratoires

Les virus respiratoires sont sujets à des mutations dans le temps. Les virus à ARN ont tendance à muter plus vite. Ces mutations peuvent modifier le fitness du virus, en

modifiant sa virulence, sa capacité à se répandre ou encore provoquer un échappement immunitaire.. Ce sont en général la cause des nouvelles épidémies et des pandémies.

C'est notamment à cause des mutations causant un échappement immunitaire, qu'il est nécessaire de mettre à jour régulièrement les vaccins Par exemple pour la grippe, il existe un rapport annuel réalisé par l'OMS visant au choix des nouvelles souches grippales à intégrer au vaccin pour chaque saison grippale [27]. Cela signifie qu'il est important de réaliser des suivis génomiques réguliers. Pour cela, il est nécessaire d'avoir des bases de données riches et représentatives de la diversité virale.

3.2. Les systèmes de PCR-multiplexés

Pour avoir des bases de données riches et représentatives de la diversité virale, il est nécessaire de réaliser le séquençage du génome complet. L'essor des technologies NGS a permis d'agrandir ces bases de données. Toutefois, afin d'obtenir un génome complet, il est nécessaire d'avoir une grande quantité d'acide nucléique. Pour cela, il faut soit avoir une forte charge virale initiale, soit réaliser une pré-amplification. Cette amplification peut être réalisée par culture virale. Mais, cette dernière reste contraignante, n'est pas toujours applicable et peut introduire des modifications de séquences [1], [28].

Il est aussi possible de réaliser des pré-amplifications ciblées du virus par PCR. Mais pour cela, il est nécessaire d'avoir un système d'amorces spécifiques. Dans le cas du Covid-19, un système ARTIC a été mis en place. Il s'agit d'un système universel, basé sur une PCR multiplexe, mis à jour régulièrement, composé de plusieurs centaines d'amorces permettant l'amplification du génome complet de tous les variants du SARS-CoV-2 [29]. Les FluA et B possèdent aussi chacune un système d'amplification [30], [31]. Pour les autres virus respiratoires, des systèmes basé sur le système ARTIC (système ARTIC-like) font peu à peu leurs apparition, c'est le cas notamment pour les HBoV avec un système développé dans notre laboratoire [32].

Objectifs et structure de la thèse

L'objectif de cette thèse est de mettre en lumière l'importance du suivi des virus respiratoires dans le temps afin de mieux comprendre et de mieux prévoir les épidémies et les pandémies, puis son application à la réalisation d'un suivi de la circulation des virus dans les eaux usées.

Une première partie s'est axée autour de ce qui a pu être réalisé pour la détection et le séquençage des virus respiratoires dans les eaux usées et est composée d'une revue bibliographique s'intéressant au SARS-CoV-2 et à d'autres virus respiratoires.

La deuxième partie s'est concentrée sur la détection du SARS-CoV-2 dans les eaux usées d'avions en provenance d'Addis-Abeba et ayant atterri à Marseille pendant la pandémie de SARS-CoV-2, dans un contexte de mesures sanitaires imposées aux voyageurs.

La troisième partie s'est focalisée sur l'évolution génomique des influenza A H3N2 et H1N1 chez les patients entre 2017 et 2022 à l'IHU en s'intéressant plus particulièrement aux réassortiments inter-sous-types.

La quatrième partie s'est intéressée à l'importance du séquençage dans le cas de potentiels faux-négatifs, avec un exemple concret sur la grippe H3N2 en 2022.

Chapitre I : De la détection au séquençage des virus respiratoires dans les eaux usées

Avant-propos :

La pandémie de Covid-19, causée par le virus SARS-CoV-2, a permis de montrer l'importance de la surveillance épidémiologique pour prévenir et gérer les épidémies, ainsi que les pandémies virales causées par les virus respiratoires.

Actuellement, la détection des virus respiratoires est réalisée directement chez les patients, majoritairement avec la réalisation de tests qPCR à partir d'échantillons nasopharyngés prélevés chez les patients dont l'état est asses sévère pour qu'on juge utile de les prélever [1], [26], [33]. Cela permet de connaître la prévalence des virus respiratoires dans la population. Toutefois, cette approche présente des limites. Les personnes qui viennent se faire tester pour des virus respiratoires sont soit des personnes symptomatiques, beaucoup plus rarement des cas contact souhaitant se faire tester. Les personnes asymptomatiques et les personnes ne souhaitant pas se faire tester ne sont donc pas recensées [34], [35]. Cela sous-évalue de fait la réelle circulation d'un virus dans la population. C'est pour cela qu'il est nécessaire de mettre en place de nouveaux moyens de surveillance indépendants des tests réalisés chez les patients symptomatiques et plutôt graves.

Plusieurs études ont montré que le SARS-CoV-2, mais aussi les autres virus respiratoires abordés dans ce manuscrit, peuvent également être excrétés dans les selles. Cette excrétion peut avoir lieu aussi bien chez les personnes symptomatiques ou non symptomatiques [17], [18], [19], [20]. De plus des études ont confirmées qu'il est possible de détecter les virus respiratoires dans les eaux usées [21], [22].

La détection de ces virus respiratoires permet de connaître les virus présents dans les eaux usées, leur évolution quantitative dans le temps et donc le niveau de leur circulation dans la population. Mais au-delà de la détection, des études ont montré qu'il était possible de séquencer ces particules virales, permettant de pouvoir détecter les virus, mais aussi d'identifier leurs clades et leurs mutations [25], [36], [37], [38]. Dans le cadre d'un suivi régulier, cela peut permettre de réaliser une détection précoce de variants avant leur détection dans la population.

Mais les eaux usées sont des matrices complexes et variables. De plus, la majorité des études sur les eaux usées concernent des études sur le SARS-CoV-2. Avant celle-ci, peu d'études ont été réalisées sur les autres virus respiratoires, et certains virus sont encore peu étudiés.

Dans le cadre de cette revue bibliographique, nous nous sommes donc intéressés aux travaux réalisés à travers le monde sur la détection des virus respiratoires dans les eaux usées que ce soit pour le SARS-CoV-2, que pour d'autres virus respiratoires, . Et que ce soit de la détection ponctuelle ou des suivis sur le long terme, allant jusqu'au séquençage de ces virus respiratoires [21], [22], [25], [36], [37], [38].

Article n°1:
From detection to sequencing of respiratory viruses in wastewater, for optimal monitoring of epidemic

Lorlane Le Targa, Nathalie Wurst, Bernard La Scola, Sarah Aherfi (en cours d'écriture, pas encore soumis)

From detection to sequencing of respiratory viruses in wastewater, for optimal monitoring of epidemic

Author list: Lorlane Le Targa[1,2], Nathalie Wurtz[1], Philippe Colson[1], Bernard La Scola[1*], Sarah Aherfi[1*]

[1] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Institut de Recherche pour le Développement (IRD), , Assistance Publique-Hôpitaux de Marseille (AP-HM), IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [2] Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France; [4], 264 rue Saint-Pierre, 13005 Marseille, France

*** Corresponding authors:** Bernard La Scola, Sarah Aherfi, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: bernard.la-scola@univ-amu.fr, aherfi.s@gmail.com

In tropical areas, Flu A and B circulate throughout the year and are not correlated with any particular seasonality. In temperate zones, Influenza epidemics are characterized by seasonality, with outbreaks occurring generally from late autumn to winter [1], [2], [3]. Nevertheless, Influenza epidemics remain unpredictable because they do not look alike from year to year. In addition, thanks to the health protocols carried out during the Covid-19 pandemics, the 2020-2021 period was marked by the absence of Influenza epidemics, only some cases are found in the world [4], [5] As for the 2021-2022 season, it was marked by a late epidemic with an epidemic peak in early spring in temperate zones [6] Influenza epidemics represent a major health issue. According to WHO data (October 2022) Influenza is responsible for 3 to 5 million severe cases of respiratory disease, 290,000 to 650,000 deaths worldwide each year [7] and for 5 to 6% of lower respiratory tract infections (LRTI) in the world [8] In France (67.5 million inhabitants) 3.0 to 8.9% of the population is affected each year, causing an average of 10,000 deaths [9] Influenza affects all population groups, but with a higher risk of complications in children under one year old, pregnant women (at any stage of pregnancy), people over 65 years old, as well as people with comorbidity such as immunocompromised people [3] According to Santé Publique France (November 2022), 90% of Influenza-related deaths occur in people over 65 years [9] demonstrating the importance of protecting high-risk groups.

In order to protect the most fragile people, the WHO [7] recommends the annual vaccination for health personnel and the populations at risk including pregnant women, children between 6 months and 5 years old, people over 65 years old, people with chronic diseases. In France, for example, in order to limit the impact of Influenza on the population, an annual vaccination program for caregivers and the frail takes place every year, generally from October to January, before the epidemic peak.

In addition to epidemics, Influenza has also been responsible for at least 4 pandemics for nearly a century: the "Spanish" Influenza (H1N1) in 1918, the "Asian" Influenza (H2N2) in 1957, the "Hong Kong" InfluenzaI (H3N2) in 1968 [10], [11] and H1N1pdm09 in 2009 [12] which is the latest Flu A epidemic to date. The great genetic variability and the large number of Influenza genotypes (respectively 10 and 13 for Flu A and B) come from the ability of Influenza segments to carry out reassortment called antigenic shift and the point mutations called antigenic drift [2], [3], [13], [14] Influenza pandemics are generally correlated with the appearance of a strain with a strong immune escape, which is linked to a genetic reassortment between different Influenza strains. Moreover, if the prediction of Influenza epidemics remains uncertain, it is even more difficult to predict the occurrence of a pandemic.

Human Coronaviruses

Human Coronaviruses are enveloped single-stranded (+)RNA viruses. They belong to the family Coronaviridae with the genus alphacoronaviruses which includes HCoV-229E and HCoV-NL63 and the genus betacoronaviruses which includes SARS-CoV, MERS-CoV, HCoV-OC43 and HCoV-HKU1 [2] SARS-CoV has two subspecies SARS-CoV and SARS-CoV-2, both of which may cause severe acute respiratory syndromes (SARS). SARS-CoV-2 is the causative agent of the current COVID-19 pandemic.

In tropical regions, seasonal coronaviruses (HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1) are present throughout the year and are not associated with any particularseasonality, whereas in temperate regions, an epidemic usually occurs during winter. SARS-CoV and MERS-CoV do not correlate with any particular seasonality, neither in tropical nor in temperate regions. Epidemics caused by coronaviruses are difficult to predict [1]

SARS-CoV has caused epidemics over one year period in 2002-2003. It appeared for the first time in China, and was detected in civet and then in bats [15] This epidemic caused 8,096 cases and 774 deaths in 27 countries. Since July 2003 no new human case has been reported [16] MERS-CoV was first detected in 2012 in Middle East, this virus originated in bats. The camel is the intermediate source between humans and MERS-CoV. Cases of MERS-CoV mainly occur in people in contact with camels. Sporadic cases of MERS-CoV still occur [16] According to the WHO since the emergence of MERS-CoV until October 2022, 2,600 cases have been confirmed with 935 deaths reported [17]

Although SARS-CoV-2 is currently the most widely monitored coronavirus in the world, patient surveillance systems already existed for seasonal coronaviruses in some countries. In the United States, the NREVSS (National Respiratory and Enteric Virus Surveillance System) carries out since 2014 a passive surveillance of seasonal coronaviruses, the results of this surveillance have shown that the seasonal coronaviruses affect in 31 to 45% of cases of people under 18 years and presents a risk of low respiratory tract infectionfollowing an infection in immunocompromised people, young children and the elderly [18].

SARS-CoV-2 appeared for the first time in 2019 in Wuhan. Its origin is still unclear but its genome shares 85% identity with bat SARS-like CoV [19].According to the WHO data [20] as of November 6, 2024 COVID-19 affected more than 750 million people (including more than 280 million in Europe) and more than 7 million deaths worldwide (including more than 2,2 million in Europe). Like Influenza, SARS-CoV-2 affects all age groups, with a higher risk of mortality and complications in people over 65 years of age, as well as in people with co-morbidities such as hypertension, diabetes, cardiovascular disease and/or respiratory disease [21]. Unlike Influenza, children are not included in the population at risk of COVID-19.

Human Parainfluenza Viruses

Human Parainfluenza Viruses or HPIV are enveloped single-stranded (-)RNA viruses. They belong to the family Paramyxoviridae with the genus Respirovirus which includes HPIV-1 and HPIV-3 and the genus Rubulavirus which includes HPIV-2 and HPIV-4 [2], [22].

HPIV serotypes are not evenly distributed in the population. HPIV-3 and HPIV-1 that belong to respiroviruses, are the most frequent serotypes with 52% and 26% of cases, respectively. Rubulaviruses represent only 14% of cases, with 12% of HPIV 2 and only 2% of HPIV 4 [23].

In tropical and subtropical regions, HPIV infections do not follow a particular seasonality. In other regions, the seasonality of HPIV epidemics depends on the serotype. HPIV-1 usually causes an epidemic every second year in autumn. HPIV-2 causes an epidemic every year also in autumn. HPIV3 causes a biannual outbreak in spring and in late autumn / early winter, especially when HPIV-1 has not caused an epidemic. And the seasonality of HPIV-4 is not well known due to the fact that it is the less isolated serotype [1], [2], [22].

HPIV affects both children and adults, with a higher risk of complication in young children and frail adults, especially those over 65 years of age and immunocompromised individuals. In frail people, the infection can cause severe and potentially fatal pneumonia. In USA, 1 to 2 million children per year are affected by any HPIV and during the epidemic season up to 40% of LRTI-related pediatric hospitalizations are related to HPIV. The age groups affected vary as for epidemics depending on the serotype, HPIV3 affects 50 to 92% of children between 1 and 3 years old while HPIV1 and 2 affect children between 2 and 6 years old. Due to its low representation it is not possible to estimate which group of people is most affected by HPIV-4 [2], [22].

Respiratory Syncytial Virus

Respiratory Syncytial Virus or RSV is an enveloped single-stranded (-)RNA virus. It belongs to the Pneumoviridae family, into the Orthopneumovirus genus [2].

In tropical regions RSV is present throughout the year without epidemics correlated with a particular season. In temperate regions RSV epidemics usually occur in autumn and winter [1], [2], [24].

According to data from the National Institute of Allergy and Infectious Diseases (NIAID) [25], RSV affects 64 million people and causes approximately 160,000 deaths per year worldwid, another studie found 118,200 (IR 94,600-149400) death per year [26]. RSV is one of the main pathogenic viruses in adults at risk (chronical ill, compromised immune systems), as well as in people over 65 years of age and children with chronic lung disease or congenital heart disease [2], [27]. RSV is also the main causative agent of bronchiolitis which is the leading cause of hospitalization in healthy

infants [28] Furthermore, infants affected early by RSV have an increased risk of developing asthma during childhood [29]. It has been estimated that in the United States RSV is responsible for 2 million cases each year in children under 5 years of age and causes nearly 58,000 hospitalizations in this age group [30].

Human Metapneumoviruses

Human Metapneumoviruses or HMPV are enveloped single-stranded (-)RNA viruses. They belong to the Pneumoviridae family, and Metapneumovirus genus [2].

In tropical regions HMPV are present throughout the year without epidemics correlated with a particular season. In temperate regions HMPV epidemics usually occur in winter or early spring [1], [2], [24].

HMPV is the second most common cause of hospitalization in children for acute respiratory infections after RSV and can be considered as severe as RSV [24]. Like RSV, HMPV affects children under 5 years of age, but the average age of the group affected by HMPV is higher than that affected by RSV, HMPV mainly affects children between 13 and 24 months while RSV mainly affects children aged 12 months and younger [24], [31].

Rhinoviruses

Rhinoviruses or RhV are non-enveloped viruses with a single stranded (+)RNA genome, they belong to the Picornavirus family. There are three identified species: A, B and C [31], [32].

Rhinoviruses circulate throughout the year without any seasonality in tropical regions, it seems to be the same for temperate regions, there would nevertheless be an increase in the number of cases in autumn and winter [1], [33]. Local epidemics of rhinoviruses can occur in health centers such as hospitals or nursing homes due to the fragility of people passing through these places.

Moreover, although they are often asymptomatic cases in children under 4 years of age, Rhinoviruses are responsible for more than half of all LRTIs worldwide and are particularly severe in the elderly, as well as in immunocompromised persons and persons with comorbidities [32], [33].

Human Bocavirus

Human Bocavirus or (HBoV) is a non-enveloped, single-stranded DNA virus. It is a member of the Parvoviridae family of the genus Bocaparvovirus. There are four genotypes of HBoV (HBoV1-4). Moreover, it is a recently discovered virus, and was first isolated in 2005 [34].

The seasonality of HBoV in temperate zones seems to vary from year to year, with some sources estimating that an increase in cases generally occurs in winter and spring, while others estimate that it occurs in autumn and winter [35]. The epidemic period for HBoV seems to vary greatly by region. In Brazil, for example, there is a higher prevalence of HBoV in April and early autumn in the southern region, whereas the prevalence is higher in spring and autumn in the northern region [1], [34], [35], [36], [37]. Studies show that HBoV1-2 are associated with respiratory infections and are the most frequently found genotypes, while HBoV3-4 are more associated with gastrointestinal disease. One of the characteristics of HBoV is that it is found in more than 50% of cases in co-infection with other viruses or bacteria [36], [38].

HBoV infections mainly affect children under 2 years of age and are rarely detected in adults or the elderly. Nevertheless, they remain one of the most frequently detected virus after RSV and Rhinovirus in young children [35].

Human Adenoviruses

Human adenoviruses or HAdV are non-enveloped double-stranded DNA viruses. They belong to the family Adenoviridae [39]. HAdVs are divided into seven families with over 60 genotypes causing infections in humans.

Since 2014 in the USA, the National Adenovirus Type Reporting System (NATRS) is a surveillance network of HAdV set up to passively monitor the different types of HAdV circulating in patients in the United States. In addition, retrospective studies have shown that HAdV-1 to HAdV-4 were the most common types between 2003 and 2016, HAdV-3, 4 and 7 are the most associated with acute respiratory illness. Since 2007, HAdV-14 has also been associated with outbreaks of acute respiratory illness in the United States in the general population and more specifically in U.S. military. Other types of HAdV affect the population but are not, or only marginally, associated with acute respiratory illness, but they are linked to ocular symptoms, such as conjunctivitis (mainly HAdV-3, 4, and 11), and digestive symptoms, such as acute gastroenteritis (primarily HAdV-3, 5, 7, 31, 40, and 41).[40], [41].

HAdV do not have a specific seasonality and circulate throughout the year in both temperate and tropical regions, epidemics are rare and usually occur in small groups. HAdV are particularly severe in immunocompromised people, especially HAdV-7 [1], [40], [42].

Method of monitoring epidemics in patients

In order to anticipate epidemics, it is important to be able to monitor the respiratory viruses in the population. The monitoring of these epidemics is usually based on the detection of these viruses in clinical samples collected from infected persons. The detection and identification of respiratory

viruses can be performed on different types of human samples, such as nasopharyngeal swabs, washes and aspirates [43], as well as ocular swab (especially for HAdV), throat swab and possible but not used rectal swab [33], [40]. Some of these viruses can also rarely be detected in blood and urine samples [44]. Diagnosis in medical laboratories of respiratory viruses are rarely performed by using the "ancestral" celle culture method. antigenic detection [45] but are less sensitive that the molecular detection, which is routinely used because it is faster and more accurate [43], [46].

Culture

Viral culture is one of the ancestral methods used for the detection and identification of respiratory viruses, it is a highly specific qualitative method allowing to characterize both known and new viruses, this method is based on the cytopathic effect of viruses. This method is little impacted by the mutations that can occur in respiratory viruses. It is also a long and tedious method, requiring special equipment and skills, it takes between 2 and 10 days to obtain information on the status of a sample, and some viruses are more difficult to isolate than others [3], [45]. A study carried out in 2003 showed that infected persons stayed in hospital for an average of 3.6 days and that it took an average of 7.7 days for the culture to become positive, concluding that the culture had no impact on the clinical decisions taken to treat the patient [45].

Antigenic

Direct Immunofluorescence Assay (DFA) and enzyme immunoassay (EIA) techniques are antigenic detection methods that allow for rapid results (10-60 min). Antigenic methods require specific antibodies for each virus to be studied, their sensitivity varies from one virus to another and there is no existing system for all respiratory viruses of interest. DFA requires specific expertise unlike EIA, commercial EIA method are easy to use[45], [46], [47].

Molecular

Molecular detection methods such as qPCR or RT-qPCR provide a rapid result within few hours. Another advantage of the qPCR is the possibility to perform multiplexes. Multiplexes allow the detection of a large panel of different viruses in order to decrease the risk of missing a diagnosis and are adapted to syndromic approach. Today, there are many commercial or home-made detection systems for respiratory viruses. Respiratory virus detection panels can play particular attention to certain types or subtypes of viruses and, depending on the design and methods used, the panels can be more or less sensitive for a given virus. For example, in the 2009/2010 panel, the FTD system had a sensitivity range of 25 (PIV2) to 100% (PIV4) depending on the virus, compared with monoplex qPCR tests. Specificity ranged from 96.8 (HBoV) to 100% (PIV1) compared with monoplex qPCR tests [48]. This method has also the possibility of being largely automated [46]. qPCR is limited by the fact

that the sequence of the virus to be detected must be known at least partially. Moreover, this method does not allow the detection of new viruses, unlike culture and metagenomics, it is an oriented method, only known viruses targeted by specific primers are searched for.Furthermore, with this method the evolution of the target viruses can reduce the sensitivity of the systems used requiring constant monitoring and evolution of the system [3], [46].

qPCR is the most widely used, but other PCR methods such as melting curve analysis, Voltammetry and Fluorescence-labeled bead array can also be used. These methods have the same limitations as qPCR [49], [50].

Among other molecular detection methods, there is the rapid molecular assay, which is based on the isothermal amplification of nucleic acids, it is a simple qualitative method that requires a simple heat source and detectiondetection by fluorescence, it is a rapid method with a result in 15 min having a good specificity and a good sensitivity. This test is particularly used for Influenza cases, the Alere I kit for the detection of flu A and B based on this system shows a sensitivity of 93.3% and 100% for Influenza A and B respectively compared to culture and 95.5 and 100% compared to qPCR [51] One of the limitations of this method is its cost [3].Monitoring epidemics in wastewater

Other epidemic monitoring systems for respiratory viruses exist, such as the detection of viruses in wastewater. This monitoring can make it possible to detect the first cases of infection in the preserved zones. But also, to carry out quantitative and/or semi-quantitative analyzes by monitoring the viral load from the start to the end of the epidemic..However, detection alone has limited value; sequencing provides more information about circulating strains. This can allow health measures to be taken as quickly and efficiently as possible and can predict to a certain level the arrival, strength and end of an epidemic. During the COVID-19 pandemic, wastewater monitoring was strongly developed. This detection system is applicable to the respiratory virus because human respiratory viruses are excreted in stools [35], [44], [52], [53], [54] in persons with and without gastrointestinal symptoms [55], [56]

Stool shedding of respiratory viruses occurs in both symptomatic, pre-symptomatic and asymptomatic individuals. In certain cases the virus can be excreted in the stools, although it is not detectable by nasopharyngeal swab, in the context of the COVID-19 pandemic a case of symptomatic patient, positive by fecal swab and negative by pharyngeal and sputum swab [57] Two cases of Influenza that were negative by nasopharyngeal swab but positive in fecal samples have also been reported [58]

It is therefore possible to detect the presence of respiratory viruses excreted in faeces in wastewater [59], [60] and those potentially before the detection of the first cases in screening centers or hospitals. In addition, this type of monitoring makes it possible to locate the areas within a city or a country where the virus is most prevalent, and thus to detect epidemic foci. The advantage of this type of monitoring is that it can be systemic, carried out on a regular basis in order to know the evolution of a virus within the population, and to prevent epidemics as well as possible.

Wastewater monitoring can also be applied to wastewater from long-haul transports such as airplanes or ships in order to track the export of epidemic viruses to different countries [61], [62]

Wastewater samples

The term wastewater covers a wide range of different types of water, which are used or produced by human activity and then discharged. DomesticD wastewater are water coming from domestic or household activities (grey water and black water). The presence of viruses in sewers in cities comes mainly from domestic wastewater. Industrial wastewater comee from the activity of industries. Agricultural and farm wastewater, from field runoff and animal waste. Stormwater and runoff. Black water from planes and boats. In addition, it is possible for several types of wastewater, from different sources (human, animal, industrial, etc.), to end up in the same network [60]

Within cities, samples can be taken at different levels of the wastewater network, both in city sewers and in WWTPs (wastewater treatment plants) [63] At the WWTP level, the sample can be taken before or after treatment, the analysis after treatment makes it possible to see if the latter allows the elimination or reduction of the viral load [64], [65] The knowledge of the wastewater networks by the cities allows to know the population living in the collection area, but also the type of community building present in the collection area (school/hospital/EPHAD/Bar). Knowledge of the topography of the collection area is essential to better understand the results, especially in areas with a high density of tourists or schools, where sampling may vary depending on the time of year. It is also important to know the rainfall and temperature of the collected areas, which can cause variations in the collection [66], [67] Moreover, depending on the water collected, it may be mixed with more or less chemical products or non-organic waste (toilet paper, etc.). But, also industrial waste, particularly from food processing companies, which may cross the urban network [62], [68]

In the case of black water from aircraft and ships, it is also important to know the chemicals used, as well as the method of cleaning the tanks and when they were cleaned, because aircraft tanks are emptied but not disinfected between each flight. In addition, the rate of non-organic waste (toilet paper) is higher in the tanks of airplanes and ships than in the wastewater from the sewers. Flight time can also have an impact on the possible detection of viruses, as detection will also depend on

283 the sanitary habits of travellers. A British study shows that in Great Britain less than 13% of people
284 defecate in the plane during a short-haul flight (<6 hours) compared with 39% of people on long-haul
285 flights (>6 hours[69] Knowing the number of passengers and crew during the flight is also an
286 important element to take into account during the analysis [62]

287 Wastewater in general is a highly variable matrix due to its broad definition, but also to collection
288 conditions, meteorology and population variation. It is therefore important in studies to take into
289 account the high variability of different wastewaters.

290 Collection method

291 The methods used to collect wastewater samples vary depending on the type of wastewater (sewers,
292 aeroplanes, etc.), but also on the collection site. Sewer systems can vary from one town to another,
293 particularly between rural and urban areas [66] but also between two towns in the same country or
294 in different countries. It is therefore necessary to base the collection method on the architecture of
295 the network; there is no single good method that is compatible with all networks.

296 For urban and rural networks, depending on the study carried out, it is possible to vary several
297 collection factors. Some studies require the collection of a single sample at a specific time, while
298 others necessitate taking multiple samples regularly over a period that can range from a few hours to
299 several months or several years. [66]Samples may be collected manually or using self-collectors., and
300 may consist of composite samples The choice of collection method may vary according to space, cost
301 and maintenance requirements [69]

302 Aircraft black water can be collected using valves either on the aircraft tank or on the tank of the
303 vacuum truck that collects the tanks [61], [62]One of the limitations in the quality of wastewater
304 collection from aircraft is the difficulty of cleaning the tanks and vacuum trucks, which are generally
305 just emptied between flights and collection, and are not systematically cleaned and disinfected. Virus
306 detection may therefore be due to viral residues from a previous aircraft [62] As for the collection of
307 wastewater from boats, this can be done at the membrane of the bioreactor at its inlet or outlet [50].

308 Passive samplers use a wide range of technologies, as electropositive filter cartridges (e.g. Zetapor®)
309 or Moore swabs, which consist of a piece of gauze tied with a thread and left in the water for 1 to 7
310 days [70] During the SARS-CoV-2 pandemic, tampon were used [71]

311 After collection, wastewater is generally stored at between -80 and 4°C [61], [62], [72], [73] but
312 sometimes at room temperature [64] For example, one of these studies showed that storage at 4°C
313 for less than two weeks was sufficient for SARS-CoV-2 [73].

Pre-treatment methods

As wastewater is a complicated matrix, it is often necessary to carry out pre-treatment, such as filtration through large-pore filters to eliminate the largest particles only or precede by centrifugation[61], [73]

As the concentration of viral particles in wastewater is rather low, it is often necessary to concentrate them before analysis. Several concentration methods exist and can be used, such as ultracentrifugation, ultrafiltration, flocculation sedimentation, polyethylene glycol (PEG) precipitation and electronegative/electropositive membrane adsorption-extraction [59], [62], [66], [74] The efficiency of the concentration methods may vary depending on the virus studied. Each virus analysed must be tested in order to find the optimum pre-treatment method.

In addition, wastewater is a complicated matrix that can contain PCR inhibitors such as chemicals contained in cleaning products, disinfectants or industrial waste [66], [68]Organic inhibitors include bile, urea, ethanol, phenol, polysaccharides, sodium dodecyl sulphates, humic acids, tannic acid and melanin [66], [75] Several strategies can be used to suppress inhibitors, such as sample dilution, addition of extraction steps, phenol-chloroform extraction, activated carbon treatment, ultrafiltration with a positively charged membrane, UV irradiation, dialysis, capture with magnetic beads or Sephadex G-100 and Chelex-100 combination treatments [75] The choice of inhibitor removal techniques will also vary according to the source of the sample and the characteristics of the wastewater network under study.

Normalization Method

When performing comparative analysis of multiple samples from different locations or time periods, normalization is necessary. As previously mentioned, wastewater is a highly variable matrix. For the same collection site at different times, variations can occur due to weather conditions, specific human or industrial activities. In the case of different collection sites, variations may also be related to different factors such as weather (distant regions), human factors (different population sizes, urban/tourist areas, etc.), industrial factors (different types of industries), and varying network structures.

To take into account these variabilities, normalization methods have to be applied. These normalization methods can standardize based on the population of the network by using human population markers. Such markers include fecal excretion markers like PMMoV (Pepper mild mottle virus), Herpes virus, Papillomavirus, and even enteric HAdV [60], [72], [76]. Other human presence markers can also be used to estimate the number of inhabitants, such as total phosphorus

concentration, total nitrogen, ammoniacal nitrogen, biochemical oxygen demand (BOD), chemical oxygen demand (COD), and ammonium-nitrogen (NH4-N) [77], [78], [79]. Normalization can also be performed per gram of biosolid [80].

Detection and identification method

The same detection methods can generally be used in wastewater as well as in clinical cases after pre-treatment and processing of samples. However, optimizations and testing may be necessary, particularly because the concentration is usually lower in water compared to that in patients, and there may also be more inhibitors present. Therefore, it is essential to use robust and sensitive techniques.

Wastewater based epidemiology

Monitoring of wastewater is already conducted for numerous substances or microorganisms, whether these are excreted by humans, animals, industries, or otherwise. Many studies exist on the detection of illicit drugs in wastewater, notably to better understand consumption rates, peak periods, etc. For example, studies show that cocaine consumption seems to peak on weekends, while cannabis consumption remains stable and comparable across several European cities [76], [81]. The consumption of legal drugs such as nicotine has also been researched [82]. There are also studies to detect harmful compounds in wastewater, such as xenoestrogens which can be associated with cancers [83], or mycotoxins [84]. In terms of microorganisms, studies have been conducted on bacterial diversity or the emergence of antibiotic resistance genes [85], [86]. Enteric viruses such as noroviruses, enteroviruses, or hepatitis A are also frequently monitored [87]. Other viruses have been studied as well, such as Monkeypox during the 2022 outbreak [88].

Viruses in wastewater

When to look, where to look and what to look for

Research into respiratory viruses in wastewater can serve several purposes, in particular prevention before an epidemic and monitoring during an epidemic. In the case of prevention, the aim is to warn the population of the risk of infection so that measures can be taken upstream to limit the epidemic and/or prepare an appropriate health response [66] Monitoring during an epidemic enables us to identify the areas most affected, the rate of presence in the general or local population and to track the evolution of the epidemic over time [66] Many respiratory viruses have a known seasonal pattern, which makes it possible to determine their likely periods of occurrence. This means that it is not necessarily needed to search for all viruses throughout the year, but only to search for them

some time before the risk period, whether in sewage, aircraft or ship wastewater. In the case of aircraft and ships, it is also possible to use health data from other countries to determine which flight or ship to follow. Routine analysis of air and sea transport can also be used to identify countries at risk of transmission and to restrict flights from at-risk regions in order to limit the entry of infected people and minimize a potential epidemic [64]

Application

Research on the detection of respiratory viruses in wastewater has been around for a long time and is conducted worldwide. Some viruses are more closely studied than others. Some studies focus on a single virus, notably in the case of SARS-CoV-2, while others examine a multitude of viruses. Epidemics and pandemics can shift interest towards different viruses over time.

SARS-CoV-2

Since the SARS-CoV-2 pandemic, research on this virus has surged rapidly. The pandemic has driven efforts to develop new methods for monitoring pandemics and epidemics. Wastewater research has become a widely applied approach for SARS-CoV-2, making it the most studied virus in wastewater.

Retrospective qPCR analyses of wastewater have notably revealed the presence of SARS-CoV-2 in Italy as early as December 18, 2019, more than a month before the first cases among tourists (January 30, 2020) and two months before the first documented local case (January 21, 2020) [89]

The detection and estimation of SARS-CoV-2 concentrations in wastewater by qPCR have been extensively conducted worldwide. This has been achieved either through direct detection of SARS-CoV-2, targeting regions such as the spike protein, nucleocapsid, or ORF1ab, or by using multiplex systems that allow for both detection and differentiation of variants of concern (VOC) [90]

One advantage of wastewater detection is its feasibility even in low- and middle-income countries. For example, an Indonesian study conducted in the Special Region of Yogyakarta province, where only 6% of households are connected to the sewer system, demonstrated a time lag of 64 hours between sample collection and results. Over a 10-week period from June 27, 2021, to late September 2021, 74% (191/258) of wastewater samples tested positive, showing that it is possible to establish a wastewater surveillance system to monitor pathogens in areas with limited sewer connectivity and low or middle income [91]

Several studies have used combined data sources, integrating wastewater concentration data with hospital data, to establish correlations between clinical cases and SARS-CoV-2 levels in wastewater. These studies indicate a time lag between the increase in wastewater concentrations and the

corresponding rise in clinical cases. For instance, a study conducted in the Piedmont region of Italy between September 2020 and February 2021 found that the correlation between wastewater concentrations and clinical cases was strongest with a delay of 7 to 21 days, with early detection in wastewater [92] In Greece's Attica region, another study conducted during a similar period (August 31, 2020, to March 21, 2024) showed the best correlation with a 5-day delay and an 8-day average concentration in wastewater [77] This study even attempted to estimate case numbers for two periods, predicting 86 (95% CI: 82–90) and 153 (95% CI: 148–159) cases, respectively, for the periods of February 15–17, 2021, and February 18–20, 2021, with observed results of 102 and 150 cases. Similar estimations were made for ICU admissions, with predicted numbers of 9 (95% CI: 8-10) and 14 (95% CI: 13-15) and actual results of 12 and 13, respectively [77] In Marseille, France, another study conducted over a similar period (July to December 2020) also found that an 8-day moving average significantly correlated clinical cases with SARS-CoV-2 RNA copies in wastewater [67] This study also observed that the different sanitary measures implemented during this period did not affect SARS-CoV-2 levels in wastewater or clinical cases, indicating that lockdowns did not slow down the spread in Marseille.

The extensive analyses performed on this virus in wastewater have also highlighted the potential limitations of qPCR tests. A study conducted in Arkansas, USA, showed that mutations in the Delta variant's S gene led to a loss of detection of this gene by the PCR method used between July and September 2021. However, the use of multiple targets allowed for continued detection of SARS-CoV-2 [90] In this study, the potential mutations causing decreased detection were identified using probes targeting known mutations in the S gene, commonly found in the circulating Delta variant at that time [90]

qPCR detection systems have also been coupled with whole genome sequencing. Sequencing of wastewater samples, as in patients, enables the identification of circulating clades. In the aforementioned Greek study, they sequenced samples from February 2021 and detected several VOCs, including B.1.1.7/UK lineage, B.1.351/South Africa lineage, and P.1/Japan-Brazil lineage, with 96.3% ± 2.2 of sequences belonging to the B.1.1.7/UK lineage [77] Additionally, potential new mutations, whether rare or dominant, were detected before their appearance in clinical cases. For instance, an American study conducted in Oregon from March 16 to April 21, 2021, detected the B.1.351 variant in Albany on March 26, although no cases had been identified in the counties or adjacent counties. Moreover, only 24 cases of this variant had been reported in the rest of the state. The first clinical case of the B.1.351 variant in Albany (Linn County) was detected on April 23, 28 days later. They showed that the sequences found in wastewater clustered with those found in the population [93] . An American study conducted in San Diego from December 2020 to September

2021, showed that Alpha and Mu genotypes were detected by sequencing in wastewater two and four weeks before clinical cases, respectively. They also observed that the increase in Delta variant levels was visible five days before the rise in clinical cases. Despite the challenging wastewater matrix, they achieved a median coverage of 96.49% (range 75.67%-100%) with Ct values up to 38 [94]. A French study conducted on April 22, 2021, performed qPCR analysis and sequencing on two sewage networks in the city of Marseille (France): a primary network covering most of the city's population, and a sub-network of the main one. The results obtained from wastewater could be compared to COVID-19 diagnoses among residents of these networks conducted three days before and ten days after wastewater sampling. Their findings demonstrated differences in variant circulation within a sub-population. In the main network, 90.5% of patients tested positive for the UK variant and 1.7% for the B.1.525 variant, while in the sub-network, patients tested positive at 57.9% and 36.9% for the UK and B.1.525 variants, respectively. These results were close to those obtained by sequencing and qPCR testing: in the main network, the UK variant was found at rates of 95% by sequencing and 70% by qPCR, while in the sub-network, the UK variant was present at 64% and 8%, respectively. The researchers concluded that the variant proportions found by wastewater sequencing appeared to more closely reflect the actual patient distribution than the qPCR results [73], [95].

Beyond sewer wastewater analysis, SARS-CoV-2 monitoring has also been conducted in airplane and ship wastewater. However, studies on these waters remain limited, especially for ships.

An Australian study tested wastewater from a ship on April 23, when only crew members were present, and passengers had disembarked a month earlier. This study demonstrated the feasibility of detecting SARS-CoV-2 in ship wastewater. It also tested wastewater from three different flights between May and April 2020 (8 to 14 hours of flight). Samples were collected at the vacuum truck after the planes were emptied. Only one flight tested positive, and none of the passengers were positive or became positive even after 14 days of isolation, suggesting possible contamination of the vacuum truck by another aircraft, as the trucks were emptied but not disinfected [62] Another Australian study tested twelve flights (10 between October and November 2021, one in April, and one in August 2021) and detected SARS-CoV-2 in only two samples. The study also conducted whole genome sequencing of the positive samples using two different amplification methods and showed that ATOPlex performed better than ARTIC V3 (99% coverage versus 61%) on certain Omicron variants. Other studies on airplane wastewater have been conducted worldwide. In France, a study on two flights from Addis Ababa (Ethiopia) on December 22 and 24, 2021, detected SARS-CoV-2, with the second flight's positive result confirmed by nasopharyngeal tests performed on passengers. Additionally, the BA.1 variant was detected in both passengers and wastewater by sequencing. For

476 the second flight, 20% of passengers (12/56) tested positive for SARS-CoV-2 during a period when a
477 negative PCR test was required to travel to France, highlighting the limitations of such protocols
478 between false positives and potentially falsified tests [61]

479 Studies on a larger number of planes have also been conducted in the United States and the United
480 Arab Emirates. The study conducted in Qatar on 198 flights between May and July 2020 found a
481 positivity rate of 13.6% [64] The U.S. study tested 88 flights at John F. Kennedy International Airport,
482 New York City, between August and September 2022, with flights from France, the United Kingdom,
483 and the Netherlands, and found 81% (65/88) of flights positive, with the same proportion of positives
484 for all origins. They obtained sequencing for 25 samples, yielding 27 genomes, including 21 BA.1
485 genomes, one BA.4.6, and one BA.2.75. Additionally, they identified four BA.5 sublineages from two
486 samples containing two genomes [96]

487 The detection of SARS-CoV-2 through antigen tests has also been explored in wastewater. This was
488 carried out using automated chemiluminescent immunoassay tests on wastewater samples collected
489 from various markets in Bangkok. The study successfully detected SARS-CoV-2 in all samples that
490 were also positive by qPCR [97]

491 ## Adenovirus

492 After SARS-CoV-2, HAdV are among the most frequently studied viruses in wastewater. This is largely
493 because most HAdV are not responsible for respiratory diseases but rather for enteric diseases. Only
494 a few adenovirus clades are associated with respiratory syndromes. In most cases, HAdV are not
495 typed in wastewater, making it difficult to identify the presence of HAdV responsible for respiratory
496 symptoms. Additionally, detection of HAdV in wastewater shows that serotype 41 is the most
497 commonly found serotype [98], [99] This serotype does not cause respiratory symptoms.

498 The detection rate of HAdV in wastewater samples varies from study to study but is often above 50%
499 throughout the year [98]. A one-year detection study conducted in the Eastern Cape Province (South
500 Africa) reported a positive sample rate of 64% using q-PCR method [98] In contrast, a similar study
501 conducted in Greater Cairo reported a 100% positivity rate for wastewater [76] In all these studies,
502 HAdV were not typed

503 For aircraft wastewater, a study conducted at Darwin International Airport (DRW) in Australia on
504 planes from India and the United Kingdom found ADV in 54.2% (any subtype) of tested planes
505 (13/24) [100]

Beyond urban or aircraft wastewater, HAdV are also commonly found in recreational waters, drinking water, rivers, and the ocean [98] The high prevalence of HAdV in humans explains why they are often used as markers of fecal contamination, alongside other enteric viruses or fecal bacteria like *E. coli*, coliform bacteria, and coliphages [76] Approximately 92% of adults are seropositive for at least one adenovirus [84].

Influenza virus

Due to the importance of the Influenza virus in human epidemics, it has been monitored in wastewater for a long time. Although the significantly increased since the rise of SARS-CoV-2, several studies show a correlation between the number of clinical cases and the rate of Influenza virus in wastewater [101], [102] While often only the differentiation between Influenza A and B is performed, detecting Influenza in wastewater also allows for understanding the prevalence of different circulating genotypes. A study conducted in Japan from October 2018 to January 2023 showed that the prevalence of Influenza A was very low during the SARS-CoV-2 pandemic. However, it also detected changes in the prevalence of circulating genotypes. During the 2018/2019 period, H3 was more prevalent than H1pdm09, whereas in 2019/2020, H1pdm09 was predominant and H3 was absent from wastewater. The same trend was observed in clinical cases. In 2022/2023, H3 became dominant again with no detection of H1pdm09 in water and in patients [103] An American study conducted in Nevada, across 6 primary schools between January 20 and June 13, 2022, as well as on a WWTP receiving wastewater from the schools, showed that it was possible to detect Influenza A in water directly from the schools but not in the WWTP [104] In this study, the feasibility of sequencing using NGS methods to detect different Influenza groups was demonstrated; they were able to detect subgroup 2a.2 sub-clade of group 3C.2a1b.3a. The prevalence rate of Influenza in wastewater varies from country to country and sometimes even from one WWTP to another. A German study conducted in Saxony from January to May 2022, across two different WWTPs, notably showed differences. In the first study, Influenza A was detected in only 5.2% of samples, compared to 41.6% in the second. For Influenza B, the rates were 36% and 57.7%, respectively. In this same study, sequencing showed that Influenza A types were H3N2, and they could not differentiate between different lineages of Influenza B [105] In an another study conducted in the USA from February 2021 to June 2022, the trend was reversed, with higher positivity for Influenza A detected in 28% of samples compared to only 6% for Influenza B [80] This demonstrates that results can vary from one country to another and from one city to another.

In airplane an Australian team showed that influenza A was detected in 16.7% of 24 airplanes coming from the UK and India, whereas Influenza B was not detected by RT-qPCR[100]

In the case of Influenza, as with COVID, it sometimes happens that Influenza is not directly searched for but rather studied through biomarkers, such as medications specifics or not such as antipyretics, antivirals (e.g. oseltamivir for Influenza, nirmatrelvir for SARS-CoV-2) and cough suppressants, to determine the prevalence of SARS-CoV-2 and Influenza in wastewater [106]

Respiratory Syncytial Virus

Before the SARS-CoV-2 epidemic, RSV was rarely monitored in wastewater. Only one study conducted in Sapporo, Japan, began researching RSV prior to the SARS-CoV-2 outbreak. This study showed that between October 2018 and April 2020, RSV A and B had positivity rates of 73% and 55% respectively. From May 2020 to January 2023, RSV A and B had positivity rates of 39% and 7%, respectively, with RSV A becoming significantly higher than RSV B during this second period. In all period, range of RSV A concentration is higher ($1{,}06 \times 10^{-6}$ –$5{,}91 \times 10^{-4}$ GC/L) than RSV B ($1{,}01 \times 10^{-6}$ –$3{,}07 \times 10^{-5}$GC/L) [103]. Since the SARS-CoV-2 outbreak, RSV is often monitored in wastewater alongside Influenza and SARS-CoV-2. Several studies show an association between the number of clinical cases and the rate of RSV in wastewater [72], [101], [107], [108] One Canadian study aimed to determine if it was possible to better detect the start and end of the RSV season to better guide the administration of palivizumab (monoclonal antibody). They showed a 12-day delay between data obtained from sewers and data obtained from clinical cases, with a strong Spearman correlation (p=0.9) [80]. A study conducted in Greece also shows an association between RSV found in wastewater and in patients, but this time with a one-week delay [108] All studies conducted on RSV detection in wastewater show a peak in positivity during the winter period. A Greek study notably shows a positivity rate of 55% (51/93), with an initial detection of RSV at the end of November, a peak at the end of December (2.8x10^11 GC/100k people), followed by a decrease in February and March before disappearance [109] To date, no NGS studies have been conducted on this virus in wastewater.

In the case of airplane wastewater, only one study has researched RSV. This Australian study, conducted on airplanes coming from the UK and India, shows an absence of RSV A detection and a positivity rate of 37.5% for RSV B [100]

Human Bocavirus

Since its discovery in 2005, the detection of HBoV in wastewater has been performed several times, mainly due to the fact that genotypes 2, 3, and 4 are known to cause gastrointestinal symptoms. HBOV 1, which causes respiratory syndromes, is the least commonly found HBoV in wastewater [88], [89], [90] An American study conducted between 2006 and 2009 in 11 states and 12 different

WWTPs showed a positivity rate of 81% (54/67) for all HBoV types combined [110] This study also demonstrated through Sanger sequencing that HBoV-2 was the most frequently found in wastewater, with 67% of identified strains belonging to this type, compared to 18% for HBoV-3 and 6% for HBoV-1. Furthermore, HBoV-1 was only found in Florida in this study, whereas other HBoVs were detected in all the tested WWTPs [110] An Italian study conducted between 2014 and 2015 in seven different regions, using qPCR and genomic sequencing, showed a similar detection rate, with 79% of samples testing positive for HBoV [111] Here too, HBoV-2 was the most commonly found, with 46% of samples positive, followed by HBoV-3 with 26%, and a complete absence of detection [111]of HBOV 1. A Thai study conducted between November 2016 and June 2018 showed similar results, with no detection of HBOV 1 in the WWTPs, a positivity rate of 79% for HBoV-2, and 7% for HBoV-3 [110] A more recent Australian study, conducted on 4 WWTPs in Queensland between January and June 2022, showed a 100% positivity rate for HBoV, with concentrations ranging from 4.96 to 7.22 log10 GC/L, but no typing was performed to differentiate the genotypes [101]

An Australian study previously cited showed that wastewater from airplanes coming from the UK and India had a 33% positivity rate for all HBoV types [100]

Rhinovirus

Rhinoviruses are rarely monitored in wastewater. It is only following the SARS-CoV-2 outbreak that they began to be studied, but there are still few studies on the subject. The primary method used is qPCR detection, with or without differentiation between RhV A and B, genotype C is not search in those studies. A study conducted over 17 months (01/02/2021 to 21/06/2022) in Santa Clara County, United States, showed a prevalence of 71.8% (155/216) for the presence of RhV in wastewater [80] In contrast, a study conducted over 6 months (04/01/2022 to 02/06/2022) in Queensland, Australia, showed a 100% prevalence (46/46) for both RhV A and RhV B [101] Furthermore, the Australian study showed that RhV A and B have similar concentrations in wastewater, with variations of 3.95–5.63 log10 GC/L and 3.74–5.61 log10 GC/L, respectively [101] Still in Australia, but in airplane wastewater, at Darwin International Airport (DRW), the prevalence of RhV A and B was also 100% among 24 airplanes coming from the UK and India [100] Sequencing was not used in the studies cited for rhinovirus.

Human Metapneumovirus

HMPV are also rarely monitored in wastewater. Studies on this subject are recent and were conducted after the SARS-CoV-2 pandemic. A study conducted in Santa Clara County (California, USA) between February 2021 and June 2022 showed that HMPV was detected in 56.1% of the samples

603 (121/216). The positivity rate was higher between November 2021 and June 2022 the end of the
604 study, with the highest viral concentration between November 2021 and January 2022, i.e., during
605 the winter period [80] Another study by the same team, conducted on samples from eight publicly
606 owned treatment works (POTWs) in the greater San Francisco Bay Area (California, USA) between
607 July 2022 and July 2023, also showed that HMPV is more prevalent in winter than at other times of
608 the year [113] This second study also revealed that the prevalence of HMPV varied from one
609 collection site to another, with periods of absence and presence, and differing peaks of presence.
610 The first detection dates in the different POTWs ranged from August to December, detection peaks
611 varied from November to April, and the last detection dates extended from March until the end of
612 the collection period [113] Currently, no NGS study has been conducted for this virus in wastewater.

613 Human Parainfluenza Virus

614 HPIV are also among the viruses for which research in wastewater only began recently. Few studies
615 exist on this subject, and they were conducted after the start of the SARS-CoV-2 epidemic. A study
616 conducted in WWTPs in Santa Clara County (California, USA) between February 2021 and June 2022
617 detected HPIV in 90.3% of the tested samples (195/216), with a significant positivity rate between
618 April 2021 and January 2022, and between April 2022 and the end of the study [80] This study also
619 showed that, out of 18 samples, the concentration of HPIV 3 was the highest, followed by HPIV 2 and
620 4B, with HPIV A and 4A being the least concentrated [80] A study conducted in Australia also
621 indicated that HPIV 3 was the most prevalent, with 19.6% of samples (9/46) testing positive, followed
622 by HPIV 2 with 10.7% (5/46), and HPIV 1 with 8.69% (4/46). HPIV 4 was not investigated in this study
623 [101] No sequencing in wastewater has been conducted in these studies, only detection.

624 Other Coronaviruses SARS-CoV-2 is currently one of the most monitored respiratory viruses in
625 wastewater, this is not the case for other coronaviruses, which have been and remain less studied in
626 this environment. In 2003, a study focused on the detection of SARS-CoV in wastewater [114] This
627 study was conducted after the SARS-CoV outbreak and aimed to detect the virus in the wastewater
628 of two hospitals in China. They were able to directly detect SARS-CoV in wastewater using semi-
629 nested qPCR, while cultures alone showed no cytotoxicity. It was only during the SARS-CoV-2
630 epidemic that a study focused on seasonal coronaviruses. This study, conducted in Santa Clara, USA,
631 on several respiratory viruses including seasonal coronaviruses (HCoV-229E, HCoV-NL63, HCoV-OC43,
632 and HCoV-HKU1), from February 2021 to June 2022, showed a 100% prevalence of HCoV in the
633 tested wastewater samples (216/216). Among 18 samples tested, HCoV-OC43 was present in 15
634 samples, HCoV-HKU1 and HCoV-NL63 in 6 samples each, and HCoV-229E in 3 samples [80] No study
635 with sequencing in wastewater has been found for this virus.

Discussion and perspective

Interest in monitoring respiratory viruses in wastewater has significantly increased since the onset of the SARS-CoV-2 pandemic. However, the surveillance and detection of respiratory viruses in wastewater has existed since the early 21st century, notably with the detection of SARS-CoV [114] Nevertheless, most respiratory viruses remain under-researched in this context. Prior to the rise of SARS-CoV-2, HAdV were among the most frequently studied viruses in wastewater, although most adenovirus serotypes do not cause respiratory symptoms, being primarily sought for their status as enteric viruses. Influenza virus was the second most researched virus in wastewater before the SARS-CoV-2 pandemic, largely due to its well-known role in causing multiple epidemics and global pandemics. Additionally, most of the respiratory viruses discussed in this review began to be studied in wastewater only after the SARS-CoV-2 pandemic. Among the other viruses discussed, only HBoV had also been studied in wastewater before the SARS-CoV epidemic. As with HAdV, this is primarily due to genotypes 2, 3, and 4 being responsible for gastrointestinal symptoms. Other respiratory viruses only started to be studied in wastewater since the SARS-CoV-2 pandemic.

Moreover, most studies focus on detection and monitoring using PCR and qPCR methods, often lacking genotyping, particularly for HAdV and HBoVs, not all of which are involved in respiratory syndromes. This is also the case for seasonal coronaviruses, which are rarely studied and seldom differentiated. Sequencing is still underutilized, whether for SARS-CoV-2 or other respiratory viruses, despite its potential to identify circulating variants and track the emergence of new ones. This is partly because direct sequencing in wastewater remains challenging. The wider use of the Artic protocol for SARS-CoV-2 (https://artic.network/) has facilitated better sequencing of this virus, both in patients and wastewater. Whole-genome sequencing methods, such as those for Influenza, have existed for a long time but are not commonly used in wastewater [115] Other systems, like one for HboV published in 2024, are too new to have been applied in this type of analysis [110].

Currently, most wastewater monitoring occurs in urban areas, though studies increasingly focus on wastewater from airplanes. As shown in various studies, sewer wastewater analysis allows for tracking pandemic progression, estimating case numbers, identifying circulating variants, and detecting new variants before clinical cases emerge. Research on airplane wastewater is less common, but such analysis could potentially monitor the spread of epidemics via travel and limit their impact. Studies on ship wastewater are even rarer.

Most studies are conducted in developed countries, though some are carried out in developing countries or regions with limited sewer networks. This demonstrates that wastewater analysis can be used almost anywhere, potentially enabling low-cost epidemic monitoring worldwide [91]

However, the various studies use different methods for sampling, extraction, normalization, and analysis, making it difficult to compare results and develop standardized methods at a regional, national, or international level. A review on SARS-CoV-2 quantification also highlights the lack of transparency in some studies regarding the tests performed, parameters used, and the quality control of these tests [117]

Conclusion

Wastewater analysis for detecting respiratory viruses is increasingly used, particularly since the SARS-CoV-2 pandemic. This approach could facilitate non-invasive, cost-effective monitoring of potential respiratory virus epidemics. However, this method remains heavily focused on SARS-CoV-2 and should be applied more frequently to other respiratory viruses that cause epidemics. Additionally, there is significant room for improvement if the goal is to conduct standardized analyses within a region, country, or globally. Sequencing, while still underutilized, provides crucial information on viral sequences and it is necessary to develop it more exhaustively in the aime to increase our knowledge on molecular epidemiology or respiratory viruses.

[1] J. M. Galván, O. Rajas, and J. Aspa, "Review of Non-bacterial Infections in Respiratory Medicine: Viral Pneumonia," *Arch Bronconeumol*, vol. 51, no. 11, pp. 590–597, Nov. 2015, doi: 10.1016/j.arbr.2015.09.015.

[2] R. L. Hodinka, "Respiratory RNA Viruses," *Microbiol Spectr*, 2016, doi: 10.1128/microbiolspec.DMIH2-0028-2016.

[3] F. Krammer *et al.*, "Influenza," *Nat Rev Dis Primers*, vol. 4, no. 1, p. 3, 2018, doi: 10.1038/s41572-018-0002-y.

[4] C. Adlhoch and R. Pebody, "What to expect for the influenza season 2020/21 with the ongoing COVID-19 pandemic in the World Health Organization European Region," *Euro Surveill*, vol. 25, no. 42, p. 2001816, Oct. 2020, doi: 10.2807/1560-7917.ES.2020.25.42.2001816.

[5] C. Adlhoch *et al.*, "Very little influenza in the WHO European Region during the 2020/21 season, weeks 40 2020 to 8 2021," *Euro Surveill*, vol. 26, no. 11, p. 2100221, Mar. 2021, doi: 10.2807/1560-7917.ES.2021.26.11.2100221.

[6] "Grippe." Accessed: Dec. 08, 2022. [Online]. Available: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/grippe

[7] "Influenza (Seasonal)." Accessed: Oct. 13, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[8] "Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017," *Lancet Respir Med*, vol. 7, no. 1, pp. 69–89, Jan. 2019, doi: 10.1016/S2213-2600(18)30496-X.

[9] "La grippe, une épidémie saisonnière." Accessed: Nov. 02, 2022. [Online]. Available: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/grippe/la-grippe-une-epidemie-saisonniere
[10] H. Brüssow, "The beginning and ending of a respiratory viral pandemic-lessons from the Spanish flu," *Microbial Biotechnology*, vol. 15, no. 5, pp. 1301–1317, 2022, doi: 10.1111/1751-7915.14053.
[11] V. St. Georgiev, K. A. Western, and J. J. McGowan, Eds., *National Institute of Allergy and Infectious Diseases, NIH: Frontiers in Research*. Totowa, NJ: Humana Press, 2008. doi: 10.1007/978-1-59745-569-5.
[12] C. Peteranderl, S. Herold, and C. Schmoldt, "Human Influenza Virus Infections," *Semin Respir Crit Care Med*, vol. 37, no. 04, pp. 487–500, Aug. 2016, doi: 10.1055/s-0036-1584801.
[13] CDC, "About Flu," Centers for Disease Control and Prevention. Accessed: Nov. 30, 2022. [Online]. Available: https://www.cdc.gov/flu/about/index.html
[14] N. J. Cox and K. Subbarao, "Global Epidemiology of Influenza: Past and Present," *Annu. Rev. Med.*, vol. 51, no. 1, pp. 407–421, Feb. 2000, doi: 10.1146/annurev.med.51.1.407.
[15] J. Cui, F. Li, and Z.-L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nat Rev Microbiol*, vol. 17, no. 3, Art. no. 3, Mar. 2019, doi: 10.1038/s41579-018-0118-9.
[16] E. de Wit, N. van Doremalen, D. Falzarano, and V. J. Munster, "SARS and MERS: recent insights into emerging coronaviruses," *Nat Rev Microbiol*, vol. 14, no. 8, pp. 523–534, 2016, doi: 10.1038/nrmicro.2016.81.
[17] "Middle East respiratory syndrome: global summary and assessment of risk - 16 November 2022." Accessed: Jan. 04, 2023. [Online]. Available: https://www.who.int/publications-detail-redirect/WHO-MERS-RA-2022.1
[18] M. E. Killerby *et al.*, "Human coronavirus circulation in the United States 2014–2017," *J Clin Virol*, vol. 101, pp. 52–56, Apr. 2018, doi: 10.1016/j.jcv.2018.01.019.
[19] N. Zhu *et al.*, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.
[20] "WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data." Accessed: Dec. 08, 2022. [Online]. Available: https://covid19.who.int/table
[21] Z. Zheng *et al.*, "Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis," *J Infect*, vol. 81, no. 2, pp. e16–e25, Aug. 2020, doi: 10.1016/j.jinf.2020.04.021.
[22] A. R. Branche and A. R. Falsey, "Parainfluenza Virus Infection," *Semin Respir Crit Care Med*, vol. 37, no. 4, pp. 538–554, Aug. 2016, doi: 10.1055/s-0036-1584798.
[23] A. M. Fry, A. T. Curns, K. Harbour, L. Hutwagner, R. C. Holman, and L. J. Anderson, "Seasonal Trends of Human Parainfluenza Viral Infections: United States, 1990–2004," *CLIN INFECT DIS*, vol. 43, no. 8, pp. 1016–1022, Oct. 2006, doi: 10.1086/507638.
[24] N. Akhras, J. B. Weinberg, and D. Newton, "Human metapneumovirus and respiratory syncytial virus: subtle differences but comparable severity," *Infect Dis Rep*, vol. 2, no. 2, p. e12, Aug. 2010, doi: 10.4081/idr.2010.e12.
[25] "Respiratory Syncytial Virus (RSV) | NIAID: National Institute of Allergy and Infectious Diseases." Accessed: Aug. 07, 2024. [Online]. Available: https://www.niaid.nih.gov/diseases-conditions/respiratory-syncytial-virus-rsv
[26] T. Shi *et al.*, "Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study," *Lancet (London, England)*, vol. 390, no. 10098, p. 946, Sep. 2017, doi: 10.1016/S0140-6736(17)30938-8.
[27] "Respiratory Syncytial Virus (RSV) | NIH: National Institute of Allergy and Infectious Diseases." Accessed: Dec. 05, 2022. [Online]. Available: https://www.niaid.nih.gov/diseases-conditions/respiratory-syncytial-virus-rsv

[28] C. A. Green *et al.*, "Admission to hospital for bronchiolitis in England: trends over five decades, geographical variation and association with perinatal characteristics and subsequent asthma," *Arch Dis Child*, vol. 101, no. 2, pp. 140–146, Feb. 2016, doi: 10.1136/archdischild-2015-308723.
[29] C. Rosas-Salazar *et al.*, "Respiratory syncytial virus infection during infancy and asthma during childhood in the USA (INSPIRE): a population-based, prospective birth cohort study," *The Lancet*, vol. 401, no. 10389, pp. 1669–1680, May 2023, doi: 10.1016/S0140-6736(23)00811-5.
[30] C. B. Hall *et al.*, "The Burden of Respiratory Syncytial Virus Infection in Young Children," *N Engl J Med*, vol. 360, no. 6, pp. 588–598, Feb. 2009, doi: 10.1056/NEJMoa0804877.
[31] S. Bosis, S. Esposito, H. G. M. Niesters, P. Crovari, A. D. M. E. Osterhaus, and N. Principi, "Impact of human metapneumovirus in childhood: Comparison with respiratory syncytial virus and influenza viruses," *Journal of Medical Virology*, vol. 75, no. 1, pp. 101–104, 2005, doi: 10.1002/jmv.20243.
[32] "Rhinoviruses and Respiratory Enteroviruses: Not as Simple as ABC - PMC." Accessed: Jan. 04, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728576/
[33] S. E. Jacobs, D. M. Lamson, K. St. George, and T. J. Walsh, "Human Rhinoviruses," *Clin Microbiol Rev*, vol. 26, no. 1, pp. 135–162, Jan. 2013, doi: 10.1128/CMR.00077-12.
[34] M. Guido *et al.*, "Human bocavirus: Current knowledge and future challenges," *World J Gastroenterol*, vol. 22, no. 39, pp. 8684–8697, Oct. 2016, doi: 10.3748/wjg.v22.i39.8684.
[35] D. Polo, A. Lema, E. Gándara, and J. L. Romalde, "Prevalence of human bocavirus infections in Europe. A systematic review and meta-analysis," *Transboundary and Emerging Diseases*, vol. 69, no. 5, pp. 2451–2461, 2022, doi: 10.1111/tbed.14233.
[36] O. Schildgen *et al.*, "Human Bocavirus: Passenger or Pathogen in Acute Respiratory Tract Infections?," *Clin Microbiol Rev*, vol. 21, no. 2, pp. 291–304, Apr. 2008, doi: 10.1128/CMR.00030-07.
[37] P. E. Silva *et al.*, "Human bocavirus in hospitalized children under 5 years with acute respiratory infection, São Paulo, Brazil, 2010," *Arch Virol*, vol. 163, no. 5, pp. 1325–1330, 2018, doi: 10.1007/s00705-017-3694-5.
[38] F. M. Moesker *et al.*, "Human bocavirus infection as a cause of severe acute respiratory tract infection in children," *Clinical Microbiology and Infection*, vol. 21, no. 10, p. 964.e1-964.e8, Oct. 2015, doi: 10.1016/j.cmi.2015.06.014.
[39] "Adenovirus Clinical Overview for Healthcare Professionals | CDC." Accessed: Nov. 30, 2022. [Online]. Available: https://www.cdc.gov/adenovirus/hcp/clinical-overview.html
[40] A. M. Binder *et al.*, "Human Adenovirus Surveillance — United States, 2003–2016," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 66, no. 39, pp. 1039–1042, Oct. 2017, doi: 10.15585/mmwr.mm6639a2.
[41] W.-J. Shieh, "Human adenovirus infections in pediatric population - An update on clinico–pathologic correlation," *Biomedical Journal*, vol. 45, no. 1, p. 38, Sep. 2021, doi: 10.1016/j.bj.2021.08.009.
[42] "Adenovirus Outbreaks | CDC." Accessed: Dec. 07, 2022. [Online]. Available: https://www.cdc.gov/adenovirus/outbreaks.html
[43] J. Mahony *et al.*, "Development of a Respiratory Virus Panel Test for Detection of Twenty Human Respiratory Viruses by Use of Multiplex PCR and a Fluid Microbead-Based Assay," *J Clin Microbiol*, vol. 45, no. 9, pp. 2965–2970, Sep. 2007, doi: 10.1128/JCM.02436-06.
[44] C. Mishra, S. Meena, J. K. Meena, S. Tiwari, and P. Mathur, "Detection of three pandemic causing coronaviruses from non-respiratory samples: systematic review and meta-analysis," *Sci Rep*, vol. 11, p. 16131, Aug. 2021, doi: 10.1038/s41598-021-95329-4.
[45] A. K. Shetty, E. Treynor, D. W. Hill, K. M. Gutierrez, A. Warford, and E. J. Baron, "Comparison of conventional viral cultures with direct fluorescent antibody stains for diagnosis of community-acquired respiratory virus infections in hospitalized children:," *The Pediatric Infectious Disease Journal*, vol. 22, no. 9, pp. 789–794, Sep. 2003, doi: 10.1097/01.inf.0000083823.43526.97.
[46] R. S. Buller, "Molecular Detection of Respiratory Viruses," *Clin Lab Med*, vol. 33, no. 3, pp. 439–460, Sep. 2013, doi: 10.1016/j.cll.2013.03.007.

[47] S. L. Johnston and C. S. Siegel, "Evaluation of direct immunofluorescence, enzyme immunoassay, centrifugation culture, and conventional culture for the detection of respiratory syncytial virus.," *J Clin Microbiol*, vol. 28, no. 11, pp. 2394–2397, Nov. 1990.
[48] S. Bierbaum *et al.*, "Detection of respiratory viruses using a multiplex real-time PCR assay in Germany, 2009/10," *Arch Virol*, vol. 159, no. 4, pp. 669–676, 2014, doi: 10.1007/s00705-013-1876-3.
[49] E. B. Popowitch, S. S. O'Neill, and M. B. Miller, "Comparison of the Biofire FilmArray RP, Genmark eSensor RVP, Luminex xTAG RVPv1, and Luminex xTAG RVP Fast Multiplex Assays for Detection of Respiratory Viruses," *Journal of Clinical Microbiology*, vol. 51, no. 5, pp. 1528–1533, Dec. 2020, doi: 10.1128/jcm.03368-12.
[50] V. M. Pierce, M. Elkan, M. Leet, K. L. McGowan, and R. L. Hodinka, "Comparison of the Idaho Technology FilmArray System to Real-Time PCR for Detection of Respiratory Pathogens in Children," *Journal of Clinical Microbiology*, vol. 50, no. 2, pp. 364–371, Dec. 2020, doi: 10.1128/jcm.05996-11.
[51] J. J. Bell and R. Selvarangan, "Evaluation of the Alere I influenza A&B nucleic acid amplification test by use of respiratory specimens collected in viral transport medium," *J Clin Microbiol*, vol. 52, no. 11, pp. 3992–3995, Nov. 2014, doi: 10.1128/JCM.01639-14.
[52] O. J. Kweon, Y. K. Lim, H. R. Kim, and T.-H. K. and M.-K. Lee, "Fecal Respiratory Viruses in Acute Viral Respiratory Infection and Nasopharyngeal Diarrheal Viruses in Acute Viral Gastroenteritis: Clinical Impact of Ectopic Viruses Is Questionable," vol. 28, no. 3, pp. 465–472, Mar. 2018, doi: 10.4014/jmb.1707.07058.
[53] "Human bocavirus genotypes 1, 2, and 3 circulating in pediatric patients with acute gastroenteritis in Chiang Mai, Thailand, 20122018 | Elsevier Enhanced Reader." Accessed: Dec. 23, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S1876034120307632?token=A27589E88D9C1C5EA42413139C2CC28F2AAD18EFD14F0B3E789E248F972EE12BA067091B337AF12AEBA4125A62E9240A&originRegion=eu-west-1&originCreation=20221223101422
[54] J. Xie *et al.*, "Influenza Virus Detection in the Stool of Children with Acute Gastroenteritis," *J Clin Virol*, vol. 131, p. 104565, Oct. 2020, doi: 10.1016/j.jcv.2020.104565.
[55] K. S. Cheung *et al.*, "Gastrointestinal Manifestations of SARS-CoV-2 Infection and Virus Load in Fecal Samples From a Hong Kong Cohort: Systematic Review and Meta-analysis," *Gastroenterology*, vol. 159, no. 1, pp. 81–95, Jul. 2020, doi: 10.1053/j.gastro.2020.03.065.
[56] S. J. Tozer *et al.*, "Detection of human bocavirus in respiratory, fecal, and blood samples by real-time PCR," *J Med Virol*, vol. 81, no. 3, pp. 488–493, Mar. 2009, doi: 10.1002/jmv.21409.
[57] L. Chen, J. Lou, Y. Bai, and M. Wang, "COVID-19 Disease With Positive Fecal and Negative Pharyngeal and Sputum Viral Tests," *Am J Gastroenterol*, vol. 115, no. 5, pp. 790–790, May 2020, doi: 10.14309/ajg.0000000000000610.
[58] A. Sangkakam *et al.*, "Detection of influenza virus in rectal swabs of patients admitted in hospital for febrile illnesses in Thailand," *SAGE Open Med*, vol. 9, p. 2050312121989631, Jan. 2021, doi: 10.1177/2050312121989631.
[59] "Viruses in wastewater: occurrence, abundance and detection methods | Elsevier Enhanced Reader." Accessed: Dec. 23, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S0048969720344399?token=407EC9807AAC66AA8ECFDD1FC4FD5F351F10C042EE767AA5F34470972CEE584B6AC61D5C7C2D30031D67230C6BE251C6&originRegion=eu-west-1&originCreation=20221223130026
[60] K. Bibby and J. Peccia, "Identification of viral pathogen diversity in sewage sludge by metagenome analysis," *Environ Sci Technol*, vol. 47, no. 4, pp. 1945–1951, Feb. 2013, doi: 10.1021/es305181x.
[61] L. Le Targa *et al.*, "SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe," *Viruses*, vol. 14, no. 7, p. 1511, Jul. 2022, doi: 10.3390/v14071511.

[62] W. Ahmed *et al.*, "Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: a surveillance tool for assessing the presence of COVID-19 infected travellers," *Journal of Travel Medicine*, vol. 27, no. 5, p. taaa116, Aug. 2020, doi: 10.1093/jtm/taaa116.
[63] D. J. Price *et al.*, "Early analysis of the Australian COVID-19 epidemic," *eLife*, vol. 9, p. e58785, Aug. 2020, doi: 10.7554/eLife.58785.
[64] A. Albastaki *et al.*, "First confirmed detection of SARS-COV-2 in untreated municipal and aircraft wastewater in Dubai, UAE: The use of wastewater based epidemiology as an early warning tool to monitor the prevalence of COVID-19," *Sci Total Environ*, vol. 760, p. 143350, Mar. 2021, doi: 10.1016/j.scitotenv.2020.143350.
[65] S. Wurtzer *et al.*, "SARS-CoV-2 genome quantification in wastewaters at regional and city scale allows precise monitoring of the whole outbreaks dynamics and variants spreading in the population," *Sci Total Environ*, vol. 810, p. 152213, Mar. 2022, doi: 10.1016/j.scitotenv.2021.152213.
[66] J. Gonçalves *et al.*, "Centralized and decentralized wastewater-based epidemiology to infer COVID-19 transmission – A brief review," *One Health*, vol. 15, p. 100405, Dec. 2022, doi: 10.1016/j.onehlt.2022.100405.
[67] N. Wurtz *et al.*, "Viral RNA in City Wastewater as a Key Indicator of COVID-19 Recrudescence and Containment Measures Effectiveness," *Front Microbiol*, vol. 12, p. 664477, May 2021, doi: 10.3389/fmicb.2021.664477.
[68] M. Bayati *et al.*, "Identification and quantification of bioactive compounds suppressing SARS-CoV-2 signals in wastewater-based epidemiology surveillance," *Water Research*, vol. 221, p. 118824, Aug. 2022, doi: 10.1016/j.watres.2022.118824.
[69] C. Schang *et al.*, "Passive Sampling of SARS-CoV-2 for Wastewater Surveillance," *Environ. Sci. Technol.*, vol. 55, no. 15, pp. 10432–10441, Aug. 2021, doi: 10.1021/acs.est.1c01530.
[70] B. Moore, "The detection of enteric carriers in towns by means of sewage examination," *J R Sanit Inst*, vol. 71, no. 1, pp. 57–60, Jan. 1951, doi: 10.1177/146642405107100109.
[71] J. L. Kevill *et al.*, "Assessment of two types of passive sampler for the efficient recovery of SARS-CoV-2 and other viruses from wastewater," *Science of The Total Environment*, vol. 838, p. 156580, Sep. 2022, doi: 10.1016/j.scitotenv.2022.156580.
[72] B. Hughes *et al.*, "Respiratory Syncytial Virus (RSV) RNA in Wastewater Settled Solids Reflects RSV Clinical Positivity Rates," *Environ. Sci. Technol. Lett.*, vol. 9, no. 2, pp. 173–178, Feb. 2022, doi: 10.1021/acs.estlett.1c00963.
[73] N. Wurtz *et al.*, "Simple Wastewater Preparation Protocol Applied to Monitor the Emergence of the Omicron 21L/BA.2 Variant by Genome Sequencing," *Viruses*, vol. 15, no. 2, p. 268, Jan. 2023, doi: 10.3390/v15020268.
[74] M. Kitajima *et al.*, "SARS-CoV-2 in wastewater: State of the knowledge and research needs," *Sci Total Environ*, vol. 739, p. 139076, Oct. 2020, doi: 10.1016/j.scitotenv.2020.139076.
[75] C. Schrader, A. Schielke, L. Ellerbroek, and R. Johne, "PCR inhibitors – occurrence, properties and removal," *Journal of Applied Microbiology*, vol. 113, no. 5, pp. 1014–1026, 2012, doi: 10.1111/j.1365-2672.2012.05384.x.
[76] M. K. Rashed, W. M. El-Senousy, E. T. A. E. Sayed, and M. AlKhazindar, "Infectious Pepper Mild Mottle Virus and Human Adenoviruses as Viral Indices in Sewage and Water Samples," *Food Environ Virol*, vol. 14, no. 3, pp. 246–257, Sep. 2022, doi: 10.1007/s12560-022-09525-0.
[77] A. Galani *et al.*, "SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions," *Science of The Total Environment*, vol. 804, p. 150151, Jan. 2022, doi: 10.1016/j.scitotenv.2021.150151.
[78] F. Been, L. Rossi, C. Ort, S. Rudaz, O. Delémont, and P. Esseiva, "Population Normalization with Ammonium in Wastewater-Based Epidemiology: Application to Illicit Drug Monitoring," *Environ. Sci. Technol.*, vol. 48, no. 14, pp. 8162–8169, Jul. 2014, doi: 10.1021/es5008388.
[79] L. Pipes, Z. Chen, S. Afanaseva, and R. Nielsen, "Estimating the relative proportions of SARS-CoV-2 haplotypes from wastewater samples," *Cell Rep Methods*, vol. 2, no. 10, p. 100313, Oct. 2022, doi: 10.1016/j.crmeth.2022.100313.

910 [80] A. B. Boehm *et al.*, "Wastewater concentrations of human influenza, metapneumovirus,
911 parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids
912 during the COVID-19 pandemic: a surveillance study," *Lancet Microbe*, vol. 4, no. 5, pp. e340–
913 e348, May 2023, doi: 10.1016/S2666-5247(22)00386-X.
914 [81] K. V. Thomas *et al.*, "Comparing illicit drug use in 19 European cities through sewage analysis,"
915 *Science of The Total Environment*, vol. 432, pp. 432–439, Aug. 2012, doi:
916 10.1016/j.scitotenv.2012.06.069.
917 [82] R. S. Mackie *et al.*, "Trends in nicotine consumption between 2010 and 2017 in an Australian city
918 using the wastewater-based epidemiology approach," *Environ Int*, vol. 125, pp. 184–190, Apr.
919 2019, doi: 10.1016/j.envint.2019.01.053.
920 [83] W. Körner *et al.*, "Input/output balance of estrogenic active compounds in a major municipal
921 sewage plant in Germany," *Chemosphere*, vol. 40, no. 9, pp. 1131–1142, May 2000, doi:
922 10.1016/S0045-6535(99)00362-8.
923 [84] E. Gracia-Lor, E. Zuccato, F. Hernández, and S. Castiglioni, "Wastewater-based epidemiology for
924 tracking human exposure to mycotoxins," *Journal of Hazardous Materials*, vol. 382, p. 121108,
925 Jan. 2020, doi: 10.1016/j.jhazmat.2019.121108.
926 [85] "Wastewater-based epidemiology for tracking bacterial diversity and antibiotic resistance in
927 COVID-19 isolation hospitals in Qatar - Journal of Hospital Infection." Accessed: Aug. 07, 2024.
928 [Online]. Available: https://www.journalofhospitalinfection.com/article/S0195-6701(23)00275-
929 X/fulltext
930 [86] M. Colomer-Lluch, W. Calero-Cáceres, S. Jebri, F. Hmaied, M. Muniesa, and J. Jofre, "Antibiotic
931 resistance genes in bacterial and bacteriophage fractions of Tunisian and Spanish wastewaters as
932 markers to compare the antibiotic resistance patterns in each population," *Environment*
933 *International*, vol. 73, pp. 167–175, Dec. 2014, doi: 10.1016/j.envint.2014.07.003.
934 [87] L. Chacón, E. Morales, C. Valiente, L. Reyes, and K. Barrantes, "Wastewater-Based Epidemiology
935 of Enteric Viruses and Surveillance of Acute Gastrointestinal Illness Outbreaks in a Resource-
936 Limited Region," *Am J Trop Med Hyg*, vol. 105, no. 4, pp. 1004–1012, Oct. 2021, doi:
937 10.4269/ajtmh.21-0050.
938 [88] I. Girón-Guzmán *et al.*, "Spanish wastewater reveals the current spread of Monkeypox virus,"
939 *Water Research*, vol. 231, p. 119621, Mar. 2023, doi: 10.1016/j.watres.2023.119621.
940 [89] G. La Rosa *et al.*, "SARS-CoV-2 has been circulating in northern Italy since December 2019:
941 Evidence from environmental monitoring," *Sci Total Environ*, vol. 750, p. 141711, Jan. 2021, doi:
942 10.1016/j.scitotenv.2020.141711.
943 [90] C. S. Silva *et al.*, "Temporal dynamics of SARS-CoV-2 genome and detection of variants of concern
944 in wastewater influent from two metropolitan areas in Arkansas," *Sci Total Environ*, vol. 849, p.
945 157546, Nov. 2022, doi: 10.1016/j.scitotenv.2022.157546.
946 [91] I. K. Murni *et al.*, "The feasibility of SARS-CoV-2 surveillance using wastewater and environmental
947 sampling in Indonesia," *PLoS One*, vol. 17, no. 10, p. e0274793, 2022, doi:
948 10.1371/journal.pone.0274793.
949 [92] A. Robotto *et al.*, "Wastewater-based SARS-CoV-2 environmental monitoring for Piedmont,
950 Italy," *Environ Res*, vol. 203, p. 111901, Jan. 2022, doi: 10.1016/j.envres.2021.111901.
951 [93] M. Sutton *et al.*, "Detection of SARS-CoV-2 B.1.351 (Beta) Variant through Wastewater
952 Surveillance before Case Detection in a Community, Oregon, USA," *Emerg Infect Dis*, vol. 28, no.
953 6, pp. 1101–1109, Jun. 2022, doi: 10.3201/eid2806.211821.
954 [94] S. Karthikeyan *et al.*, "Wastewater sequencing reveals early cryptic SARS-CoV-2 variant
955 transmission," *Nature*, vol. 609, no. 7925, pp. 101–108, Sep. 2022, doi: 10.1038/s41586-022-
956 05049-6.
957 [95] N. Wurtz *et al.*, "Monitoring the Circulation of SARS-CoV-2 Variants by Genomic Analysis of
958 Wastewater in Marseille, South-East France," *Pathogens*, vol. 10, no. 8, p. 1042, Aug. 2021, doi:
959 10.3390/pathogens10081042.
960 [96] R. C. Morfino *et al.*, "Notes from the Field: Aircraft Wastewater Surveillance for Early Detection
961 of SARS-CoV-2 Variants - John F. Kennedy International Airport, New York City, August-

962 September 2022," *MMWR Morb Mortal Wkly Rep*, vol. 72, no. 8, pp. 210–211, Feb. 2023, doi:
963 10.15585/mmwr.mm7208a3.
964 [97] S. Thongpradit, S. Prasongtanakij, S. Srisala, S. Chanprasertyothin, E. Pasomsub, and B.
965 Ongphiphadhanakul, "The Detection of SARS-CoV2 Antigen in Wastewater Using an Automated
966 Chemiluminescence Enzyme Immunoassay," *Int J Environ Res Public Health*, vol. 19, no. 13, p.
967 7783, Jun. 2022, doi: 10.3390/ijerph19137783.
968 [98] O. Osuolale and A. Okoh, "Incidence of human adenoviruses and Hepatitis A virus in the final
969 effluent of selected wastewater treatment plants in Eastern Cape Province, South Africa," *Virol J*,
970 vol. 12, p. 98, Jun. 2015, doi: 10.1186/s12985-015-0327-z.
971 [99] "Microorganisms | Free Full-Text | Mastadenovirus Molecular Diversity in Waste and
972 Environmental Waters from the Lisbon Metropolitan Area." Accessed: Aug. 14, 2024. [Online].
973 Available: https://www.mdpi.com/2076-2607/10/12/2443
974 [100] W. Ahmed, W. J. M. Smith, A. Tiwari, A. Bivins, and S. L. Simpson, "Unveiling indicator,
975 enteric, and respiratory viruses in aircraft lavatory wastewater using adsorption-extraction and
976 Nanotrap® Microbiome A Particles workflows," *Sci Total Environ*, vol. 896, p. 165007, Oct. 2023,
977 doi: 10.1016/j.scitotenv.2023.165007.
978 [101] W. Ahmed *et al.*, "Occurrence of multiple respiratory viruses in wastewater in Queensland,
979 Australia: Potential for community disease surveillance," *Sci Total Environ*, vol. 864, p. 161023,
980 Mar. 2023, doi: 10.1016/j.scitotenv.2022.161023.
981 [102] T. de Melo, G. Islam, D. B. D. Simmons, J.-P. Desaulniers, and A. E. Kirkwood, "An alternative
982 method for monitoring and interpreting influenza A in communities using wastewater
983 surveillance," *Front Public Health*, vol. 11, p. 1141136, Jul. 2023, doi:
984 10.3389/fpubh.2023.1141136.
985 [103] H. Ando, W. Ahmed, R. Iwamoto, Y. Ando, S. Okabe, and M. Kitajima, "Impact of the COVID-
986 19 pandemic on the prevalence of influenza A and respiratory syncytial viruses elucidated by
987 wastewater-based epidemiology," *Sci Total Environ*, vol. 880, p. 162694, Jul. 2023, doi:
988 10.1016/j.scitotenv.2023.162694.
989 [104] V. Vo *et al.*, "Identification and genome sequencing of an influenza H3N2 variant in
990 wastewater from elementary schools during a surge of influenza A cases in Las Vegas, Nevada,"
991 *Science of The Total Environment*, vol. 872, p. 162058, May 2023, doi:
992 10.1016/j.scitotenv.2023.162058.
993 [105] "IJERPH | Free Full-Text | Simultaneous Detection of SARS-CoV-2 and Influenza Virus in
994 Wastewater of Two Cities in Southeastern Germany, January to May 2022." Accessed: Aug. 13,
995 2024. [Online]. Available: https://www.mdpi.com/1660-4601/19/20/13374
996 [106] X.-T. Shao *et al.*, "Surveillance of COVID-19 and influenza A(H1N1) prevalence in China via
997 medicine-based wastewater biomarkers," *Water Research*, vol. 247, p. 120783, Dec. 2023, doi:
998 10.1016/j.watres.2023.120783.
999 [107] E. Mercier *et al.*, "Wastewater-based surveillance identifies start to the pediatric respiratory
1000 syncytial virus season in two cities in Ontario, Canada," *Front Public Health*, vol. 11, p. 1261165,
1001 2023, doi: 10.3389/fpubh.2023.1261165.
1002 [108] M. Koureas *et al.*, "Wastewater Levels of Respiratory Syncytial Virus Associated with
1003 Influenza-like Illness Rates in Children—A Case Study in Larissa, Greece (October 2022–January
1004 2023)," *International Journal of Environmental Research and Public Health*, vol. 20, no. 6, Art. no.
1005 6, Jan. 2023, doi: 10.3390/ijerph20065219.
1006 [109] A. Zafeiriadou, L. Kaltsis, M. Kostakis, V. Kapes, N. S. Thomaidis, and A. Markou, "Wastewater
1007 surveillance of the most common circulating respiratory viruses in Athens: The impact of COVID-
1008 19 on their seasonality," *Science of The Total Environment*, vol. 900, p. 166136, Nov. 2023, doi:
1009 10.1016/j.scitotenv.2023.166136.
1010 [110] O. Blinkova *et al.*, "Frequent detection of highly diverse variants of cardiovirus, cosavirus,
1011 bocavirus, and circovirus in sewage samples collected in the United States," *J Clin Microbiol*, vol.
1012 47, no. 11, pp. 3507–3513, Nov. 2009, doi: 10.1128/JCM.01062-09.

1013 [111] M. Iaconelli, M. Divizia, S. Della Libera, P. Di Bonito, and G. La Rosa, "Frequent Detection and
1014 Genetic Diversity of Human Bocavirus in Urban Sewage Samples," *Food Environ Virol*, vol. 8, no.
1015 4, pp. 289–295, Dec. 2016, doi: 10.1007/s12560-016-9251-7.
1016 [112] K. Kumthip, P. Khamrin, A. Yodmeeklin, H. Ushijima, and N. Maneekarn, "Contamination of
1017 Human Bocavirus Genotypes 1, 2, 3, and 4 in Environmental Waters in Thailand," *Microbiol*
1018 *Spectr*, vol. 9, no. 3, p. e0217821, Dec. 2021, doi: 10.1128/spectrum.02178-21.
1019 [113] A. B. Boehm, M. K. Wolfe, B. J. White, B. Hughes, D. Duong, and A. Bidwell, "More than a
1020 Tripledemic: Influenza A Virus, Respiratory Syncytial Virus, SARS-CoV-2, and Human
1021 Metapneumovirus in Wastewater during Winter 2022–2023," *Environ. Sci. Technol. Lett.*, vol. 10,
1022 no. 8, pp. 622–627, Aug. 2023, doi: 10.1021/acs.estlett.3c00385.
1023 [114] X.-W. Wang *et al.*, "Excretion and detection of SARS coronavirus and its nucleic acid from
1024 digestive system," *World J Gastroenterol*, vol. 11, no. 28, pp. 4390–4395, Jul. 2005, doi:
1025 10.3748/wjg.v11.i28.4390.
1026 [115] B. Zhou *et al.*, "Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine
1027 Production for Classical and Swine Origin Human Influenza A Viruses," *Journal of Virology*, vol.
1028 83, no. 19, pp. 10309–10313, Oct. 2009, doi: 10.1128/jvi.01109-09.
1029 [116] H. Hikmat *et al.*, "Sequencing and characterization of human bocavirus genomes from
1030 patients diagnosed in Southern France between 2017 and 2022," *Journal of Medical Virology*,
1031 vol. 96, no. 6, p. e29706, 2024, doi: 10.1002/jmv.29706.
1032 [117] A. Bivins *et al.*, "Variability in RT-qPCR assay parameters indicates unreliable SARS-CoV-2 RNA
1033 quantification for wastewater surveillance," *Water Res*, vol. 203, p. 117516, Sep. 2021, doi:
1034 10.1016/j.watres.2021.117516.
1035

Chapitre II : Détection du SARS-CoV-2 dans les eaux usées des avions

Avant-propos :

Pendant la pandémie de Covid-19, les voyages en avion et en bateau ont été associés à la propagation du SARS-CoV-2, ainsi qu'à la circulation de nouveaux variants. À travers le monde, des mesures sanitaires ont été mises en place, telles que des quarantaines, une vaccination obligatoire et des restrictions d'entrée, afin de limiter la propagation par le biais des transports aériens et maritimes.. Pendant la première phase de propagation du Covid-19 lorsque l'épidémie était concentrée à Wuhan (Chine), des études ont constatées que ces mesures sanitaires ont seulement contribué à retarder la propagation du SARS-CoV-2 [39], [40]. Pourtant, des articles ont montrées que les mesures sanitaires mises en place ne suffisaient pas à endiguer complètement l'import du SARS-CoV-2 par les transports aériens et maritimes [41], [42].

Le variant Omicron a été décrit pour la première fois en Afrique du Sud et au Botswana début novembre 2021 [43]. Ce variant a été rapidement classé comme variant préoccupant (VOC) par l'ECDC, dès le 26 novembre 2021. En décembre 2021 le variant omicron était encore minoritaire en France et le variant Delta représentait la majorité des infections.

Des études ont montré qu'il était possible de détecter l'ARN du SARS-CoV-2 dans les selles des personnes infectées, qu'elles soient symptomatiques ou asymptomatiques [44]. Par conséquent, l'analyse du SARS-CoV-2 dans les eaux usées a été considérée comme une approche intéressante dans la surveillance du virus. Depuis le premier rapport de la détection du SARS-CoV-2 dans les eaux usées par Medema et al. aux Pays-Bas [45], la détection et la surveillance dans des échantillons d'eaux usées ont été expérimentées dans de nombreux pays [24], [46], [47]. Il a même été montré que le séquençage du SARS-CoV-2 dans l'eau usée est possible et peut permettre l'identification des variants présents dans la communauté, que ce soit des variants déjà détectés ou non dans la régions[47], [48], [48], [49], [50].

Lorsque cette étude a été réalisée, seules trois études avait montré qu'il était possible de détecter le SARS-CoV-2 dans les eaux usées des avions [24], [51], [52]. De plus, une de ces études avait montré qu'il était possible de réaliser le séquençage de l'ARN viral du SARS-CoV-2 dans cette matrice [52]. Dans le cas de cette partie, nous nous sommes donc intéressés à la recherche du virus du SARS-CoV-2 et notamment des variants Delta et

Omicron dans deux avions en provenance d'Addis-Abeba (Éthiopie) ayant atterri à Marseille (France) les 22 et 24 décembre 2021. La détection a été réalisée par méthode PCR et suivie d'un séquençage génomique. Pour le second avion, des tests nasopharyngés ont été proposés aux passagers, afin de corréler les souches retrouvées dans l'avion à celle des passagers qui s'avéreraient positifs. La conclusion était que cette méthode permettait une surveillance génomique de l'introduction de nouveaux variants du SARS-CoV-2 indépendante de tests qui auraient été réalisés chez les patients mais aussi de l'inefficacité des mesures restrictives de type pass sanitaire.

Article n°2:
SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe

Lorlane Le Targa, Nathalie Wurtz, Alexandre Lacoste, Gwilherm Penant, Priscilla Jardot, Alexandre Annessi, Philippe Colson, Bernard La Scola, Sarah Aherfi (Publier dans Viruses)

viruses

MDPI

Brief Report

SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe

Lorlane Le Targa [1,2,3], Nathalie Wurtz [1,2], Alexandre Lacoste [4], Gwilherm Penant [1,2], Priscilla Jardot [1,2], Alexandre Annessi [4], Philippe Colson [1,2], Bernard La Scola [1,2,*] and Sarah Aherfi [1,2,*]

1 Microbes Evolution PHylogénie et Infections, Institut de Recherche pour le Développement (IRD), Assistance Publique-Hôpitaux de Marseille (AP-HM), Aix-Marseille Université, 13005 Marseille, France; lorlane.le-targa@biosellal.com (L.L.T.); nathalie.wurtz@univ-amu.fr (N.W.); gwilherm.penant@ap-hm.fr (G.P.); priscilla.jardot@univ-amu.fr (P.J.); philippe.colson@univ-amu.fr (P.C.)
2 Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France
3 Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France
4 Bataillon des Marins Pompiers de la ville de Marseille, 13005 Marseille, France; alexandre.lacoste@bmpm.gouv.fr (A.L.); alexandre.annessi@bmpm.gouv.fr (A.A.)
* Correspondence: bernard.la-scola@univ-amu.fr (B.L.S.); aherfi.s@gmail.com (S.A.); Tel.: +33-413-732-401 (B.L.S. & S.A.); Fax: +33-413-732-402 (B.L.S.); +33-413-732-052 (S.A.)

check for updates

Citation: Le Targa, L.; Wurtz, N.; Lacoste, A.; Penant, G.; Jardot, P.; Annessi, A.; Colson, P.; La Scola, B.; Aherfi, S. SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe. *Viruses* **2022**, *14*, 1511. https://doi.org/10.3390/v14071511

Academic Editors: Matthew Scotch, Arvind Varsani and Crystal Marie Hepp

Received: 15 June 2022
Accepted: 7 July 2022
Published: 9 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Background: Most new SARS-CoV-2 epidemics in France occurred following the importation from abroad of emerging viral variants. Currently, the risk of new variants being imported is controlled based on a negative screening test (PCR or antigenic) and proof of up-to-date vaccine status, such as the International Air Transport Association travel pass. Methods: The wastewater from two planes arriving in Marseille (France) from Addis Ababa (Ethiopia) in December 2021 was tested by RT-PCR to detect SARS-CoV2 and screen for variants. These tests were carried out between landing and customs clearance and were then sequenced by MiSeq Illumina. Antigenic tests and sequencing by NovaSeq were carried out on respiratory samples collected from the 56 passengers on the second flight. Results: SARS-CoV-2 RNA suspected of being from the Omicron BA.1 variant was detected in the aircraft's wastewater. SARS-CoV2 RNA was detected in 11 [20%) passengers and the Omicron BA.1 variant was identified. Conclusion: Our work shows the efficiency of aircraft wastewater testing to detect SARS-CoV-2 cases among travellers and to identify the viral genotype. It also highlights the low efficacy of the current control strategy for flights entering France from outside Europe, which combines a requirement to produce a vaccine pass and proof of a negative test before boarding.

Keywords: SARS-CoV2; wastewater monitoring; variant spread; aircraft; surveillance

1. Introduction

Coronavirus disease 2019 (COVID-19], caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2], emerged in Wuhan (China) in December 2019. Since then, it has become a pandemic, with more than 336 million confirmed cases globally and 5.5 million deaths as of 20 January 2022. Most national SARS-CoV-2 epidemics that occurred successively or concurrently resulted from the importation from abroad of emerging viral variants [1]. Air travel and cruises have been associated with the spread of SARS-CoV-2, including the spread of new variants via infected passengers. Since the beginning of the COVID-19 pandemic, many countries and regions have imposed restrictions, including quarantine periods, entry bans, compulsory vaccination and travel restrictions. For information on the restrictions imposed in different countries, see the COVID-19 Travel Regulations Map developed by the International Air Transport Association (IATA), powered by Timatic. This tool provides real-time information on travel requirements according to itineraries

anywhere in the world (https://www.iatatravelcentre.com/world.php; accessed on 14 June 2022). Previous studies have examined the effect of travel restrictions and travel-related measures imposed during the pandemic. Most of these studies focused on the initial phase of the spread of COVID-19 when the epidemic was concentrated in Wuhan (China) [2,3]. All these studies found that travel restrictions in the early part of the epidemic helped to delay the spread of COVID-19. Other studies found that the restrictions were insufficient to completely control the global spread [4,5]. With the emergence of new SARS-CoV-2 variants, many countries have reinforced border control measures, including pre-travel and post-travel screening tests, to avoid the importation of these variants. In recent months, new variants have spread worldwide including, most recently, the Omicron variant which was first described in South Africa and Botswana [6]. Its clinical manifestations are similar to those of other respiratory viral infections with a dry cough, fever, tiredness, myalgia and breathing difficulties [6], but can also include gastrointestinal symptoms such as diarrhoea, nausea, abdominal pain and vomiting in between two and ten per cent of cases [7]. High concentrations of SARS-CoV-2 RNA have been found in the stools of infected asymptomatic and symptomatic people [8] and it has been shown that the virus remained infectious [9]. Therefore, analysis of SARS-CoV-2 in wastewater appears to be an interesting approach for monitoring the disease burden within communities. Since the first report of the detection of SARS-CoV-2 in wastewater by Medema et al. in the Netherlands [10], detection and monitoring in wastewater samples have been reported in many countries [11–14]. A few studies have performed SARS-CoV-2 genome sequencing from sewage to identify viral genotypes circulating within a community and to study genetic diversity [14–19]. Some of these showed a match between variants found in clinical isolates during the same period, while others identified genotypes not yet reported in clinical samples.

Until now, only three studies revealed that SARS-CoV-2 RNA had been detected in wastewater from passenger aircraft [20–22] and this monitoring demonstrated a high positive predictive value for SARS-CoV-2 infection among passengers. One of these works reported the successful detection by genome sequencing of variants in aircraft wastewater [20]. In our study, we report the detection of SARS-CoV-2 variants in the wastewater of aircraft travelling from Addis Ababa (Ethiopia) to Marseille (France). Two methods, including full-length genome sequencing and real-time reverse transcription (RT)-PCR (qPCR), to detect different variants by using the Bio-T Kit® FiveStar COVID-19 (Biosellal, Dardilly, France) were used. Following the detection of a high concentration of the Omicron variant in the wastewater of the first aircraft, a wastewater sample from a second flight was tested and passengers were offered a test upon disembarking, to assess whether the results correlated. This indicated the widespread importation of the Omicron variant into France from Africa. This confirms that monitoring aircraft wastewater provides precious public health information on the global spread of emerging SARS-CoV-2 variants and shows that production of a negative SARS-CoV-2 test before boarding is no guarantee that passengers are not carrying the virus.

2. Materials and Methods

2.1. Samples

A volume of 100 mL of aircraft wastewater samples from two flights arriving from Addis Ababa (Ethiopia) to Marseille (France) on 22 and 24 December (hereinafter referred to as flights "2212" and "2412", respectively) (Supplementary File SF1) were collected from the aircraft on the airport tarmac by the Bataillon des Marins Pompiers de Marseille (BMPM) via a special extraction valve which was washed between each sample by soaking for 15 min in bleach then 15 min in clean water. The samples were then stored at 4 °C until arrival at the laboratory. Samples were first passed through a paper filter to remove large particles, then a volume of 30 mL of the filtrate was filtered on a Millex sterile syringe filter with a pore size of 5 µm (SLSV025LS, Merck Millipore, Burlington, MA, USA).

For the flight on 24 December 2021, all passengers were offered a nasopharyngeal swab to test for SARS-CoV-2, in line with a joint initiative of the regional prefecture and the

regional health agency and in compliance with Decree No. 2020-551 of 12 May 2020 on the information systems (Article 11 of Law No. 2020-546 of 11 May 2020 extending the state of health emergency for people arriving from countries experiencing active circulation of the virus) [23]. These nasopharyngeal swabs were taken by the BMPM staff from 56 passengers and were tested using a rapid antigenic diagnosis test COVID-VIRO®, AAZ (Boulogne-Billancourt, France). All positive samples were transported to our laboratory at 4 °C for further RT-PCR testing and sequencing (see below).

2.2. Nucleic Acid Extraction

Prior to DNA/RNA extraction, 10 µL of Bio-T Kit® FiveStar COVID-19 internal positive control (Biosellal, Dardilly, France) and 10 µL of magnetic silica were added to each wastewater sample. Nucleic acids from 1 mL of each wastewater sample were extracted using the eGENE-UP® Lysis and RNA/DNA Purification (BioMérieux, Marcy l'Etoile, France) to obtain a volume of 100µL of eluate. Negative controls consisted of RNase Free water which was extracted following the same protocol.

For clinical samples, viral RNA was extracted from a volume of 200µL using the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA) to obtain a volume of 80 µL of eluate.

2.3. RT-PCR Detection and Variant Screening

Direct screening of SARS-CoV-2 variants in wastewater was performed using an RT-PCR Quantstudio5 device (Thermo Fisher, France) and a combination of the Bio-T Kit FiveStar COVID-19 and the Bio-T kit "Environmental Δ & O" (Biosellal, Dardilly, France). In addition, screening of N Gene (primers and probes of the CDC: www.cdc.gov/eid, accessed on 14 June 2022) was performed by RT-PCR, allowing assessment of the viral load. SARS-CoV-2 positive lateral flow results were confirmed by RT-PCR, as previously described [24].

2.4. Sample Preparation for NGS Sequencing

The first RNA/DNA extract from the 2212 sample was used without pre-treatment for further RT-PCR. For the 2412 sample, 1 mL was freeze-dried and then rehydrated in 30 µL of water. The reverse transcription step was performed in duplicate using the SuperScript VILO cDNA synthesis kit (11754-250, Thermo Fisher, Waltham, MA, USA) according to the supplier's recommendations in a final volume of 20 µL per reaction. The ARTIC v3 PCR (ARTIC nCoV-2019 V3 Panel and 500 rxn of IDT 10006788, Integrated DNA Technologies, Inc., Coralville, IA, USA) was carried out under the following conditions for one reaction for each of the pools 1 and 2: 2.5 µL of reaction mix 10 × 0.5 µL dNTP [10 mM), 0.4 µL of forward primer [100 nM), 0.125 µL of HotStart qDNA Polymerase (Qiagen 203205, Hilden, Germany), water PCR grade (qsp 25µL) and 2 µL of template. Eight replicates were made per extract and per pool. The eight replicates were then pooled (final volume of 200 µL) before purification on a NucleoFast 96-well plate (Macherey Nagel ref 743100.50, Hoerdt, France). The purification products were eluted in 30 µL of TE 1X and then placed on a 2% agarose gel (migration for 30 min, 100 V). The 400 base pair (bp) bands were cut out of the gel and purified according to the supplier protocol using the Monarch DNA Gel Extraction Kit (New England BioLabs, ref T1020L, Évry-Courcouronnes, France) with a final elution volume of 40 µL.

For passengers' samples, cDNA was amplified using the Illumina COVIDSeq protocol, including a multiplex PCR protocol with ARTIC nCoV-2019 V3 Panel primers (Integrated DNA technologies) according to the ARTIC procedure (https://artic.network/; accessed on 14 June 2022).

2.5. NGS Sequencing

For wastewater samples, final purification products were sequenced using the paired-end strategy with the Nextera XT DNA sample preparation kit (Illumina Inc., San Diego, CA,

USA). These samples were barcoded for mixing with other projects. Libraries were prepared following the Illumina protocol (Illumina Inc., San Diego, CA, USA). PCR amplification to complete tag adapters and introduce dual index barcodes was performed over 12 cycles followed by purification with 0.8 × AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, USA). Libraries were then normalised on specific beads according to the Nextera XT protocol (Illumina Inc., San Diego, CA, USA) and were then pooled. Automated cluster generation and pairwise sequencing with dual-index reads were performed in 2 × 250 bp using the Miseq Reagent Kit (V2-500 cycles) (Illumina Inc., San Diego, CA, USA). We chose to sequence wastewater samples using a MiSeq Illumina instrument to avoid possible cross contaminations with other clinical samples received in our laboratory and routinely sequenced on the NovaSeq 6000 instrument.

Viral genomes from clinical samples were sequenced using a NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA), as previously described [25].

2.6. Sequence Analysis

Reads of wastewater samples were analysed as previously described [15]. Briefly, the reads from pool1 and pool2 provided by the ARTIC procedure were mapped together against the Wuhan-Hu-1 SARS-CoV-2 isolate genome (GenBank accession number NC_045512.2] using the CLC genomics softwarev7.5 (Qiagen Digital Insights, Hilden, Germany) with the default parameters. Non-synonymous mutations present in more than 10% of the reads were taken into account. For each sample, non-synonymous mutations were individually compared with classifying mutations that matched with 225 SARS-CoV-2 variants and sub-variants that had been in circulation since the beginning of the pandemic, including those that were circulating at this time. We refer to the non-synonymous mutations mapped during the analysis as 'mutation patterns'. When a mutation pattern occurred in more than one variant or sub-variant, all variants and sub-variants were added to the results (Supplementary File SF2).

For clinical samples, genome consensus sequences were generated using the CLC Genomics workbench v.7 by mapping on the SARS-CoV-2 genome GenBank accession No. NC_045512.2 with the following thresholds: 0.8 for coverage and 0.9 for similarity. Sequences from complete genomes and clade assignments were analysed using the Nextclade web tool (https://clades.nextstrain.org/ accessed on 14 June 2022) [26].

3. Results

The results of screening the wastewater from flight 2212 were positive with a cycle threshold value (Ct) of 31.2, 29.8 and 34.4 for systems targeting the E gene, N gene and the E484A mutation, respectively, with a viral load of 171,699 copies/mL. The Ct value of the internal controls was 27. Wastewater from flight 2412 was also positive with a Ct value of 32.7, 30.6 and 34.9 for systems targeting the E gene, N gene and the E484A mutation, respectively, with a viral load of 95,846 copies/mL. The Ct value of the internal controls was 27.

For the 2212 sample, RNA concentration after extraction was 21.3 ng/µL. A total of 1,419,298 reads were obtained, and 97.3% were mapped, covering 75% of the reference genome. When considering a threshold of 10%, 31 non-synonymous mutations were present, 14 of them being signature mutations of SARS-CoV-2 variants (Supplementary File SF3). Furthermore, P4715L and D614G, which are present in the majority of SARS-CoV-2 variants, were found in 100% of the reads. Of 27 cumulative Omicron BA.1/21 K and BA.1.1 subvariant mutations covered by the reads, 13 were present at a frequency ranging from 51% to 100%. The following specific mutations of these subvariants were found: K856R, S3673_G3676S, I3758V, T547K and N856K. Of the 20 and 32 mutation patterns of the Omicron BA.1.1.529 and BA.2/21L subvariants covered by the reads, eight were found at a frequency ranging from 17% to 100%.

A total of 1,274,982 reads were obtained for the 2412 sample, of which 83.7% were mapped, covering 71.6% of the reference genome. Twenty-four non-synonymous mutations

were present when considering a threshold of 10%. The analysis of the reads revealed 15 non-synonymous mutation patterns specific to variants (Supplementary File SF4). The D614G mutation was found in 100% of the reads. Of the 34 Omicron BA.1 and BA.1.1 mutation patterns covered by the reads, nine were present, with a frequency of between 36% and 100%. The K856R, S2083_L2084delinsIle, S3673_G3676S and T547K mutations specific to these sub-variants were found. Of a total of 27 and 39 mutation patterns of Omicron BA.1.1.529 and BA.2, the sub-variants covered by the reads, six were present, with a frequency ranging from 25% to 100%. Six mutation patterns specific to the Delta variant (comprising between 10 and 18 of the mapped mutations) were present, with a frequency ranging from 11% to 100%.

For the aircraft that arrived on 24 December 2021, of the 56 passengers who were administered a rapid antigenic diagnostic test upon disembarking the plane, 12 [21%) were detected as positive. The result was obtained within 20 min and was communicated directly to the passengers. Eleven of these positive tests were confirmed by RT-PCR, the results of Ct for each patient are summarized in supplementary file 5 (Supplementary File SF5). SARS-CoV-2 next-generation genome sequencing using the COVIDSeq protocol was performed on these 11 samples, all of which were identified as the Omicron BA.1 variant.

4. Discussion

In this study, we screened the aircraft wastewater from two flights from Ethiopia to France. RT-PCR screening revealed the presence of the Omicron variant. Combined ARTIC and Illumina sequencing revealed the presence of mutation patterns of the Omicron variant. Interestingly, 12 passengers on the flight of 24 December 2021 tested positive using a lateral flow test and full genome sequencing of 11 of them revealed the presence of the Omicron variant. This number may also be underestimated since antigen tests are a little less sensitive than PCR tests. Thus, despite the requirement to have a negative PCR test within 72 h of boarding, as required even for vaccinated French nationals, one-fifth of the passengers on board this flight were infected. Furthermore, the congruent results of qPCR screening and variant detection by NGS of the aircraft wastewater and the genotype obtained on clinical samples showed that aircraft wastewater monitoring by NGS is effective for monitoring the circulation of variants, and is potentially a possible powerful strategy for preventing the widespread importation of new variants of concern from abroad. Of course, it would be important in future work to carry out an interrogation of the passengers to find out who went to the toilet during the trip, as well as the viral load of the positives, to determine as much as possible about the sensitivity of the procedure. However, since it is unthinkable to test all passengers on arrival, the detection of SARS-CoV-2 in the aircraft's wastewater is still necessarily proof of the presence of infected people.

Monitoring SARS-CoV-2 circulation in wastewater has already proven to be an effective tool for tracking infections at the community level and has been correlated with the number of individual cases [27]. Applying such an approach to aircraft wastewater may be a powerful tool for controlling SARS-CoV-2 importation and exportation, a risk which exists despite strict measures to control passengers through mandatory clinical negative testing. Recent studies have shown that SARS-CoV-2 monitoring of wastewater from international flights and cruise ships is a useful way of prioritizing testing of passengers, and of improving the management of contact tracing [20,22]. Ahmed et al. recently detected the Omicron variant through NGS of aircraft wastewater samples collected from a flight arriving in Darwin (Northern territory, Australia) from Johannesburg (Gauteng, South Africa) [23]. They combined the ARTIC approach with Oxford Nanopore-GridION technology and ATOPlex combined with the DNBseq-g400 sequencing. In this study, we also detected the Omicron variant in wastewater samples from a long-haul (nine hours) flight from Addis Ababa to Marseille. Although the wastewater tanks were cleaned, drained, washed with detergent and drained again between flights, the possible contamination by remaining traces of SARS-CoV-2 RNA in the blackwater tanks from previous flights cannot be excluded. However, the Ct obtained by testing wastewater has to be taken into account.

Here, the Ct were 27.99 and 30.24 (respectively for 2212 and 2412 flights) for the E gene corresponding to a viral load of 790,503 copies/mL and 169,462 copies/mL, respectively. This high viral load cannot be the result of remaining SARS-CoV2 RNA after cleaning tanks and is clearly linked to a viral excretion by onboard passengers, or even possibly the aircraft crew that was not tested for SARS-CoV-2 infection. The Omicron variant was designated as a variant of concern by the World Health Organization on 26 November 2021 and is described as highly transmissible, with a potential for immune escape as assessed by the reduced efficiency of the protective immunity developed after COVID-19 vaccination. Within three weeks of the first declared cases in Botswana, the Omicron variant had been detected in 87 countries [28]. Similarly, a previous study carried out in our laboratory showed that of sixteen SARS-CoV-2 variants identified in Marseille since the beginning of pandemic, seven had been imported through travel from abroad [1]. In order to learn from this experience, identify any potential pandemic waves from new emerging variants at an early stage and implement preventive measures, it has become indispensable to track the circulation of variants. The importation to Europe of the Omicron variant from South Africa illustrates perfectly that air traffic is a threatening and powerful entry point for new variants, despite public health policies including the "green" passport and compulsory RT-PCR testing for travelers (wrong certificates). This study shows that despite a requirement to show a negative RT-PCR test result within the 72 h before crossing the border, 20% of passengers were positive, highlighting a major failure of such prevention policies. Aircraft wastewater screening can be performed within an hour, a timeframe which is short enough to inform passengers before they clear customs, and for initiating nasopharyngeal testing and strict quarantine until the results are available. Such an approach could be an effective tool and is likely to be more powerful than presenting evidence of a negative test, which can be falsified. Aircraft and cruise wastewater monitoring may be of particular interest during inter-epidemic periods and in remote areas, where the massive importation of new SARS-CoV2 variants would have considerable public health consequences. Based on these findings, we propose SARS-CoV2 screening of wastewater followed by variant monitoring by NGS as a global strategy for preventing the importation of new SARS-CoV-2 variants to unaffected regions, especially in isolated areas such as islands or during periods of low viral circulation. We hope that the publication of such a manuscript on the subject of aircraft wastewater testing will allow us to convince our health authorities to carry out these tests on a large scale and not on an ad hoc basis but on an experimental basis, as the systematic screening and NGS of aircraft and boat wastewater may help policymakers to target management strategies by testing and isolating passengers in the event of the positive detection of SARS-CoV-2 in wastewater.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v14071511/s1. **SF1.** Summary of the results. **SF2.** Table of variants. **SF3.** Complete mutation matrix, showing all the signature mutations of the SARS-CoV2 variants, for which at least one signature mutation was detected in the wastewater sample from flight 2212. Grey cells are signature mutations of the SARS-CoV2 variant indicated in the corresponding row. Orange cells are signature mutations of the SARS-CoV2 variant indicated in the corresponding row and detected in the 2212 sample. The number indicated in the orange cells is the frequency of the detection of the mutation. These results were obtained by mapping the 2212 sample to the reference SARS-CoV-2 genome. (GenBank accession number NC_045512.2] using the CLC genomics software v7.5 (Qiagen Digital Insights, Germany) with default settings. Only non-synonymous mutations present on at least 10% of reads were recovered. Variant labels include the local nomenclature (IHU) and Nextstrain, and Pangolin nomenclatures. **SF4.** Complete mutation matrix, showing all the signature mutations of the SARS-CoV2 variants, for which at least one signature mutation was detected in the wastewater sample from flight 2412. Grey cells are signature mutations of the SARS-CoV2 variant indicated in the corresponding row. Orange cells are signature mutations of the SARS-CoV2 variant indicated in the corresponding row and detected in the 2412 sample. The number indicated in the orange cells is the frequency of the detection of the mutation. These results were obtained by mapping sample 2412 to the reference SARS-CoV-2 genome. (GenBank accession number

NC_045512.2] using the CLC genomics software v7.5 (Qiagen Digital Insights, Germany) with default settings. Only non-synonymous mutations present on at least 10% of reads were recovered. Variant labels include the local nomenclature (IHU) and Nextstrain and Pangolin nomenclatures. **SF5.** Results of Ct obtained from clinical samples collected from passengers of the flight of 24 December 2021.

Author Contributions: L.L.T. analysed the data and wrote the manuscript. N.W. and A.A. wrote the manuscript. A.L. collected the samples and performed PCR screening on wastewater and clinical samples. G.P. prepared samples for NGS and wrote the manuscript. P.J. performed NGS and wrote the manuscript. P.C. revised the manuscript. B.L.S. conceived the project, revised the manuscript and supervised the work. S.A. conceived data analysis, wrote the manuscript and supervised the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the French Government under the "Investissements d'avenir" (Investments for the Future) programme managed by the Agence Nationale de la Recherche (ANR, French National Agency for Research), reference: Méditerranée Infection 10-IAHU-03.

Institutional Review Board Statement: Since July 2021: according to Decree No. 2020-551 of 12 May 2020 on the information systems mentioned in Article 11 of Law No. 2020-546 of 11 May 2020 extending the state of public health emergency for people arriving from countries classified at risk if the virus circulates actively, all passengers arriving at Marseille airport are offered a nasopharyngeal swab for the antigenic detection of SARS-CoV-2 n according to a joint initiative of the regional prefecture and the regional health agency. Sampling and antigenic detection were performed by the Bataillon des Marins Pompiers de Marseille. The positive samples collected for diagnostic purposes were anonymously reused for retrospective genotyping. According to French law (the so-called "Loi Jardé"), anonymous retrospective studies do not require approval from an institutional review board nor informed consent. This study was approved by the Ethics Committee of the IHU Méditerranée Infection under number 2022-013.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: Lorlane Le Targa is an employee of Biosellal.

References

1. Colson, P.; Fournier, P.-E.; Chaudet, H.; Delerce, J.; Giraud-Gatineau, A.; Houhamdi, L.; Andrieu, C.; Brechard, L.; Bedotto, M.; Prudent, E.; et al. Analysis of SARS-CoV-2 Variants From 24,181 Patients Exemplifies the Role of Globalization and Zoonosis in Pandemics. *Front. Microbiol.* **2022**, *12*, 786233. [CrossRef] [PubMed]
2. Wells, C.R.; Sah, P.; Moghadas, S.M.; Pandey, A.; Shoukat, A.; Wang, Y.; Wang, Z.; Meyers, L.A.; Singer, B.H.; Galvani, A.P. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7504–7509. [CrossRef] [PubMed]
3. Chinazzi, M.; Davis, J.T.; Ajelli, M.; Gioannini, C.; Litvinova, M.; Merler, S.; Piontti, Y.; Pastore, A.; Mu, K.; Rossi, L.; et al. The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak. *Science* **2020**, *368*, 395–400. [CrossRef]
4. Gonzalez-Reiche, A.S.; Hernandez, M.M.; Sullivan, M.J.; Ciferri, B.; Alshammary, H.; Obla, A.; Fabre, S.; Kleiner, G.; Polanco, J.; Khan, Z.; et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **2020**, *369*, 297–301. [CrossRef]
5. Russell, T.W.; Wu, J.T.; Clifford, S.; Edmunds, W.J.; Kucharski, A.J.; Jit, M. Effect of internationally imported cases on internal spread of COVID-19: A mathematical modelling study. *Lancet Public Health* **2020**, *6*, e12–e20. [CrossRef]
6. Petersen, E.; Koopmans, M.; Go, U.; Hamer, D.H.; Petrosillo, N.; Castelli, F.; Storgaard, M.; Al Khalili, S.; Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **2020**, *20*, e238–e244. [CrossRef]
7. Alberca, G.G.F.; Solis-Castro, R.L.; Solis-Castro, M.E.; Alberca, R.W. Coronavirus disease—2019 and the intestinal tract: An overview. *World J. Gastroenterol.* **2021**, *27*, 1255–1266. [CrossRef]
8. Gu, S.; Chen, Y.; Wu, Z.; Chen, Y.; Gao, H.; Lv, L.; Guo, F.; Zhang, X.; Luo, R.; Huang, C.; et al. Alterations of the Gut Microbiota in Patients With Coronavirus Disease 2019 or H1N1 Influenza. *Clin. Infect. Dis.* **2020**, *71*, 2669–2678. [CrossRef]
9. Dergham, J.; Delerce, J.; Bedotto, M.; Scola, B.L.; Moal, V. Isolation of Viable SARS-CoV-2 Virus from Feces of an Immunocompromised Patient Suggesting a Possible Fecal Mode of Transmission. *J. Clin. Med.* **2021**, *10*, 2696. [CrossRef]
10. Medema, G.; Heijnen, L.; Elsinga, G.; Italiaander, R.; Brouwer, A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environ. Sci. Technol. Lett.* **2020**, *7*, 511–516. [CrossRef]

11. Ahmed, W.; Angel, N.; Edson, J.; Bibby, K.; Bivins, A.; O'Brien, J.W.; Choi, P.M.; Kitajima, M.; Simpson, S.L.; Li, J.; et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci. Total Environ.* **2020**, *728*, 138764. [CrossRef] [PubMed]
12. Wurtzer, S.; Marechal, V.; Mouchel, J.M.; Maday, Y.; Teyssou, R.; Richard, E.; Almayrac, J.L.; Moulin, L. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Eurosurveillance* **2020**, *25*, 2000776. [CrossRef] [PubMed]
13. Wurtz, N.; Lacoste, A.; Jardot, P.; Delache, A.; Fontaine, X.; Verlande, M.; Annessi, A.; Giraud-Gatineau, A.; Chaudet, H.; Fournier, P.-E.; et al. Viral RNA in City Wastewater as a Key Indicator of COVID-19 Recrudescence and Containment Measures Effectiveness. *Front. Microbiol.* **2021**, *12*, 664477. [CrossRef] [PubMed]
14. Izquierdo-Lara, R.; Elsinga, G.; Heijnen, L.; Munnink, B.B.O.; Schapendonk, C.M.E.; Nieuwenhuijse, D.; Kon, M.; Lu, L.; Aarestrup, F.M.; Lycett, S.; et al. Monitoring SARS-CoV-2 Circulation and Diversity through Community Wastewater Sequencing, the Netherlands and Belgium. *Emerg Infect. Dis.* **2021**, *27*, 1405–1415. [CrossRef]
15. Wurtz, N.; Revol, O.; Jardot, P.; Giraud-Gatineau, A.; Houhamdi, L.; Soumagnac, C.; Annessi, A.; Lacoste, A.; Colson, P.; Aherfi, S.; et al. Monitoring the Circulation of SARS-CoV-2 Variants by Genomic Analysis of Wastewater in Marseille, South-East France. *Pathogens* **2021**, *10*, 1042. [CrossRef] [PubMed]
16. Agrawal, S.; Orschler, L.; Lackner, S. Metatranscriptomic Analysis Reveals SARS-CoV-2 Mutations in Wastewater of the Frankfurt Metropolitan Area in Southern Germany. *Microbiol. Resour. Announc.* **2021**, *10*, e00280-21. [CrossRef]
17. Crits-Christoph, A.; Kantor, R.S.; Olm, M.R.; Whitney, O.N.; Al-Shayeb, B.; Lou, Y.C.; Flamholz, A.; Kennedy, L.C.; Greenwald, H.; Hinkle, A.; et al. Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* **2021**, *12*, e02703–e02720. [CrossRef]
18. Rios, G.; Lacoux, C.; Leclercq, V.; Diamant, A.; Lebrigand, K.; Lazuka, A.; Soyeux, E.; Lacroix, S.; Fassy, J.; Couesnon, A.; et al. Monitoring SARS-CoV-2 variants alterations in Nice neighborhoods by wastewater nanopore sequencing. *LANCET Reg. Health Eur.* **2021**, *10*, 100202. [CrossRef]
19. Wurtzer, S.; Waldman, P.; Levert, M.; Cluzel, N.; Almayrac, J.L.; Charpentier, C.; Masnada, S.; Gillon-Ritz, M.; Mouchel, J.M.; Maday, Y.; et al. SARS-CoV-2 genome quantification in wastewaters at regional and city scale allows precise monitoring of the whole outbreaks dynamics and variants spreading in the population. *Sci. Total Environ.* **2022**, *810*, 152213. [CrossRef]
20. Ahmed, W.; Bivins, A.; Smith, W.J.M.; Metcalfe, S.; Stephens, M.; Jennison, A.V.; Moore, F.A.J.; Bourke, J.; Schlebusch, S.; McMahon, J.; et al. Detection of the Omicron (B.1.1.529) variant of SARS-CoV-2 in aircraft wastewater. *Sci. Total Environ.* **2022**, *820*, 153171. [CrossRef]
21. Ahmed, W.; Bertsch, P.M.; Angel, N.; Bibby, K.; Bivins, A.; Dierens, L.; Edson, J.; Ehret, J.; Gyawali, P.; Hamilton, K.A.; et al. Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: A surveillance tool for assessing the presence of COVID-19 infected travellers. *J. Travel Med.* **2020**, *27*, taaa116. [CrossRef] [PubMed]
22. Albastaki, A.; Naji, M.; Lootah, R.; Almeheiri, R.; Almulla, H.; Almarri, I.; Alreyami, A.; Aden, A.; Alghafri, R. First confirmed detection of SARS-COV-2 in untreated municipal and aircraft wastewater in Dubai, UAE: The use of wastewater based epidemiology as an early warning tool to monitor the prevalence of COVID-19. *Sci. Total Environ.* **2020**, *760*, 143350. [CrossRef] [PubMed]
23. COVID 19: International Travel. Ministère de l'Intérieur 12 February 2022. Available online: https://www.interieur.gouv.fr/covid-19-international-travel (accessed on 14 June 2022).
24. Amrane, S.; Tissot-Dupont, H.; Doudier, B.; Eldin, C.; Hocquart, M.; Mailhe, M.; Dudouet, P.; Ormières, E.; Ailhaud, L.; Parola, P.; et al. Rapid viral diagnosis and ambulatory management of suspected COVID-19 cases presenting at the infectious diseases referral hospital in Marseille, France, - January 31st to March 1st, 2020: A respiratory virus snapshot. *Travel Med. Infect. Dis.* **2020**, *36*, 101632. [CrossRef] [PubMed]
25. Colson, P.; Lagier, J.C.; Baudoin, J.P.; Bou Khalil, J.; La Scola, B.; Raoult, D. Ultrarapid diagnosis, microscope imaging, genome sequencing, and culture isolation of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.* **2020**, *39*, 1601–1603. [CrossRef]
26. Aksamentov, I.; Roemer, C.; Hodcroft, E.B.; Neher, R.A. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* **2021**, *6*, 3773. [CrossRef]
27. Fernandez-Cassi, X.; Scheidegger, A.; Banziger, C.; Cariti, F.; Tunas, C.A.; Ganesanandamoorthy, P.; Lemaitre, J.; Ort, C.; Julian, T.R.; Kohn, T. Wastewater monitoring outperforms case numbers as a tool to track COVID-19 incidence dynamics when test positivity rates are high. *Water Res.* **2021**, *200*, 117252. [CrossRef]
28. Viana, R.; Moyo, S.; Amoako, D.G.; Tegally, H.; Scheepers, C.; Althaus, C.L.; Anyaneji, U.J.; Bester, P.A.; Boni, M.F.; Chand, M.; et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **2022**, *603*, 679–686. [CrossRef]

Chapitre III : Suivis du réassortiment de la grippe A/H3N2 et A/H1N1 chez les patients de 2017 à 2022

Avant-propos :

Les influenza virus font partie de la famille des *Orthomyxoviridae* [1]. Les influenzas virus A et B sont les plus répandus, dans les régions tempérées et causent chaque année des épidémies pendant l'hiver. Dans les régions tropicales, la grippe circule toute l'année sans période épidémique particulière avec toutefois un plus grand nombre de cas dans les périodes humides. Les épidémies à influenza virus A sont causées en particulier par les sous-types H1N1 et H3N2 [1], [2], [53].

Le virus influenza A, est un virus enveloppé à ARN simple-brin négatif d'environ 13,7kb. Une des particularités de ce virus est son génome segmenté, constitué de huit segments d'une longueur variant entre environ 840 et 2340 pb. Ces huit segments codent pour dix protéines, dont l'hémagglutinine (HA) et la neuraminidase (NA) codées respectivement par les segments n°4 et n°6 [1], [54].

L'une des difficultés dans la lutte aux *Alphainfluenzavirus* est leur capacité à évoluer rapidement et à causer des échappements immunitaires. Ceci est dû notamment à la capacité de la grippe à apporter des modifications aux épitopes viraux HA et NA (Antigenic drift : dérive antigénique) et à des échanges complets de segments par réassortiment HA et NA entre deux influenza A (Antigenic shift : un type de recombinaison) [54], [55].

Les réassortiments ont aussi lieu avec les autres segments. Ces réassortiments peuvent avoir lieu entre des grippes A de même sous-types ou de sous-types différents (ex : H1N1/H3N2) [54]. Les échanges entre sous-types différents peuvent notamment causer l'apparition de nouvelles épidémies, comme l'épidémie de grippe porcine H1N1 en 2009 (H1N1pdm09) [56].

Actuellement, la majorité des études sur l'influenza virus A sont réalisées sur les segments HA et NA [57], [58]. Cela est notamment dû au fait que HA et NA jouent un rôle majeur dans l'antigénicité de la grippe et sont les segments utilisés dans la construction des vaccins contre la grippe. Les autres segments sont moins étudiés et leurs réassortiments moins connus. Pourtant, avec l'augmentation de l'utilisation du NGS, les génomes complets sont de plus en plus accessibles, que ce soit par séquençage direct ou avec une pré amplification.

Ce sont les réassortiments des HA et NA entre deux sous-types différents qui sont les plus connus et étudiés. Les réassortiments intra-sous-type sont moins étudiés et les réassortiments des segments autre que HA et NA le sont encore moins.

Dans un premier temps, nous avons réalisé un suivi de l'évolution de la grippe dans les eaux usées, mais les charges virales faibles et dégradation des virus rendaient compliquée cette tâche. De plus, la recherche des réassortiments n'avait jamais été faite dans notre laboratoire et est peu étudiée en France et dans le monde en général. C'est pour cela que nous avons décidé de commencer par réaliser cette étude sur des échantillons de patients.

C'est dans ce cadre que nous nous sommes intéressés à mieux comprendre le réassortiment de tous les segments inter-sous-types des grippes A H3N2 et H1N1. Pour détecter et décrire les réassortiments, il est nécessaire d'avoir une large plage de données. C'est pour cela que dans le cadre de cette partie nous nous sommes intéressés à une période allant de 2017 à 2022 en séquençant plus de 3 000 échantillons positifs à la grippe, prélevés à Marseille (France). La période étudiée et de plus intéressante, car pendant la saison hivernale 2020-2021, très peu de cas de grippes ont été recensés dans les pays tempérés et en France [59], [60]. Ce faible nombre de cas est probablement lié aux mesures sanitaires imposées pendant l'épidémie de Covid-19.

Nous nous sommes dans un premier temps concentrés sur les grippes A/H3N2 présentes dans nos échantillons, en raison de l'évolution plus fréquente de ce sous-type [57]. La grippe A/H3N2 est composée de plus de clades que la grippe A/H1N1, avec 33 clades H3N2 en circulation entre 2017 et 2022, contre 15 H1N1 pendant la même période. C'est dans un second temps que nous nous sommes intéressés, avec une autre étudiante de l'équipe, aux grippes A/H1N1 présentes dans nos échantillons.

Article n°3 :
Monitoring of Influenza A/H3N2 reassortment in the southern France between 2017-2022

Lorlane Le Targa, Houmadi Hikmat, Celine Boschi, Laurence Thomas, Raphael Tola, Bernard La Scola, Philippe Colson (en cours d'écriture)

Monitoring of Influenza A/H3N2 reassortment in the southern France between 2017-2022

Short title (for the running head): Influenza A/H3N2 reassortment

Author list: Lorlane Le Targa[1,2], Houmadi Hikmat[1], Celine Boschi[1] Laurence Thomas[1], Raphael[1] Tola, Bernard La Scola[1*], Philippe Colson[1*]

Affiliations:

[1] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Institut de Recherche pour le Développement (IRD), , Assistance Publique-Hôpitaux de Marseille (AP-HM), IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [2] Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France; [4], 264 rue Saint-Pierre, 13005 Marseille, France

*** Corresponding authors:** Bernard La Scola, Philippe Colson IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: bernard.la-scola@univ-amu.fr, philippe.colson@univ-amu.fr

Key words: Influenza; H3N2; Reassortement

ABSTRACT

The influenza virus is a major human health issue, causing annual epidemics and sometimes global pandemics. Influenza evolves rapidly through the appearance of minor modifications in the HA and NA antigenic sites (antigenic drift) as well as through its ability to exchange entire gene segments (antigenic shift: a type of reassortment). To date, intra-subtype reassortments of influenza are still not well studied. In this study, we analyzed 843 complete genomes of A/H3N2 influenza collected from 2017 to 2022 in Marseille, France, before and after the COVID-19 pandemic, and demonstrated that the intra-subtype reassortment rate varied from year to year. The public health measures implemented during the COVID-19 pandemic also appear to have influenced the circulation of variants and the reassortment rate.

INTRODUCTION

Influenza A viruses is a virus of the Orthomyxoviridae family of the Alphainfluenzavirus genus. It is an enveloped virus, with negative single-stranded RNA. It is made up of eight segments ranging from approximately 840 to 2340bp, i.e. a total genome of approximately 13.7kb. These eight segments code for ten proteins [1]. The antigenicity of influenza is mainly due to hemagglutinin (HA) and neuraminidase (NA) respectively encoded by segments no. 4 and no. 6

Influenza is a major health problem. In temperate regions, influenza A generally causes annual epidemics during the winter, while in tropical regions Influenza circulates throughout the year. [2]–[4]. These epidemics are caused in particular by the H1N1 and H3N2 subtypes. Each year, the Influenza virus is responsible for 5 to 7 million severe cases of respiratory illness and 290,000 to 650,000 deaths [5]. In France in particular, Influenza is responsible for 2 to 6 million cases and on average 10,000 deaths per year [6].

One of the major difficulties associated with influenza is its capacity for rapid evolution and immune escape. This is due in particular to influenza's ability to make minor modifications to HA and NA viral epitopes (antigenic drift) and complete HA and NA segment exchanges between two influenza A viruses (antigenic shift: a type of reassortment) [1], [7]. In addition to antigenic shifts, influenza can also reassort with other segments. Reassortment can take place between two influenzas of different subtypes (H3N2/H1N1) or of the same subtype (exchange between clades) [1]. The significant animal reservoir of influenza further increases the appearance of new strains, particularly with the possibility of shifting between strictly animal strains and those circulating in humans [7].

Currently, the majority of influenza evolution studies are carried out on HA and NA, due to their important roles in antigenicity [8], [9]. Nowadays, with the rise of NGS especially during Covid-19 pandemic, Whole genome sequencing is becoming more and more accessible,

making it easy to study the evolution of a genome in its entirety. As HA and NA are the segments used in the annual construction of flu vaccines, the other segments are less studied, and their evolution and potential reassortment are less well known. Moreover, Influenza vaccines are often less effective against strains of subtype H3N2, due to the more frequent evolution of influenza A/H3N2 [8]. Influenza A/H3N2 is composed of more clades than Influenza A/H1N1, with 33 H3N2 clades circulating between 2017 and 2022, compared with 15 H1N1. It therefore seems more likely to detect intrasubtype reassortments in Influenzas A/H3N2.

In this study, it was decided to focus on the evolution of H3N2 influenza and particularly intra-subtype reassortment, from sample stored at -80C°, from January 2017 to October 2022, thus spanning a period before, during and after Covid-19 pandemic.

MATERIAL AND METHODS

RT-PCR detection:

Diagnoses of Influenza A virus infections are routinely carried out in the diagnostic laboratory of university and public hospitals of Marseille, southeastern France, using FTD Respiratory pathogens 21 (FTD) (Fast Track Diagnosis, Luxembourg), GeneXpert Flu/RSV (GeneXpert) (Cepheid, Sunnyvale, CA, USA) or BioFire FilmArray Respiratory panel 2 plus (BioFire) (Biomérieux, Marcy-l'Etoile, France), all of which target the M1 gene (plus two other genes for the GeneXpert test). The FTD allows detection of influenza A and B, and identification of A/H3, A/H1 and A/H1-2009, as well as obtaining Cts. The BioFire allows detection of influenza A and B and identification of the A/H1 variant, but does not allow acquisition of Cts. The GeneXpert can obtain Cts, but only detects influenza A and B, viruses without identifying any particular viral type.

Extraction method:

Viral RNA was extracted from respiratory samples using the EZ1 Virus Mini v2.0 kit

on an EZ1 Advanced XL instrument (Qiagen, Courtaboeuf, France) or the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's instructions.

Amplification method:

Amplification of influenza A RNA segments was performed using the primer system of method B written by P.K. Mitchell et al. [10]. This method uses two forward primers (MBTuni-12-GTTACGCGCCAGCAAAAGCAGG and MBTuni-12-SNV-GTTACGCGCCAGCGAAAGCAGG) and one reverse primer (MBTuni-13-GTTACGCGCCAGTAGAAACAAGG). This system is based on the optimization of a system originally designed in 2009 by Zhou et al. [11]. This system enables the complete sequencing of all 8 segments of human or animal influenza A. In this study, RT-PCR conditions were optimized to enable sequencing of human influenza samples with a Ct of 33 or less. Primers MBTuni-12 and MBTuni-12-SNV were used at 200nm each and MBTuni-13 at 400nm. The PCR cycle was performed with a first RT PCR step at 55°C for 30s, then 42°C for 50min and 94°C for 2min. This was followed by a pre-amplification step with 5 cycles of 94°C 15s, 42°C 30s and 68°C 3min. Amplification is carried out over 35 cycles of 94°C 15s, 42°C 30s, 68°C 3min with a final extension at 68°C for 5min.

Sequencing

Sequencing was carried out with Illumina technology (Mi-seq and Nova-seq) following the protocol used by Colson et al. (2021) [12] Sequence analysis was performed on CLC workbench v.7 software (QIAGEN). Mapping was performed using as reference the 8 influenza A H3N2 segments of the A/Michigan/173/2020(H3N2) sequence (MW796044.1 to MW796051.1), segment 4 of which belongs to clade 3C.2a1b.2a.2. Mapping was performed with a fraction length of 0.5 and a fraction similarity of 95%. Consensus sequences were extracted with a Threshold of 15. The resulting sequences were then analyzed using Nextstrain

CLI and Nextclade CLI pipelines on local instances [13], [14].

Sequence alignment was performed with MAFFT v7.505 [17] with the following parameters : thread = 8; threadtb = 5; threadit = 0; reorder; auto.

Phylogenetic analysis

Phylogenetic analysis with IQ-TREE v2.2.0 COVID-edition [18] .The parameters used are 1000 replicates for ultrafast bootstraps [19] and also 1000 replicates for SH approximate likelihood ratio test, the statistical method used for the substitution model is Neighbor-joining (Tamura and Nei, 1993) [20], the model used is Maximum Likelihood and the rate among site is the Uniform Rates (alrt = 1000 ; B = 1000 et m = TN+FO+I). The visualization and annotation of the trees was carried out with Itol v.6 [21], the annotation was also carried out in part with the drawing software inkscape v1.3.

Clades were assigned from segment 4, using Nextclade v2.14.1 (WEB version) with reference A/Wisconsin/67/2005 (CY163680) [13].

Reassortant detection

To detect reassortants, the alignments of each segment carried out by MAFFT were used with MrBayes software V.3.2.6 [22] in order to generate trees, 200,000 generations with a sampling frequency of 200. The trees generated with Mr.Bayes is used with GiRaF (Graph-incompatibility-based Reassortment Finder) software v1.02 [23]. This tool makes it possible to find reassortants in an automated manner by searching for phylogenetic incongruences. The default parameters have been kept for the burnin and the treashold (The confidence treshold for reporting a reassortment), only the cull (Fraction of tree in which the reassortant is found) has been changed to be set at 10%.

RESULTS

Quality of sequençing samples

In this study, 3475 samples between January 2017 and october 2022 were sequenced regardless of whether the subtype had been previously identified (Figure 1). Only the sequences of samples identified as H3N2 after sequence analysis were used for further analysis.

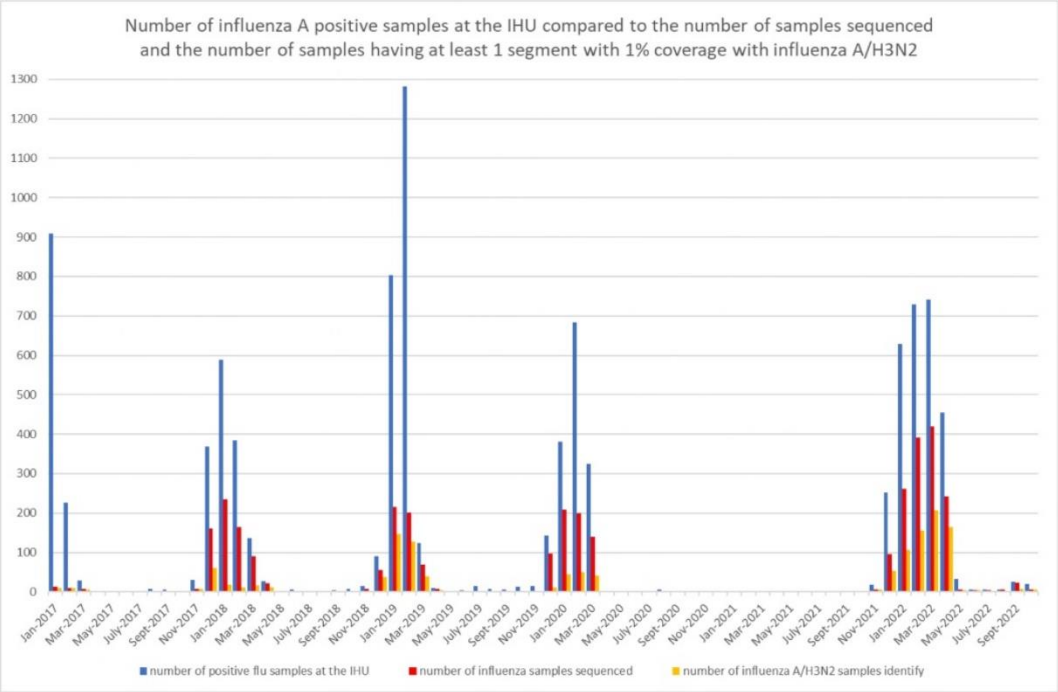


Figure 1. Number of Influenza A positive samples at the IHU compared to the number of samples sequenced and the number of samples having a least one segment with 1% coverage with influenza A/H3N2

Of the 3475 samples sequenced, 1285 provided at least one influenza A/H3N2 segment with at least 90% coverage (Table 1). Analysis of the 8 segments from these 1285 samples shows that coverage ranges from 2% to 99%, with a median strictly above 90% (Table 2).

A

Nb sequence by segment with 90% min coverage :								
Total	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8
1285	1023	850	1023	1084	1072	1097	1278	1197

B

Nb segment by sample with 90% min coverage :								
	1 Segment	2 Segment	3 Segment	4 Segment	5 Segment	6 Segment	7 Segment	8 Segment
	52	86	30	32	34	44	140	843

Table 1. Sequence with at least 90% coverage. A. Number of sequences with at least 90% coverage obtained for each segment. B. Number of samples with x segments have at least 90% coverage.

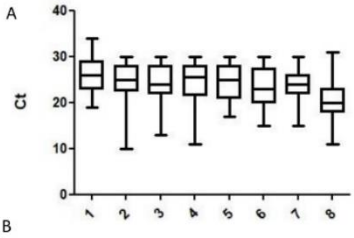
Table 2. Overall quality of sequences with at least 1% coverage.

Overall, 1024, 851, 1246, 1085, 1073, 1098, 1279 and 1198 sequences from segments 1 to 8 with at least 90% coverage were recovered. The best recovered segment is segment 7 and the least recovered is segment 2.

sequence by segment with 1% min coverage :								
	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8
Nb	1256	1187	1245	1228	1190	1211	1345	1282
Min coverage	0,02	0,02	0,02	0,03	0,04	0,04	0,05	0,06
Max coverage	0,99	0,99	0,99	0,99	0,98	0,99	0,91	0,97
Median	0,99	0,99	0,99	0,99	0,98	0,99	0,91	0,97
average coverage	0,87	0,83	0,88	0,91	0,92	0,94	0,89	0,94
Std. Deviation	0,27	0,28	0,26	0,23	0,21	0,18	0,11	0,15

From the samples with known Ct, the mean Ct of samples with eight segments was significantly lower than for samples with one to seven segments, the lower the Ct of the samples the greater the chance of having the eight segments with at least 90% coverage (Figure 2A, B and C). There is also a significant difference between the average Ct for six or seven segments compared with samples where only one segment is obtained. Of the 843 samples obtained with eight segments, 786 have a known Ct, ranging from 11 to 31 (mean 20.6, standard deviation 3.45) (Figure 2 A&B).

A



B

Number of segment	1	2	3	4	5	6	7	8
Number of values	66	77	27	32	31	42	125	786
Minimum	19	10	13	11	17	15	15	11
25% Percentile	23	22,5	22	21,5	21	20	22	18
Median	26	25	24	25,5	25	23	24	20
75% Percentile	29	28	28	28	28	27,25	26	23
Maximum	34	30	30	30	30	30	30	31
Mean	25,89	24,94	24,15	24,53	24,23	23,52	23,67	20,56
Std. Deviation	3,134	3,861	4,102	4,614	4,072	4,139	3,364	3,453
Std. Error	0,386	0,440	0,789	0,816	0,731	0,639	0,301	0,123
Lower 95% CI	25,12	24,06	22,53	22,87	22,73	22,23	23,08	20,32
Upper 95% CI	26,66	25,81	25,77	26,19	25,72	24,81	24,27	20,8

C

		1	2	3	4	5	6	7	8
1	Mean Difference	---	0.958	1.746	1.363	1.668	2.370*	2.222***	5.335***
	P-value	---	0.726	0.503	0.798	0.481	0.043	< .001	< .001
2	Mean Difference		---	0.787	0.404	0.709	1.411	1.263	4.377***
	P-value		---	0.987	1,000	0.99	0.609	0.264	< .001
3	Mean Difference			---	-0.383	-0.078	0.624	0.4761	3.59**
	P-value			---	1,000	1,000	0.999	0.999	< .001
4	Mean Difference				---	0.305	1.007	0.859	3.973***
	P-value				---	1,000	0.977	0.974	< .001
5	Mean Difference					---	0.702	0.554	3.667***
	P-value					---	0.996	0.997	< .001
6	Mean Difference						---	-0.148	2.965***
	P-value						---	1,000	< .001
7	Mean Difference							---	3.113***
	P-value							---	<.001
8	Mean Difference								---
	P-value								---

Note. * p < .05, ** p < .01, *** p < .001

153 Figure 2. Analysis of complete sequences. A. Boxplot representation of the average Ct per
154 sample having between one and eight sequences with at least 90% coverage. B. Statistical
155 description of the Ct of groups having between one and eight sequences with at least 90%
156 coverage. C. Post-hoc Games-Howels test, analysis of the correlation between the Ct and the
157 number of segments obtained.

158 ***1.Sequence analysis***

159 From the sequences of the 843 samples with eight segments, segments 4 and 6 have the
160 most unique mutations (present at least once in a genome) per 100bp, with 38.37 and 36.06
161 nucleotide mutations per 100bp respectively (Table 3). Segments 5 and 7 have the fewest single
162 mutations per 100bp, with 22.30 and 25.67 mutations per 100bp respectively. The majority of
163 mutations are present in less than 10% of sequences.

	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6	Segment 7	Segment 8	Whole genome
Reference segment size	2341	2341	2233	1762	1566	1467	1027	890	13627
NT unique mutation number (Total)	751	727	715	676	402	529	229	258	4287
Unique mutation number (Total) per 100 bp	32,08	31,06	32,02	38,37	25,67	36,06	22,30	28,99	31,46
NT unique mutation number (>90%)	53	65	58	63	30	50	13	12	344
NT unique mutation number (10=< and >90%)	41	41	41	63	32	24	14	16	272
NT unique mutation number (<10%)	657	645	652	625	689	677	724	723	5392
AA unique mutation number (Totale)	200	165	300	255	70	201	63	147	1401
AA unique mutation number (>90%)	6	6	18	20	2	24	2	6	84
AA unique mutation number (10=< and >90%)	5	3	7	32	4	12	1	7	71
AA unique mutation number (<10%)	740	742	726	699	745	715	748	738	5853

164 Table 3. Analysis of NT and AA mutations in 843 complete influenza genomes.

165 Segment 4 coding for HA is the segment with the most mutations on average in
166 nucleotides (104.5 ± 1.5) and amino acids (38.91 ± 0.41) (Tab.4.A&B). Segment 8, coding for
167 2 NS1 and NEP genes, has the fewest nucleotide mutations on average (22.0 ± 0.2), while
168 segment 7, coding for the M1 and M2 genes, has the fewest amino acid mutations on average
169 (2.90 ± 0.04).

170 Table 4. Descriptive statistics of the number of mutations per NT (A) and AA (B) of the eight
171 segments of the 843 complete influenza genomes

Statistiques descriptives

	Moyenne	Intervalle de confiance à 95%		Médiane	Ecart-type	Minimum	Maximum
		Borne inf	Borne sup				
Nb mutation NT seg1	81.6	81.1	82.0	83.0	6.53	60.0	101.0
Nb mutation NT seg2	91.5	90.9	92.0	95.0	8.11	71.0	105.0
Nb mutation NT seg3	80.4	80.1	80.6	81.0	3.75	70.0	91.0
Nb mutation NT seg4	104.5	103.5	105.5	115.0	14.44	77.0	125.0
Nb mutation NT seg5	44.5	44.1	44.9	46.0	5.50	30.0	54.0
Nb mutation NT seg6	67.9	67.6	68.3	69.0	4.95	52.0	78.0
Nb mutation NT seg7	23.0	22.7	23.4	26.0	4.52	15.0	30.0
Nb mutation NT seg8	22.0	21.8	22.2	22.0	2.94	15.0	31.0

Note. The CI of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

Statistiques descriptives

	Moyenne	Intervalle de confiance à 95%		Médiane	Ecart-type	Minimum	Maximum
		Borne inf	Borne sup				
Nb mutation AA seg1	9.06	8.97	9.15	9.00	1.353	1.00	12.00
Nb mutation AA seg2	8.06	7.93	8.18	9.00	1.827	1.00	13.00
Nb mutation AA seg3	21.69	21.50	21.89	23.00	2.852	3.00	27.00
Nb mutation AA seg4	38.91	38.50	39.33	43.00	6.121	27.00	48.00
Nb mutation AA seg5	4.87	4.75	4.98	6.00	1.672	2.00	10.00
Nb mutation AA seg6	31.49	31.32	31.65	32.00	2.433	24.00	37.00
Nb mutation AA seg7	2.90	2.86	2.95	3.00	0.647	2.00	6.00
Nb mutation AA seg8	10.73	10.64	10.83	11.00	1.411	7.00	16.00

Note. The CI of the mean assumes sample means follow a t-distribution with N - 1 degrees of freedom

Segment 8 had the lowest ratio of nucleotide to amino acid mutations (2.05), while segment 2 had the highest ratio (11.35) (Table 4B). Almost half of the nucleotide mutations in segment 8 are non-synonymous mutations, while less than 1/10 of the nucleotide mutations in segment 2 are non-synonymous mutations.

Clade and sub-clade of sequençing samples

Influenza A/H3N2 is subdivided into 33 clades and subclades (gisaid/nextclade) assigned from the hemagglutinin gene (segment 4). According to gisaid data, between 08-2018 and 10-2022, 31 clades and sub-clades circulated worldwide. Over the same period at IHU 20 clades were detected with the 843 complete samples (Figure3) and 21 since 2017 (Table5 & Figure3).

H3N2 (Total)	01-2017 to 10-2022
3C.2a1	7
3C.2a1a	7
3C.2a1b.1	79
3C.2a1b.1a	35
3C.2a1b.1b	9
3C.2a1b.2	43
3C.2a1b.2a.2	23
3C.2a1b.2a.2a.1	386
3C.2a1b.2a.2a.1a	2
3C.2a1b.2a.2a.1b	2
3C.2a1b.2a.2a.2	1
3C.2a1b.2a.2a.3	16
3C.2a1b.2a.2a.3a	10
3C.2a1b.2a.2a.3b	8
3C.2a1b.2a.2b	4
3C.2a1b.2a.2c	16
3C.2a1b.2b	13
3C.2a2	51
3C.2a3	69
3C.2a4	1
3C.3a1	61

Table 5. Number of samples per clade detected between January 2017 and October 2022

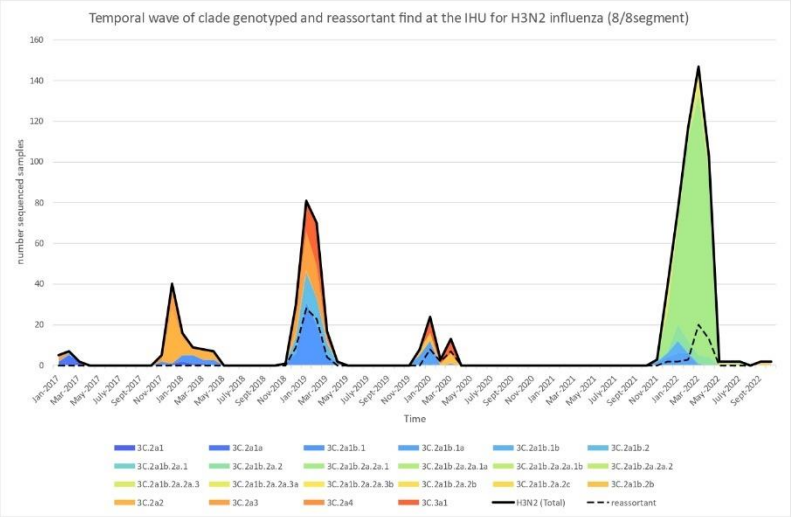


Figure 3. Clade of Influenza A virus detected between January 2017 and October 2022. The dotted curve represents the number of total reassortants detected.

Clade and sub-clade circulation

Clade 3C.2a1a, a subclade of calde 3C.2a1 (3C2a1 + HA1 :N121K + HA1 :T135K + HA2 :G150E) was detected at the IHU only in January and February 2017 for 7 samples, all of

189 which had the HA1 N121K and T135K mutations and the HA2 G150E mutation [24] (Figure
190 3).

191 Over the 2017-2018 epidemic season at the IHU (November 2017 to April 2018), clade
192 3C.2a2 (54%) was in the majority, followed by clades 3C.2a3 (22%) 3C.2a1b.1 (16%), 3C.2a1
193 (6%) and 3C.2a4 (1%) (Figure 3). Sub-clades 3C.2a2, 3C.2a3 and 3C.2a4 were not defined in
194 2017-2018 and were grouped under clade 3C.2a, as was clade 3C.2a1b.1, which was grouped
195 under clade 3C.2a1.

196 In 2018-2019 (November 2018 to April 2019) several clades circulated at similar rates
197 at IHU: 3C.2a1b.1 (32%), 3C.2a3 (23%), 3C.3a1 (22%) and 3C.2a1b.2 (21%). The 3C.2a2
198 clade was detected in 2% of cases (Figure 3). At that time, the sub-clades 3C.2a1b.1 and
199 3C.2a1b.2 were not defined and grouped together under the 3C.2a1b clade.

200 In 2019-2020 (December to March 2020), clades 3C.2a1b.1a (35%) and 3C.3a1 (35%)
201 were in the majority at IHU, followed by clades 3C.2a1b.2b (27%) and 3C.2a1b.1 (1%) (Figure
202 3). At that time, sub-clade 3C.3a1 was integrated with clade 3C.3a, clades 3C.2a1b.1a and
203 3C.2a1b.2b with clade 3C.2a1b.

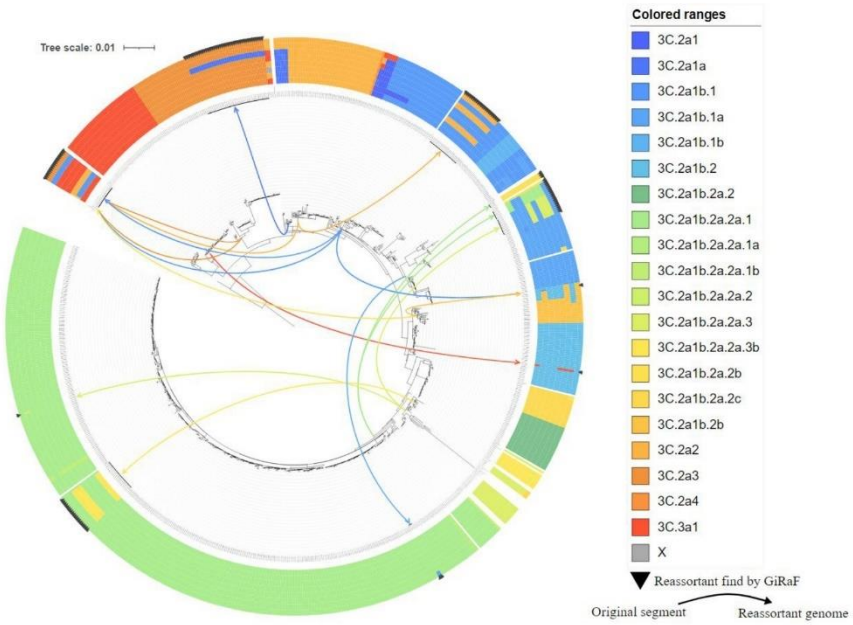
204 In 2021-2022 at the IHU (November 2021 to April 2022), clade 3C.2a1b.2a.2a.1 was
205 the most detected (79%), followed by clades 3C.2a1b.2a.2 (5%), 3C.2a1b.1a (4%),
206 3C.2a1b.2a.2a.3 (3%), 3C.2a1b.2a.2c (3%), 3C.2a1b.1b (2%), 3C.2a1b.2a.2a.3a (2%),
207 3C.2a1b.2a.2a.3b (1%) (Figure 3). At that time, the sub-clades 3C.2a1b.2a.2a.1,
208 3C.2a1b.2a.2a.3, 3C.2a1b.2a.2a.3a, 3C.2a1b.2a.2a.3b, 3C.2a1b.2a.2c were grouped together
209 under the clade 3C.2a1b.2a.2.

210 ***2. Phylogeny and intrasubtype-reassortant***The search for reassortants with GiRaF
211 made it possible to find 126 reassortants or 14.95% (126/843) of the analyzed genomes between
212 2017 and 2022. In this study no reassortants were found at the beginning of 2017 and during
213 the 2017-2018 epidemic. During the 2018-2019 and 2019-2020 epidemics, respectively 31.84%

and 35.42% of the sequential genomes were reassortants. In 2021-2022, 8.89% of genomes were reassortants.

Based on the eight segments of the 843 complete samples, a phylogeny was performed for each segment (Supplementary 1).

The concatenation of the eight segments makes it possible to create a phylogenetic tree



showing where the reassortants are placed in the phylogeny (Figure 4). The origin of the reassorted segments of the reassortant sequences detected by Giraf was carried out from the manual annotation of the trees of each segment from the clades identified by nextclade on segment 4 and from the tree of the eight concatenate segments. The reassortants found by GiRaF concern seven different clades: 3C.2a1b.1, 3C.2a1b.2, 3C.2a1b.2a.2a.1, 3C.2a1b.2a.2a.3, 3C.2a1b.2a.2b, 3C.2a3 and 3C.3a1.Figure 4. Phylogeny of the concatenation of the 8 segments of the 843 complete influenza genomes

Of the 843 complete genomes, 79 belong to the 3C.2a1b.1 clade, which has been detected at the IHU during three epidemic periods: 2017-2018, 2018-2019 and 2019-2020 (Figure 5). Twenty two genomes from this clade are probable reassortants according to GiRaF

and were sampled between January and March 2019. All 3C.2a1b.1 reassortants were found during the epidemic 2018-2019 and represented 34.38% (22/64) of the sequences of this clade obtained during this period, or 27.85% (22/79) of all the genomes of this clade detected at the IHU.

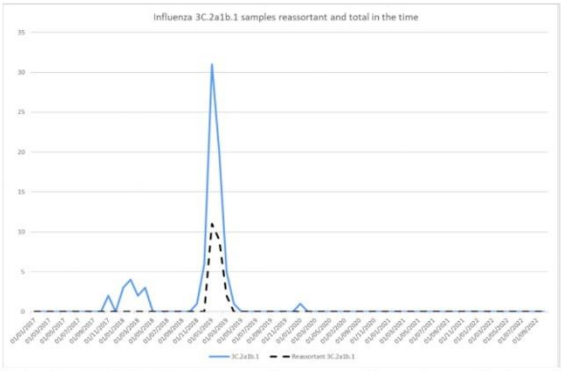


Figure 5. number of clades 3C.2a1b.1 samples as a function of time (solid line: total number, dotted line number of reassortants)

On the 43 genomes obtained from the 3C.2a1b.2 clade, 42 were detected over the 2018-2019 year and one in January 2017 (Figure 6). Two reassortants were detected among this clade, the first in February and the second in March 2019, representing 4.65% (2/43) of the clade's members.

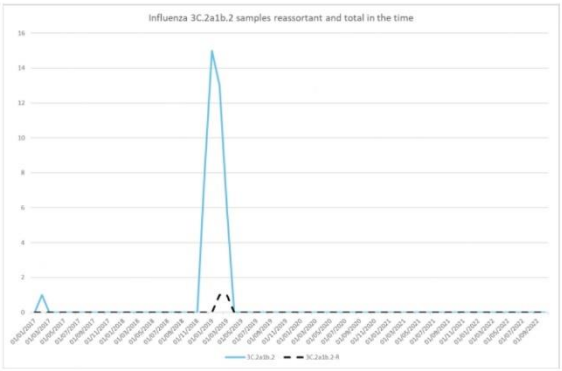


Figure 6. number of clades 3C.2a1b.2 samples as a function of time (solid line: total number, dotted line number of reassortants)

The 3C.2a1b.2a.2a.1 clade, attributed to 386 genomes, appeared at the IHU during the 2021-2022 epidemic, and was first detected in a sample from December 2021 (Figure 7). The

first reassortant of this clade was found on a sample in February 2022. In all, 7.51% (29/386) of this clade's members were detected as reassortants.

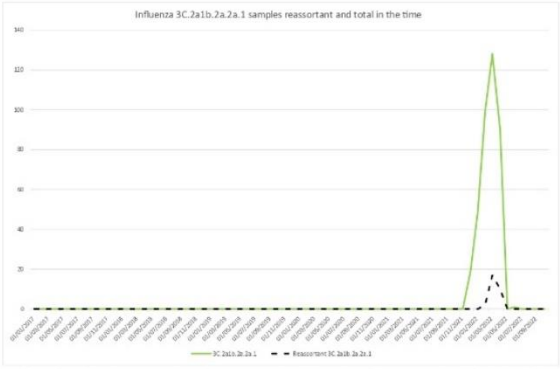


Figure 7. number of clades 3C.2a1b.2a.2a.1 samples as a function of time (solid line: total number, dotted line number of reassortants)

The 3C.2a1b.2a.2a.3 clade emerged during the same epidemic period as the 3C.2a1b.2a.2a.1 clade. Its first occurrence at IHU was in November 2021, and the first reassortant in December 2022 (Figure 8). Overall, 68.75% (11/16) of the sequences obtained from this clade would be reassortant.

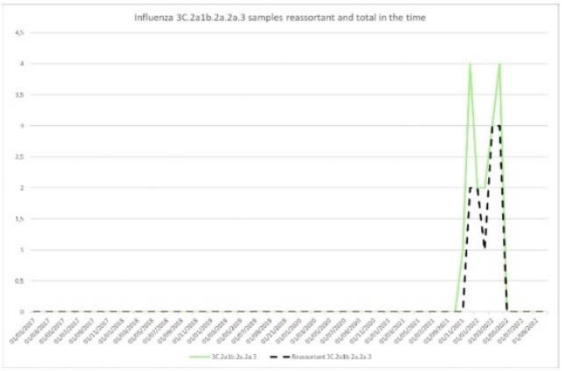


Figure 8. number of clades 3C.2a1b.2a.2a.3 samples as a function of time (solid line: total number, dotted line number of reassortants)

Only 4 genomes of clade 3C.2a1b.2a.2b were found, this clade was found only in October and November 2022 and would all be reassortant (Figure 9).

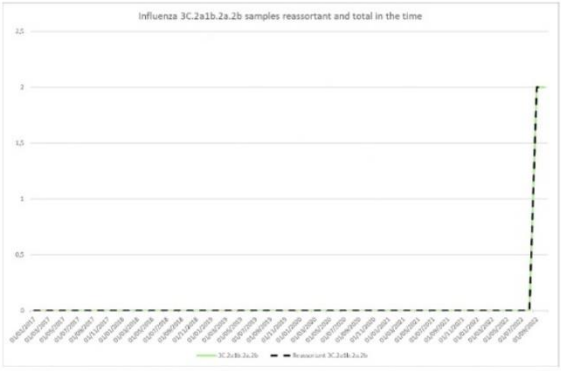


Figure 9. number of clades 3C.2a1b.2a.2b samples as a function of time (solid line: total number, dotted line number of reassortants)

Of the genomes analyzed here, 69 belong to clade 3C.2a3 and were detected in early 2017 and during the 2017-2018 and 2018-2019 epidemics. Of these, 40 are reassortant, i.e. 80.67% (40/69) (Figure 10). All the reassortants in this clade were detected during the 2018-2019 period, during which time reassortant and non-reassortant or both were detected from December 2018 onwards. The reassortants represented 86.96% of the genomes in this clade during this period.

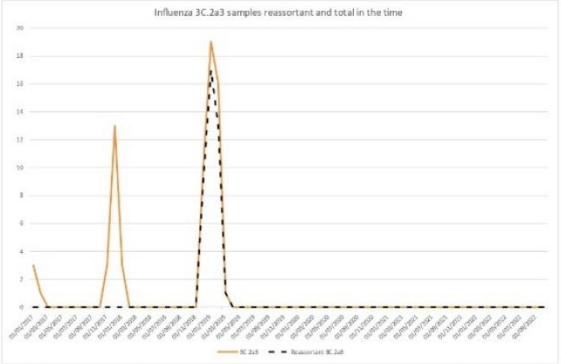


Figure 10. number of clades 3C.2a3 samples as a function of time (solid line: total number, dotted line number of reassortants)

The 3C.3a1 clade was detected at IHU during the 2018-2019 and 2019-2020 epidemics (Figure11). 61 of the 843 genomes in this study belong to this clade, and 17 of these are considered reassortants by GiRaF. The 17 reassortants were all detected during the 2019-2020 epidemic and represent 100% (17/17) of the 3C.3a1 influenzas in this period and 27.87% (17/61) of the genomes in this clade.

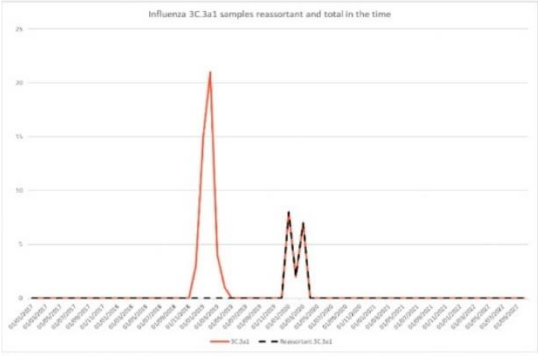


Figure 11. number of clades 3C.3a1 samples as a function of time (solid line: total number, dotted line number of reassortants)

The 126 reassortants detected by GiRaF can be separated into thirteen groups with reassortants of similar origins (Figure 4). Groups no. 1 and no. 2 are both part of clade 3C.2a1 (Table 6). Four of their reassorted segments. Segments 8 of these two groups come from clade 3C.2a3 and their segments 7 from clade 3C.2a1b.1. Segments 2 and 3 of group no. 1 come from clade 3C.2a1b.2b and for group no. 2 they come from clades 3C.2a1b.1 and 3C.2a2 respectively. Group 3 is part of clade 3C.2a3. Only segment 6 reassorts and comes from clade 3C.2a1a. Group no. 4 is part of clade 3C.2a1b.1. Segments 3, 6 and 8 reassort with clade 3C.2a2. Groups no. 5 and no. 6 are part of clade 3C.2a1b.2a.2b. For group no. 5 it is difficult to know which segments are reassorting and with whom they are reassorting. Segment 2 of group no. 6 reassorts with clade 3C.2a1b.2a.2a. Groups no. 7, no. 11, no. 12 and no. 13 are part of the 3C.2a1b.2a.2a clade. Segments 8 of groups no. 7 and no. 11 reassort with clade 3C.2a1b.1a, group no. 7 also carries out an exchange with the same clade at segments 2 and 3. Group no. 12 reassorts at level of segments 1, 5 and 6 with clade 3C.2a1b.2a.2a.3b. Group No. 13 reassociates

with clade 3C.2a1b.2a.2a.3 at segment 7. Group No. 8 is part of clade 3C.2a1b.2a.2a and reassorts at segments 1, 2, 3, 5 and 8 with clade 3C.2a1b.1. Groups no. 9 and no. 10 are part of clade 3C.2a1b.2. Group no. 9 reassorts at segment 2 with clade 3C.2a1b.1 and with clade 3C.2a1b.2b with segments 6 and 8. Group no. 10 reassorts at segments 1, 5, 6 and 7 with clade 3C.3a1.

The reassortant clades with groups no. 4 and 6 to 13 are present in the temporality as the original clade. Segments 7 and 8 of groups no. 1 and 2 and segments 2 and 3 of group no. 2 reassort with clades not (3C.2a3 and 3C.2a2) or little (3C.2a1b.1) detected on the year the reassortant are present. Segments 2 and 3 of group no. 1 reassort with a clade present in the same temporality. For group no. 3, reassortments took place in 2019 with one clade only present in 2017.

DISCUSSION

Few studies on intra-subtype reassortants of A/H3N2 influenza in humans have been carried out and no study on intra-subtype reassortants in the pre- and post-covid period in France has been carried out, whereas SARS-CoV-2 epidemic has had an impact on influenza epidemics across the world. The strength of this study lies in the fact that it was carried out with a large panel of good quality and complete genomes, over a period extending from 2017 to 2022.

It should be noted that the clades assigned during this study were assigned in 2023, with all clades existing in 2023. After an epidemic, it is possible for new subclades to be defined, and therefore for a sample labeled in one clade at the time of the epidemic to be subsequently labeled in a descending subclade.

Clade 3C.2a1a appears at first glance to be absent from the Gisaid database, but this seems to be a difference in classification, linked to the use of Nextclade WEB. On Nextclade WEB, when using reference A/Darwin/6/2021 (EPI1857216), samples classified as 3C.2a1a with reference A/Wisconsin/67/2005 (CY163680) end up classified as 3C.2a1. The same problem occurs with the 3C.2a4 samples, which are assigned 3C.2a.

During the 2017-2018 epidemic season, according to Rondy et al. it was the 3C.2a (63%) and 3C.2a1 (35%) subclades that were in the majority in Europe, while at the IHU the rates of 3C.2a were higher and those of 3C.2a1 lower than for the rest of Europe [25].

In 2018-2019, the rate of sub-clade 3C.2a3 is very different between the IHU and the rest of Europe (23 and 1% respectively) [26].

During the 2019-2020 season, clade 3C.2a1b was in the majority at the IHU (63%), its prevalence being higher than for the rest of Europe (45%). In Marseille, clade 3C.3a1 was secondary to the WHO european region result (35 vs. 55%) [27].

In 2020-2021, no samples have been diagnosed [28], therefore sequenced, at the IHU over this period, as few cases of influenza have been found in Europe or elsewhere in the world [29], [30]. During this period, the 3C.2a1b and 3C.2a subclades circulated with a predominance for the 3C.2a1b subclade [31].

During 2021-2022, according to the WHO European region, the 3C.2a1b.2a.2 clade was the one majoritary, with 98% of viruses characterized belonging to this clade, close to the results obtained here [32].

New clades and subclades appeared and disappeared over time. None of the Clades present in 2017 were still present in 2022. Some of the Clades evolved to make way for a new lineage, while others disappeared without giving rise to a new lineage. Clades 3C.2a2, 3C.2a3, 3C.2a4 and 3C.3a1 (first and last detected in December 2017 and March 2019, January 2017 and March 2019, March 2018 and March 2018, December 2018 and March 2020) have all disappeared at IHU and are no longer detected. These Clades have also stopped circulating in Europe since the 2019-2020 epidemic period, according to ecdc reports [33]–[36]. Only subclades descending from Clade 3C.2a1 are still circulating.

Phylogenetic analysis of the eight segments individually shows that the topology between the trees varies with or without reassortment. The rate of mutations is different from one segment to another.

Influenza A are known to carry out intra-subtype reassortments. For H3N2 influenza, the reassortment rate varies greatly between studies. The latter varies depending on the epidemics and the regions studied [37]–[41]. With analysis methods similar to that used here, rates ranging from 6.7% in Scotland between 2014 and 2015 [38] to 64.1% in Togo between 1994 and 2020 [39]. A study by Berry et al. on more than 2000 H3N2 influenza sequences across the world between 2008 and 2014 shows a reassortant rate of 41.3% with strong variation between regions of the globe (less than 15% reassortant in North America, more than 75 % in South America) [37].

In the case of the H3N2 sequences analyzed here, 14.95% of the 843 genomes were reassortant, with a variation from 0 to 35.42% depending on the epidemic seasons. This variability can in particular be explained by the fact that from one epidemic season to another the circulating clades are not the same, reassortments can only take place if there is co-infection of a cell by the two subclades of flu. Furthermore, the probability of intrasubtype reassortment between two influenzas would depend on the subclade [42] and in particular on the packaging signals allowing the eight segment to regroup in the viral particle [43].

According to the data obtained in this study, reassortant genomes generally appear in a second phase. In the case of reassortant clades: 3C.2a1b.1, 3C.2a3 and 3C.3a1, reassortant genomes appear during the second epidemic period of their presence at the IHU. For reassortant clades present only during a single epidemic period: 3C.2a1b.2, 3C.2a1b.2a.2a.1 and 3C.2a1b.2a.2a.3 reassortants appear between two and three months after the start of the epidemic. For clade 3C.2a1b.2a.2b, reassortants were detected in all representatives of this

clade, as soon as this clade was detected at the IHU, but only four genomes of this clade could be studied, limiting the data.

For the clades that emerged over several epidemic periods (3C.2a1b.1, 3C.2a3 and 3C.3a1) it is not possible to know when and where reassortment took place. These clades appeared and circulated only during the pre-covid period, so fewer samples from this period were obtained and sequenced as part of this study, and therefore fewer genomes were obtained. It is therefore difficult to know whether reassortments were already present during the first epidemic period of their circulation at the IHU, but in a minority manner and would then have escaped our study, or whether reassortments appeared in another region of the world. Notably in the case of clades 3C.2a3 and 3C.3a1, 86.96 to 100% of which were reassortants over the period when the reassortants were identified and appeared at the start of the epidemic, whereas in the previous year no reassortant was detected despite the fact that the clade was already circulating. For clade 3C.2a1b.1, it is possible that reassortants appeared during the 2018-2019 epidemic period, since their first occurrences took place two months after the clade's first appearance during this epidemic.

In all thirteen similar reassortment groups have been detected, these groups vary from year to year depending on the clade that is circulating. Furthermore, reassortment groups are never presented over several epidemic periods. As explained previously for those appearing pre-covid it is difficult to know when and where the segment exchange was carried out. For groups no. 1 and 2 it seems that the reassortment did not for the most part take place during the epidemic period of their appearance and not in Europe, because the last one took place with clades which did not circulate during this period in Europe. The 3C.2a2 and 3C.2a3 clade were no longer circulating in Europe in 2019-2020 [27]. For group no. 3 which reassorts with the 3C.2a1a clade, it is possible but unlikely that the segment exchange took place in Europe during the 2018-2019 epidemic period because the 3C.2a1a clade was still detected in Europe, but in

less than 1% of cases [26]. For groups no. 5 and 6, the GiRaF analysis detects it as reassortant, but it is more difficult to know with which clade they reassort. Notably because all the samples from these groups belong to the 3C.2a1b.2a.2b clade which is composed of only reassortants and is very mobile in the phylogenies of the eight segments.

Currently, most studies of influenza evolution and characterization are based on segment 4. In particular, the ecdc uses segments 4 to describe a subclade. In the case of ecdc, some of the strains used do not have a complete genome and only their segment 4 is used in the reports. For example, the sequence A/Pakistan/GIHSN/ICT/969/2023 used in the trees for clade 3C.2a1b.2a.2a.3b is made up of 7 H3N2 influenza segments and one segment (segment 5) of H1N1 influenza [9]. Focused the study of the influenza genome on segment 4 and segment 6, limits our vision of the impact of intra-subtype reassortment from one epidemic to the next.

CONCLUSION

Studies on reassortment among all segments of A/H3N2 influenza viruses are limited. This is the first study to examine infra-subtype reassortment of A/H3N2 influenza in France, particularly during the pre- and post-COVID periods. In this project, we found that over the studied period and location, the reassortment rate was 15%, with significant variation across years. This study highlights the lack of information currently available on the role of reassortment among segments other than HA and NA in the emergence of new genotypes. Moving forward, it would be valuable to investigate whether reassortments affect transmissibility, infectivity, and immune escape to better understand their impact on the population and the potential for future epidemics and pandemics.

Institutional Review Board Statement

The present study has been approved by the ethics committee of the University Hospital Institute (IHU) Méditerranée Infection (N°2022-021) and was registered in the General Data Protection Regulation (GDPR) registry no. ZJ6W5X from Assistance Publique-Hôpitaux de Marseille (AP-HM, Marseille public hospitals) institution.

Informed Consent Statement

The present study has been conducted on influenza A virus genome sequences obtained from respiratory samples sent to our clinical diagnostic laboratory for the purpose of diagnosis of respiratory viruses. No supplementary clinical sample was used that was not sent to our laboratory in this setting. All clinical samples and data have been fully anonymized.

Data Availability Statement

Viral genomes analyzed in the present study have been deposited in the NCBI GenBank sequence database under Accession nos. (Not already yet)

Conflicts of Interest

Lorlane Le Targa is an employee of Biosellal (Dardilly, France). The other authors have no conflict of interest to declare relative to the present study.

BIBLIOGRAPHY

[1] W. Shao, X. Li, M. U. Goraya, S. Wang, and J.-L. Chen, "Evolution of Influenza A Virus by Mutation and Re-Assortment," *Int J Mol Sci*, vol. 18, no. 8, p. 1650, Aug. 2017, doi: 10.3390/ijms18081650.

[2] F. Krammer *et al.*, "Influenza," *Nat Rev Dis Primers*, vol. 4, no. 1, p. 3, 2018, doi: 10.1038/s41572-018-0002-y.

[3] R. L. Hodinka, "Respiratory RNA Viruses," *Microbiol Spectr*, 2016, doi: 10.1128/microbiolspec.DMIH2-0028-2016.

[4] J. M. Galván, O. Rajas, and J. Aspa, "Review of Non-bacterial Infections in Respiratory Medicine: Viral Pneumonia," *Arch Bronconeumol*, vol. 51, no. 11, pp. 590–597, Nov. 2015, doi: 10.1016/j.arbr.2015.09.015.

[5] "Influenza (Seasonal)." Accessed: Oct. 13, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[6] "La grippe, une épidémie saisonnière." Accessed: Nov. 02, 2022. [Online]. Available: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/grippe/la-grippe-une-epidemie-saisonniere

[7] H. Kim, R. G. Webster, and R. J. Webby, "Influenza Virus: Dealing with a Drifting and Shifting Pathogen," *Viral Immunology*, vol. 31, no. 2, pp. 174–183, Mar. 2018, doi: 10.1089/vim.2017.0141.

[8] M. Hayati, B. Sobkowiak, J. E. Stockdale, and C. Colijn, "Phylogenetic identification of influenza virus candidates for seasonal vaccines," *Sci Adv*, vol. 9, no. 44, p. eabp9185, doi: 10.1126/sciadv.abp9185.

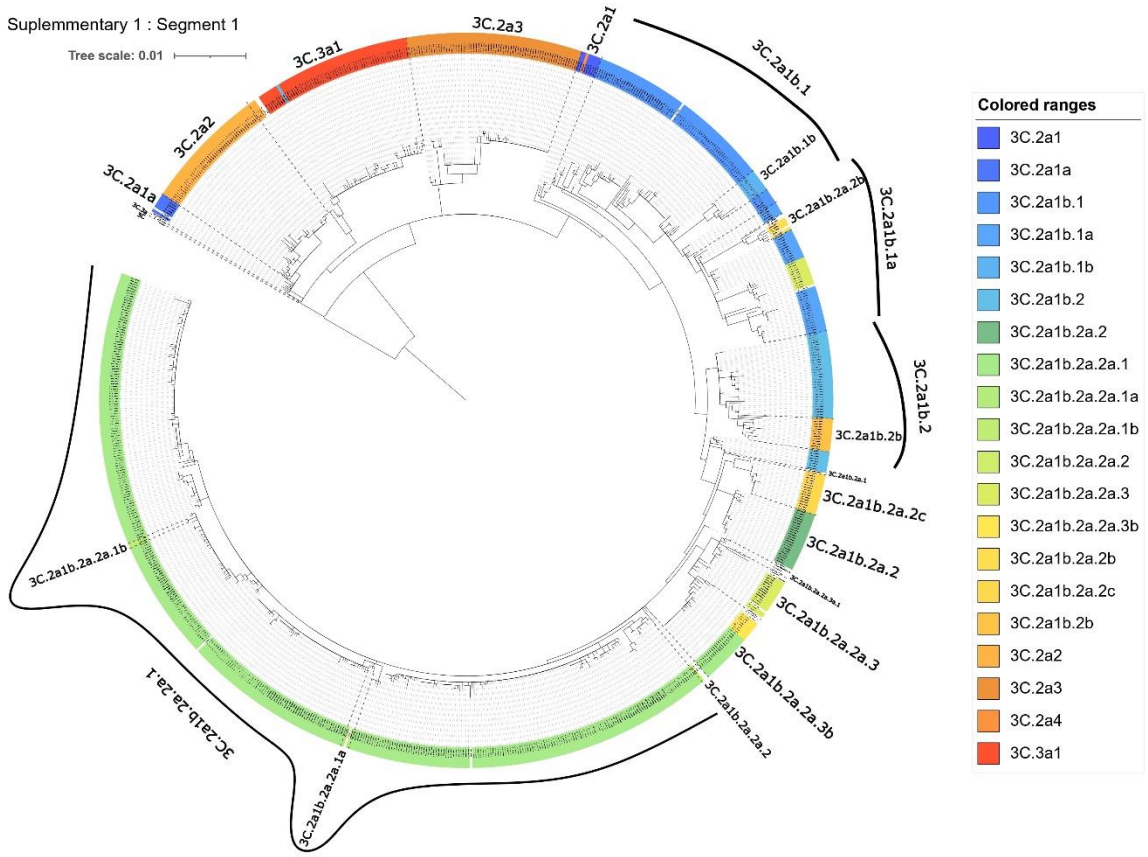
[9] "Influenza characterisation February 2023", [Online]. Available: https://www.ecdc.europa.eu/sites/default/files/documents/Influenza-characterisation-February-2023.pdf

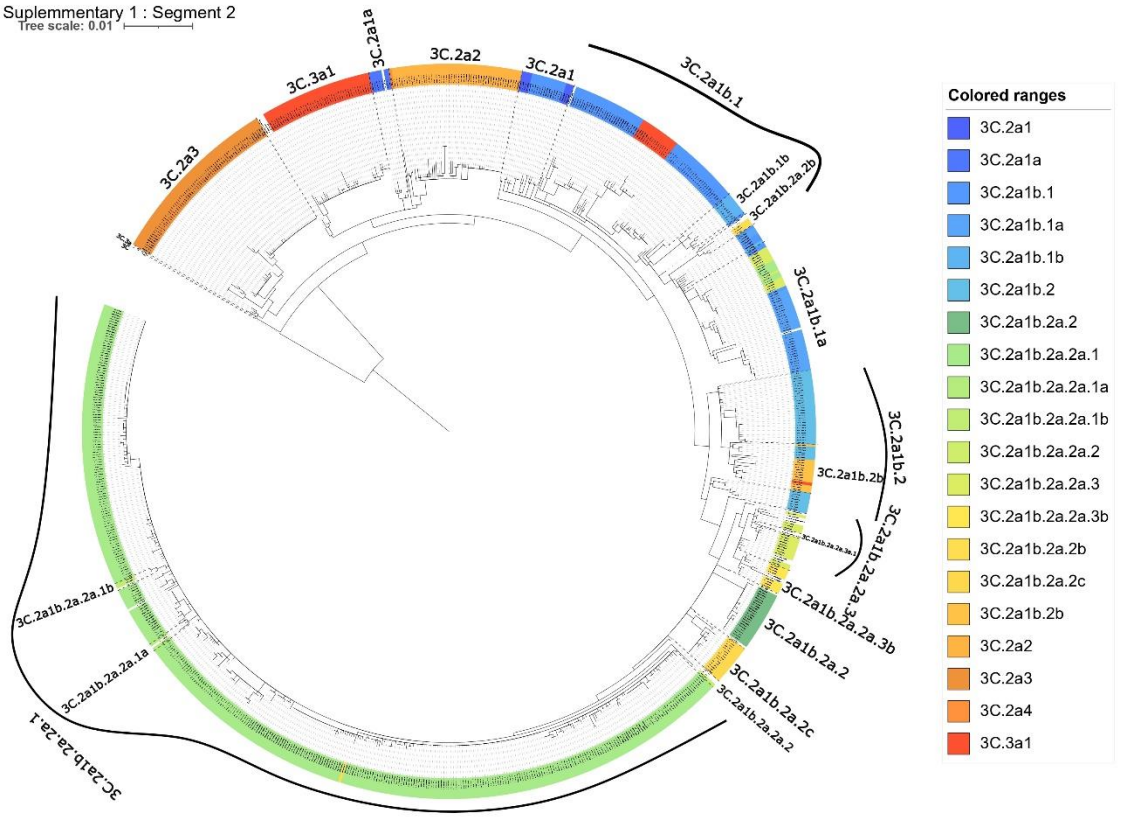
[10] P. K. Mitchell *et al.*, "Method comparison of targeted influenza A virus typing and whole-genome sequencing from respiratory specimens of companion animals," *J Vet Diagn Invest*, vol. 33, no. 2, pp. 191–201, Mar. 2021, doi: 10.1177/1040638720933875.
[11] B. Zhou *et al.*, "Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses," *J Virol*, vol. 83, no. 19, pp. 10309–10313, Oct. 2009, doi: 10.1128/JVI.01109-09.
[12] P. Colson *et al.*, "Analysis of SARS-CoV-2 Variants From 24,181 Patients Exemplifies the Role of Globalization and Zoonosis in Pandemics," *Front Microbiol*, vol. 12, p. 786233, Feb. 2022, doi: 10.3389/fmicb.2021.786233.
[13] I. Aksamentov, C. Roemer, E. Hodcroft, and R. Neher, "Nextclade: clade assignment, mutation calling and quality control for viral genomes," *JOSS*, vol. 6, no. 67, p. 3773, Nov. 2021, doi: 10.21105/joss.03773.
[14] J. Hadfield *et al.*, "Nextstrain: real-time tracking of pathogen evolution," *Bioinformatics*, vol. 34, no. 23, pp. 4121–4123, Dec. 2018, doi: 10.1093/bioinformatics/bty407.
[15] "jamovi - open statistical software for the desktop and cloud." Accessed: Nov. 07, 2023. [Online]. Available: https://www.jamovi.org/
[16] "The Comprehensive R Archive Network." Accessed: Nov. 07, 2023. [Online]. Available: https://cran.r-project.org/
[17] K. Katoh and D. M. Standley, "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, Apr. 2013, doi: 10.1093/molbev/mst010.
[18] B. Q. Minh *et al.*, "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era," *Molecular Biology and Evolution*, vol. 37, no. 5, pp. 1530–1534, May 2020, doi: 10.1093/molbev/msaa015.
[19] D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh, "UFBoot2: Improving the Ultrafast Bootstrap Approximation," *Molecular Biology and Evolution*, vol. 35, no. 2, pp. 518–522, Feb. 2018, doi: 10.1093/molbev/msx281.
[20] K. Tamura and M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.," *Molecular Biology and Evolution*, vol. 10, no. 3, pp. 512–526, May 1993, doi: 10.1093/oxfordjournals.molbev.a040023.
[21] I. Letunic and P. Bork, "Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation," *Nucleic Acids Research*, vol. 49, no. W1, pp. W293–W296, Jul. 2021, doi: 10.1093/nar/gkab301.
[22] F. Ronquist *et al.*, "MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space," *Syst Biol*, vol. 61, no. 3, pp. 539–542, May 2012, doi: 10.1093/sysbio/sys029.
[23] N. Nagarajan and C. Kingsford, "GiRaF: robust, computational identification of influenza reassortments via graph mining," *Nucleic Acids Research*, vol. 39, no. 6, p. e34, Mar. 2011, doi: 10.1093/nar/gkq1232.
[24] D. M. Skowronski *et al.*, "Influenza Vaccine Effectiveness by A(H3N2) Phylogenetic Subcluster and Prior Vaccination History: 2016–2017 and 2017–2018 Epidemics in Canada," *J Infect Dis*, vol. 225, no. 8, pp. 1387–1398, Mar. 2020, doi: 10.1093/infdis/jiaa138.
[25] M. Rondy *et al.*, "Interim 2017/18 influenza seasonal vaccine effectiveness: combined results from five European studies," *Euro Surveill*, vol. 23, no. 9, pp. 18–00086, Mar. 2018, doi: 10.2807/1560-7917.ES.2018.23.9.18-00086.
[26] A. Melidou *et al.*, "Predominance of influenza virus A(H3N2) 3C.2a1b and A(H1N1)pdm09 6B.1A5A genetic subclades in the WHO European Region, 2018–

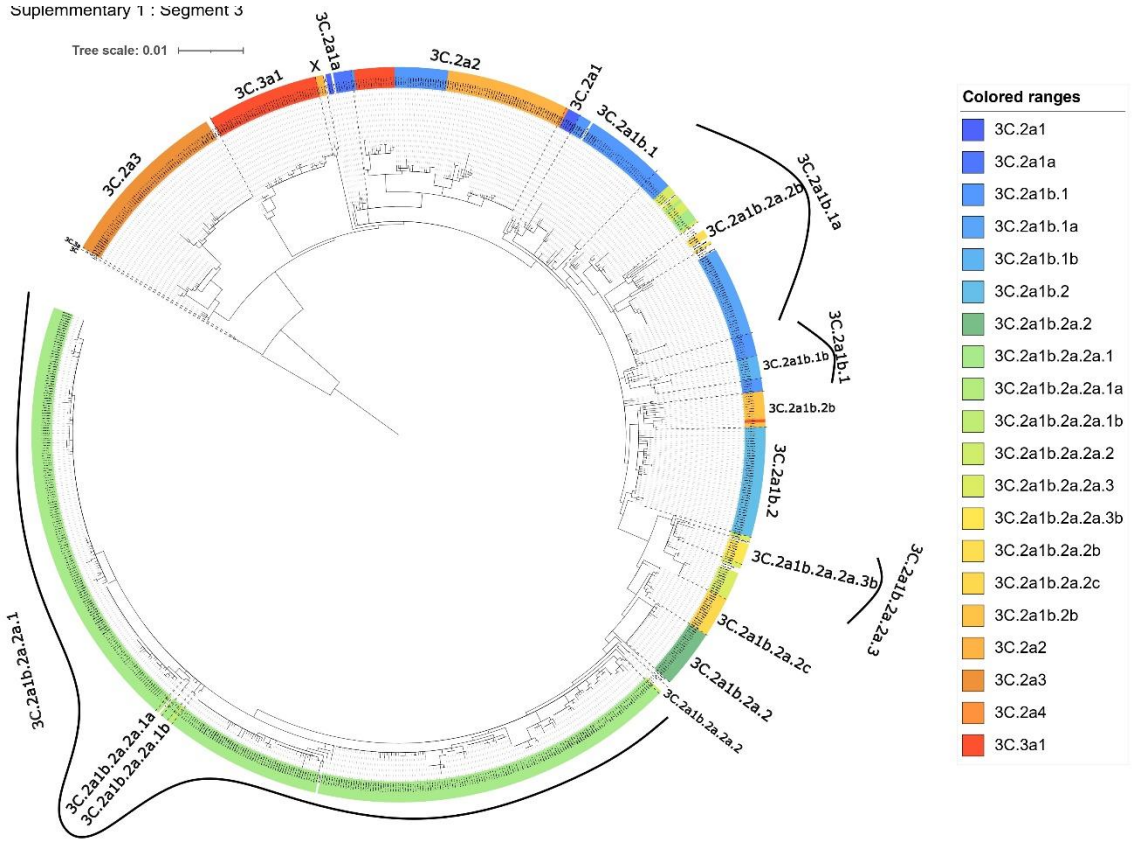
2019," *Vaccine*, vol. 38, no. 35, pp. 5707–5717, Jul. 2020, doi: 10.1016/j.vaccine.2020.06.031.
[27] A. Melidou *et al.*, "Virological surveillance of influenza viruses in the WHO European Region in 2019/20 – impact of the COVID-19 pandemic," *Euro Surveill*, vol. 25, no. 46, p. 2001822, Nov. 2020, doi: 10.2807/1560-7917.ES.2020.25.46.2001822.
[28] A. Giraud-Gatineau *et al.*, "Control of common viral epidemics but not of SARS-CoV-2 through the application of hygiene and distancing measures," *Journal of Clinical Virology*, vol. 150–151, p. 105163, Jun. 2022, doi: 10.1016/j.jcv.2022.105163.
[29] C. Adlhoch and R. Pebody, "What to expect for the influenza season 2020/21 with the ongoing COVID-19 pandemic in the World Health Organization European Region," *Euro Surveill*, vol. 25, no. 42, p. 2001816, Oct. 2020, doi: 10.2807/1560-7917.ES.2020.25.42.2001816.
[30] C. Adlhoch *et al.*, "Very little influenza in the WHO European Region during the 2020/21 season, weeks 40 2020 to 8 2021," *Euro Surveill*, vol. 26, no. 11, p. 2100221, Mar. 2021, doi: 10.2807/1560-7917.ES.2021.26.11.2100221.
[31] "Influenza characterisation report April 2021", [Online]. Available: https://www.ecdc.europa.eu/sites/default/files/documents/Influenza-characterisation-report-April-2021.pdf
[32] A. Melidou *et al.*, "Influenza returns with a season dominated by clade 3C.2a1b.2a.2 A(H3N2) viruses, WHO European Region, 2021/22," *Eurosurveillance*, vol. 27, no. 15, Apr. 2022, doi: 10.2807/1560-7917.ES.2022.27.15.2200255.
[33] "Seasonal influenza 2022-2023 Annual Epidemiological Report 2023," 2023.
[34] "Seasonal influenza 2021-2022 - Annual Epidemiological Report," 2021.
[35] "Seasonal influenza - Annual Epidemiological Report for 2020-2021." Accessed: Jul. 19, 2023. [Online]. Available: https://www.ecdc.europa.eu/en/publications-data/seasonal-influenza-annual-epidemiological-report-2020-2021
[36] "Seasonal influenza - Annual Epidemiological Report for 2019–2020." Accessed: Jul. 19, 2023. [Online]. Available: https://www.ecdc.europa.eu/en/publications-data/seasonal-influenza-annual-epidemiological-report-2019-2020
[37] I. Maljkovic Berry *et al.*, "Frequency of influenza H3N2 intra-subtype reassortment: attributes and implications of reassortant spread," *BMC Biol*, vol. 14, p. 117, Dec. 2016, doi: 10.1186/s12915-016-0337-3.
[38] E. J. Goldstein *et al.*, "Integrating patient and whole-genome sequencing data to provide insights into the epidemiology of seasonal influenza A(H3N2) viruses," *Microb Genom*, vol. 4, no. 1, p. e000137, Dec. 2017, doi: 10.1099/mgen.0.000137.
[39] G. Nabakooza, A. Pastusiak, D. P. Kateete, J. J. Lutwama, J. M. Kitayimbwa, and S. D. W. Frost, "Whole-genome analysis to determine the rate and patterns of intra-subtype reassortment among influenza type-A viruses in Africa," *Virus Evol*, vol. 8, no. 1, p. veac005, Jan. 2022, doi: 10.1093/ve/veac005.
[40] K. B. Westgeest *et al.*, "Genomewide Analysis of Reassortment and Evolution of Human Influenza A(H3N2) Viruses Circulating between 1968 and 2011," *J Virol*, vol. 88, no. 5, pp. 2844–2857, Mar. 2014, doi: 10.1128/JVI.02163-13.
[41] L. A. E. Van Poelvoorde *et al.*, "Whole-genome-based phylogenomic analysis of the Belgian 2016–2017 influenza A(H3N2) outbreak season allows improved surveillance," *Microb Genom*, vol. 7, no. 9, p. 000643, Sep. 2021, doi: 10.1099/mgen.0.000643.
[42] K. Y. Taylor *et al.*, "Influenza A virus reassortment is strain dependent," *PLoS Pathog*, vol. 19, no. 3, p. e1011155, Mar. 2023, doi: 10.1371/journal.ppat.1011155.
[43] C. Gavazzi *et al.*, "A functional sequence-specific interaction between influenza A virus genomic RNA segments," *Proc Natl Acad Sci U S A*, vol. 110, no. 41, pp. 16604–16609, Oct. 2013, doi: 10.1073/pnas.1314419110.

SUPPLEMENTARY DATA

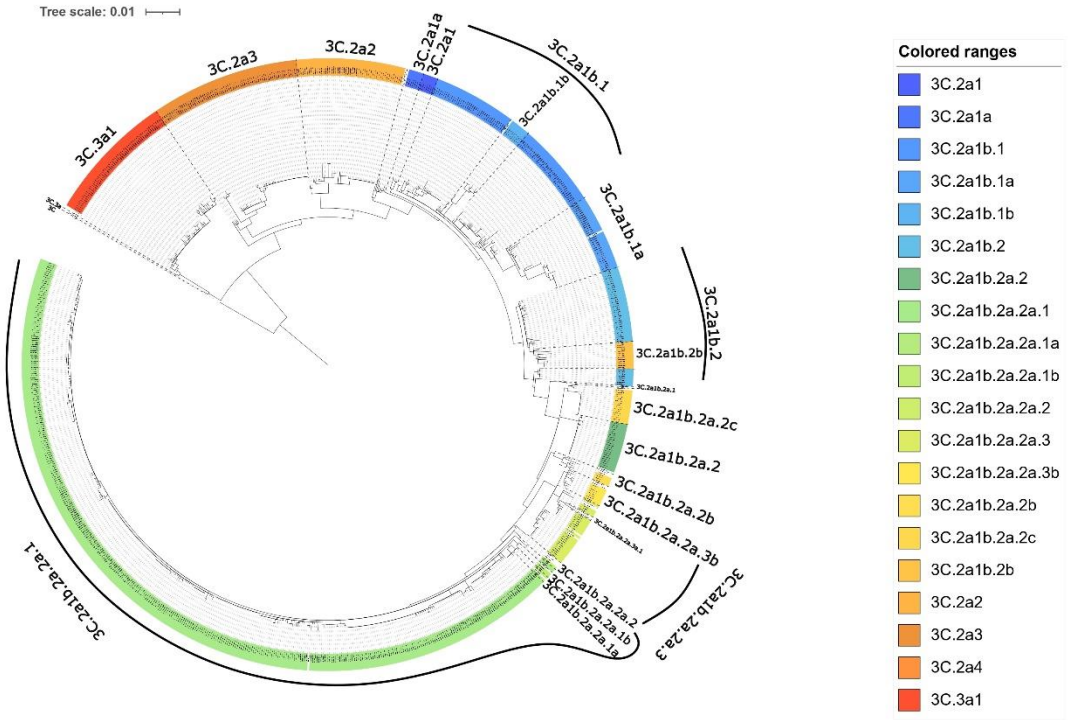
Supplementary 1



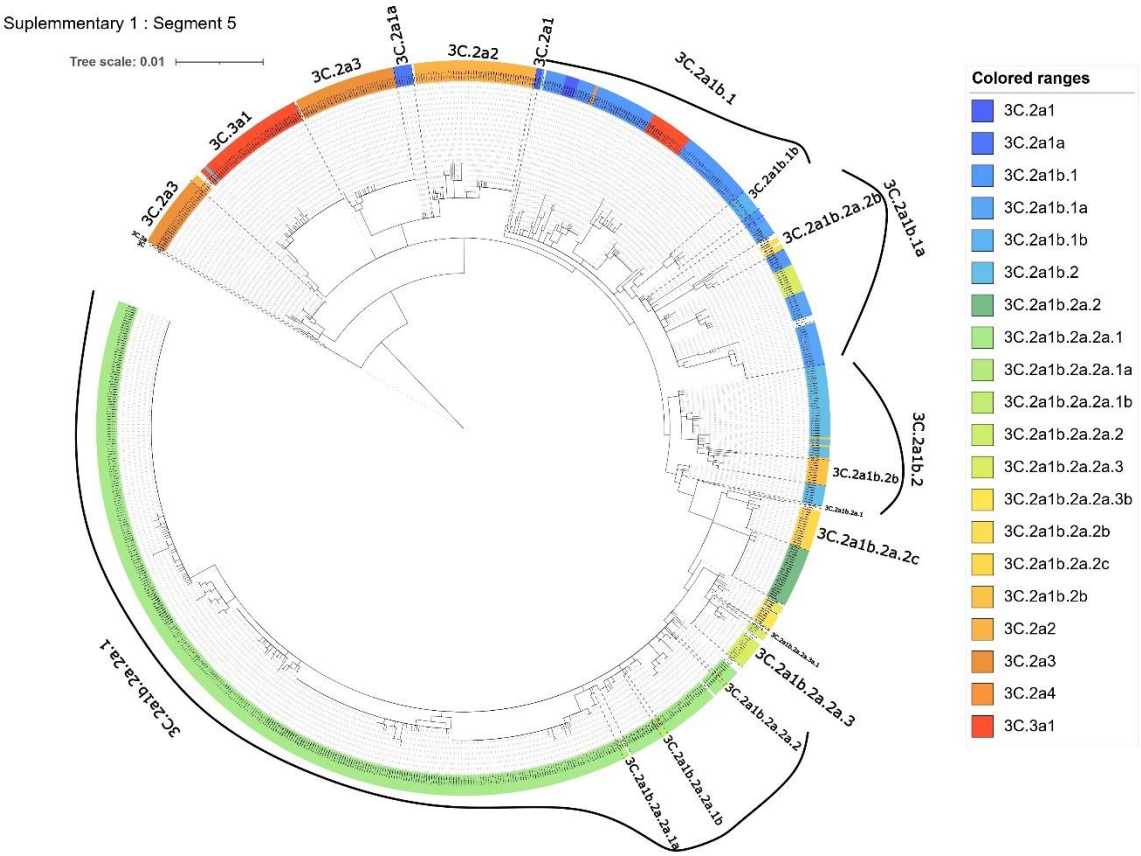

Suplemmentary 1 : Segment 2
Tree scale: 0.01
3C.2a1a
3C.2a2
3C.2a1
3C.3a1
3C.2a3
3C.2a1b.1
3C.2a1b.1b
3C.2a1b.2a.2b
3C.2a1b.1a
3C.2a1b.2
3C.2a1b.2b
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2a.1b
3C.2a1b.2a.1a
3C.2a1b.2a.2c
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
Colored ranges
3C.2a1
3C.2a1a
3C.2a1b.1
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2b
3C.2a1b.2a.2c
3C.2a1b.2b
3C.2a2
3C.2a3
3C.2a4
3C.3a1

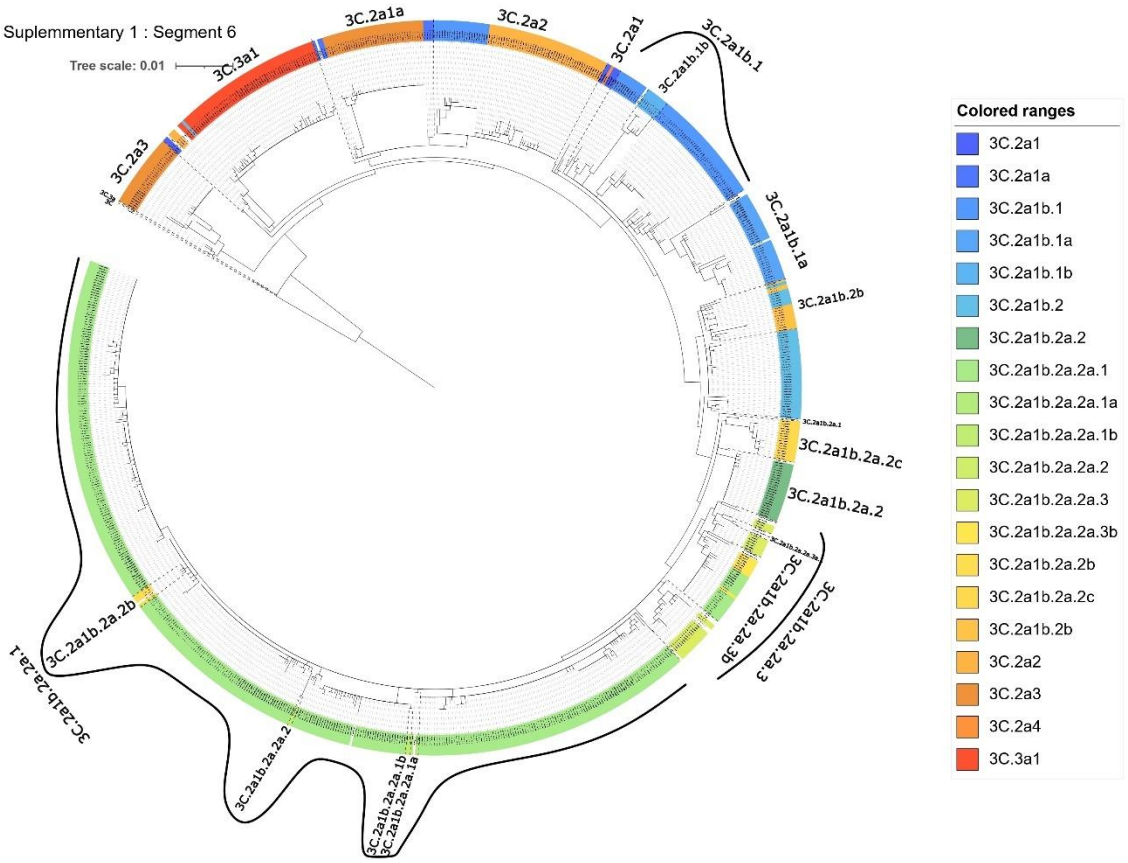

Suplemmentary 1 : Segment 3
Tree scale: 0.01
3C.2a1a
3C.2a2
3C.2a1
3C.3a1
3C.2a3
3C.2a1b.1
3C.2a1b.2a.2b
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2b
3C.2a1b.2
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2a.3
3C.2a1b.2a.2c
3C.2a1b.2a.2
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
Colored ranges
3C.2a1
3C.2a1a
3C.2a1b.1
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2b
3C.2a1b.2a.2c
3C.2a1b.2b
3C.2a2
3C.2a3
3C.2a4
3C.3a1

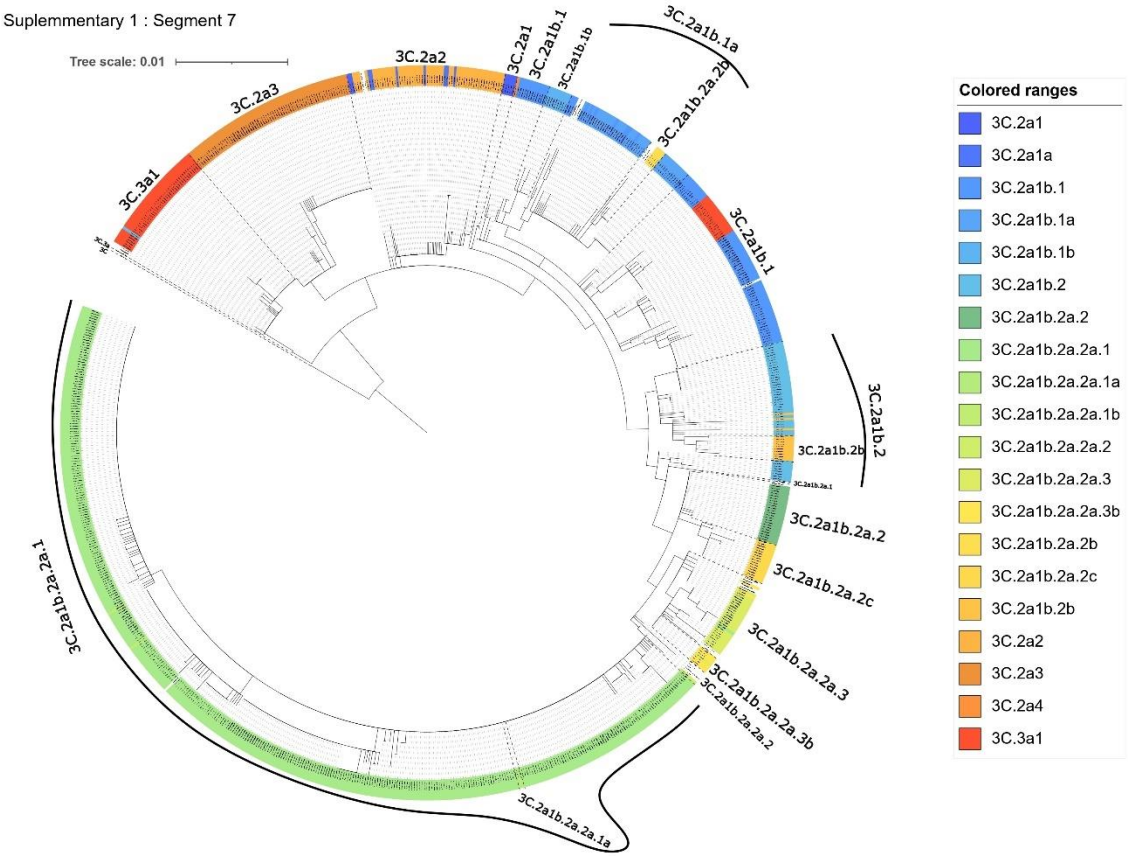
Suplemmentary 1 : Segment 4

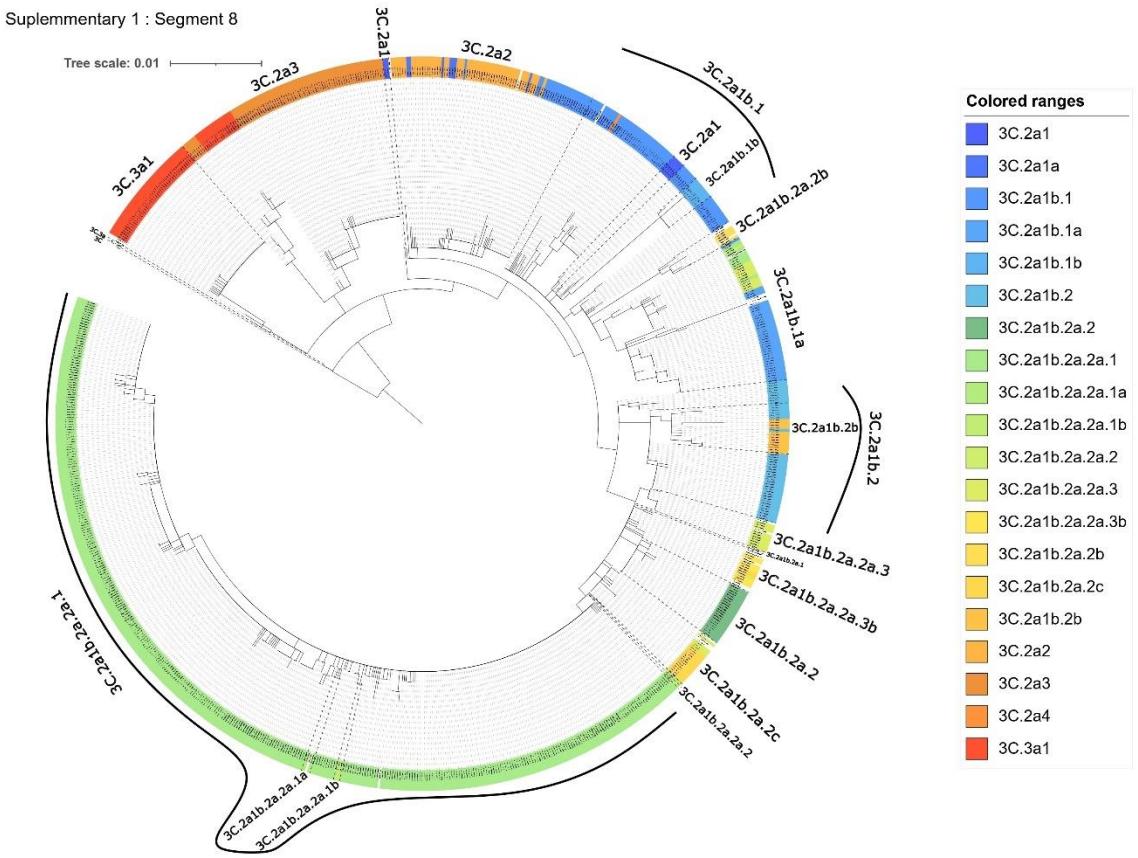


Suplemmentary 1 : Segment 5




Suplemmentary 1 : Segment 6
Tree scale: 0.01
3C.2a1a
3C.2a2
3C.2a1
3C.2a1b.1
3C.3a1
3C.2a1b.1b
3C.2a3
3C.2a1b.1a
3C.2a1b.2b
3C.2a1b.2a.2c
3C.2a1b.2a.2
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2a.3
3C.2a1b.2a.2b
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.1a
Colored ranges
3C.2a1
3C.2a1a
3C.2a1b.1
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2b
3C.2a1b.2a.2c
3C.2a1b.2b
3C.2a2
3C.2a3
3C.2a4
3C.3a1


Suplemmentary 1 : Segment 7
Tree scale: 0.01
3C.2a2
3C.2a1
3C.2a1b.1
3C.2a1b.1b
3C.2a1b.1a
3C.2a1b.2b
3C.2a3
3C.3a1
3C.2a1b.1
3C.2a1b.2
3C.2a1b.2b
3C.2a1b.2a.2
3C.2a1b.2a.2c
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1
Colored ranges
3C.2a1
3C.2a1a
3C.2a1b.1
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2b
3C.2a1b.2a.2c
3C.2a1b.2b
3C.2a2
3C.2a3
3C.2a4
3C.3a1


Suplemmentary 1 : Segment 8
Tree scale: 0.01
3C.2a1
3C.2a2
3C.2a3
3C.3a1
3C.2a1b.1
3C.2a1
3C.2a1b.1b
3C.2a1b.2a.2b
3C.2a1b.1a
3C.2a1b.2b
3C.2a1b.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2
3C.2a1b.2a.2c
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.1
Colored ranges
3C.2a1
3C.2a1a
3C.2a1b.1
3C.2a1b.1a
3C.2a1b.1b
3C.2a1b.2
3C.2a1b.2a.2
3C.2a1b.2a.2a.1
3C.2a1b.2a.2a.1a
3C.2a1b.2a.2a.1b
3C.2a1b.2a.2a.2
3C.2a1b.2a.2a.3
3C.2a1b.2a.2a.3b
3C.2a1b.2a.2b
3C.2a1b.2a.2c
3C.2a1b.2b
3C.2a2
3C.2a3
3C.2a4
3C.3a1

Article n°4 :
Monitoring of Influenza A/H1N1 recombination and reassortment in the southern France between 2017-2022

Safietou Dialo, **Lorlane Le Targa**, Houmadi Hikmat, Bernard La Scola, Philippe Colson (en cours d'écriture, pas encore soumis)

Monitoring of Influenza A/H1N1 recombination and reassortment in the southern France between 2017-2022

Author list: Safietou Diallo[1,2], Lorlane Le Targa[1,2,3] , Houmadi Hikmat[1,2], La Scola Bernard[1,2,4] Philippe, Colson[1,2,4*]

Affiliations:

[1] IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [2] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Institut de Recherche pour le Développement (IRD), Microbes Evolution Phylogeny and Infections (MEPHI), Assistance Publique-Hôpitaux de Marseille (AP-HM), IHU Méditerranée Infection, 19-2127 boulevard Jean Moulin, 13005 Marseille, France; [3] Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France; [4] Assistance Publique-Hôpitaux de Marseille (AP-HM), 264 rue Saint-Pierre, 13005 Marseille, France

* Corresponding authors:

Philippe Colson IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France.

Tel.: +33 413 732 401, Fax: +33 413 732 402; email: philippe.colson@univ-amu.fr

17 # I. INTRODUCTION

18 Viruses have the ability to exchange genetic material during co-infection, a phenomenon that
19 often gives rise to new strains with implications for viral evolution and public health (1). In
20 influenza, this exchange is known as reassortment, a process by which influenza viruses
21 exchange genetic segments. This genetic exchange is made possible by the segmented nature
22 of the viral genome and occurs when two different influenza viruses co-infect the same cell.
23 The viral diversity generated by reassortment is vast and plays an important role in the evolution
24 of influenza viruses, making it a real global public health issue, as demonstrated by the 2009
25 pandemic caused by the H1N1 virus of swine origin (2). In metropolitan France, as in all
26 countries with a temperate climate, influenza causes an epidemic every year during the winter
27 season, generally between December and March. Influenza epidemics occur every year and are
28 responsible for considerable morbidity and absence from school among healthy children and
29 adolescents (3).

30 According to the WHO, seasonal influenza causes around a billion cases worldwide each year,
31 including 3 to 5 million cases of severe illness, and between 290,000 and 650,000 deaths (4).
32 In France, influenza affects 2 to 6 million people and causes around 10,000 deaths every year
33 (5). Mortality from seasonal influenza most often occurs in the elderly, the
34 immunocompromised and people with serious risk factors such as chronic illness (6). Prior to
35 covid 19, influenza was the most frequent viral etiology in intensive care (7). However, a
36 significant decline in the number of influenza cases was observed after the emergence of the
37 SARS-CoV-2 pandemic worldwide. In Japan and the Republic of Korea, low influenza activity
38 was maintained throughout the Covid 19 pandemic (8). In Marseille, the number of influenza-
39 related deaths fell significantly in winter 2019-2020 compared with the previous winter (9).
40 The most significant decrease in the number of influenza deaths occurred in winter 2020-2021,

41 when the ECDC surveillance report recorded more than 150 times fewer influenza cases in
42 WHO regions (10).

43 The aim of this work is to characterize H1N1 influenza at the IHU from 2017 to 2022 in order
44 to gain insight into the evolution of H1N1 influenza before and after the Covid 19 pandemic.
45 To this end, we are interested in the clade present at the IHU during each influenza epidemic
46 during this period, and in the presence of recombinations and/or reassortments.

47

48

II. Materials and methods

1. RT-PCR detection:

Influenza is routinely diagnosed in the diagnostic laboratory of the Institut Hospitalo-Univertaire de Marseille (IHU). Several different methods are used for diagnosis:

- FTD Respiratory pathogens 21 (FTD) (Fast Track Diagnosis, Luxembourg), which detects influenza A and B and identifies the A/H1 variant, but does not provide a CT.
- GeneXpert Xpert Flu/RSV (GeneXpert) (Cepheid, Sunnyvale, CA, USA), which detects influenza A and B and identifies the A/H3, A/H1 and A/H1-2009 variants, as well as CTs.
- Or the BioFire FilmArray Respiratory panel 2 plus (BioFire) (Biomérieux, Marcy-Etoile, France), which targets all M1 genes (plus two genes for the GeneXpert test).

2. Extraction method:

Viral RNA was extracted from respiratory samples using either EZ1 Virus Mini v2.0 kits on an EZ1 Advanced XL instrument (Qiagen, Courtaboeuf, France) or the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA,USA), following the manufacturer's instructions.

3. Amplification method:

Amplification was performed using the B-method primer system described by P.K.Mitchell et al. (11). This system, based on the optimization of a system initially designed in 2009 by Zhou et al. (12), uses two forward primers (MBTuni-12- GTTACGCGCCAGCAAAAGCAGG and MBTuni-12 SNV-GTTACGCGCCAGCGAAAGCAGG) and one reverse primer (MBTuni-13-GTTACGCGCCAGTAGAAACAAGG) and enables complete sequencing of all 8 human or animal influenza A segments. For this study, RT-PCR conditions were optimized to obtain a CT of 33 or less. Primers MBTuni-12 and MBTuni-12-SNV were used at 200nm each and MBTuni-13 at 400nm. Thermal cycling conditions began with a first RT-PCR step cycling at

55°C for 30s, then 40°C for 50min and 94°C for 2min. Then a pre-amplification step was carried out in 5 cycles of 94°C 15s, 42°C 30s, and 68°C 3min. Then amplification is carried out over 35 cycles of 94°C 15s, 42°C 30s, 68°C 3min with a final extension at 68°C for 5min.

4. Sequencing method: (In progress)

Sequencing was performed using Illumina technology (Mi_seq and Nova_seq) according to the prototype used by Colson et al. (2021).

5. Clades and subclades :

Clades and subclades were assigned using Nexclades (WEB version) based on segment 4, which codes for hemagglutinin (HA), and segment 6, which codes for neuraminidase (NA).

6. Phylogeny:

Sequences were aligned using Mafft v7.505 (13) with the following parameters:

thread = 8; threadtb = 5; threadit = 0; reorder; auto.

Phylogenetic trees were constructed with IQTREE v2.2.0 COVID-edition (14), using the Neighbor-joining method with boostrap analysis for 1000 replicas and Maximum Likelihood as the model. The rate among sites is +I. The command line used is: iqtree2 -s [ALIGNMENT] --alrt 1000 -B 1000 -m TN+FO+I.

The phylogenetic trees were visualized and annotated with ITOL (15).

7. Detection of recombination and reassortment:

a. Recombinations

Recombination detection was performed using RDP4 software, which contains a wide range of methods including RDP, GENECOV, BootScan, MaxChi, Chimera, SiScan and 3seq to detect and visualize recombination events (16) (17). For our analysis, all default parameters were

retained, and at least four detection techniques were considered in order to reduce false positives.

b. Reassortments

The search for reassortments was carried out using GiRAF, a tool that enables automated reassortment searches for phylogenetic incongruence (18). To do this, the alignments of each segment made with MAFFT were submitted to MrBayes V.3.2.6 software to generate trees, 200,000 generations with a sampling frequency of 200 (19). These trees were then used with GiRAF software to find potential reassortments. The default parameters were retained for burnin and treashold (the confidence threshold for signaling a reassortment), with only cull (the fraction of the tree in which the reassortment is found) modified to 10%.

III. Results and discussion

1. Results

a. Clades and subclades

In total, our study covered 857 complete genomes with all 8 segments. Influenza A/H1N1 is subdivided into several clades and subclades according to segment 4 (HA), which codes for hemagglutinin, and segment 6 (NA), which codes for neuraminidase. Over the period from 2017 to 2022, the IHU detected 10 clades according to segment 4 and 8 clades according to segment 6. For segment 4, subclade 6B.1A.5a.1 was in the majority with (58.9%), followed by subclades 6B.1A with (17.8%), 6B.1A.5a (9.1%), 6B.1A.5b (8.2%). Subclades 6B.1A.5a.2, 6B.1A.7, 6B.1A.1, 6B.1A.6, 6B.1A.5a.2a.1, 6B.1A.2 were poorly represented, with (2.2%), (2.1%), (0.9%) (0.3%) (0.3%) (0.1%) respectively (Table 1). For segment 6 clades, over the period December 2017 to December 2022, the majority subclade was B.3.1.2 (50%), followed

by A.1.1 (30.1%), B.1 (8.5%), B.2.1 (5.6%), B.3 (2.2%), B.3.1 (2.1%), B.2 (1%) and C.5.2 (0.3%). Subclades A.1 and C were not detected at this time at the IHU (table 2).

b. Evolution of clades and subclades

Over the period from December 2018 to March 2020, all the sub-clades detected with segment 4 circulated at the IHU. Between June 2020 and September 2021, no subclades were detected at the IHU. However, in December 2021, only subclade 6B.1A.5a.1 reappeared at a high rate (Value) and subclade 6B.1A.5a.2a.1 at a low rate.

For the clades found with segment 6, over the period from December 2017 to April 2018, only clade A.1.1 was present. Between December 2018 and April 2020, several sub-clades circulated. However, in December 2021, a new sub-clade (B.3.1.2) appeared with a high rate.

c. Phylogenie

- ***Recombination:*** the search for recombination by RDP4, we considered the results found by at least 4 methods. The results showed no recombination but 17 strains reassorting in the same way between segment 2 and segment 3. These 17 strains all belonged to the same clade 6B.1A.5a.1 and were all at the IHU in winter 2022.

- ***Reassortment:*** Reassortment: the search for reassortants with GiRaF revealed the same result as that obtained with RDP4: out of the 857 sequences in our analysis, only 17 reassortants were found, all detected during the 2021-2022 epidemic. All reassortants detected by GiRAF belong to clade 6B.1A.5a.1. This clade was present at the IHU during winter 2019 and reappeared during the 2021-2022 influenza season. A phylogeny was performed for each segment, and a phylogenetic tree was created with the concatenation of the eight segments to show where the 17 potential reassortant strains found lie. A/California/07/2009 (A/H1N1pdm09), was considered as the root for the 8 trees made for each segment and the

concatenated tree. The 26 reference sequences used to make the trees were retrieved from GISAID and each was composed of the 8 complete A/H1N1 segments. All these trees were made with ITOL.

2. Discussion

In recent years, three subtypes of influenza A virus (H1N1, H2N2 and H3N2) have been circulating in the human population. These viruses were the cause of epidemics against which humans lacked neutralizing antibodies, resulting in fulminant worldwide spread, with millions of infections and increased mortality (20). The arrival of SARS-CoV-2 has had a major impact on influenza epidemics worldwide. Our study of 857 complete H1N1 influenza genomes covers the period from 2017 to 2022, corresponding to the SARS-CoV-2 pandemic. In our study, we found different types of clades and sub-clades that had circulated at IHU during the period from 2017 to 2022. These clades and sub-clades were assigned based on segment 4 (HA) and segment 6 (NA). Among the clades and subclades found with segment 4, the most frequent was subclade 6B.1A.5a.1 with 505 samples or 58.9%. Parallel to our study, this majority subclade was found by Kissling. E et al in a multicenter study conducted in 2021 and 2022 (21). According to the 2022 ECDC-WHO report, few influenza A/H1N1 viruses were detected during the 2021-2022 season. The sub-clades detected during this period were 6B.1A.5a.1 and 6B.1A.5a.2, with a much higher frequency in Europe for sub-clade 6B.1A.5a.1. Subclade 6B.1A.5a.2 has been detected worldwide (22). At IHU, in addition to subclade 6B.1A.5a.1, we also detected subclade 6B.1A.5a.2.a.1 in a small percentage (0.3%) over the same period.

Between 2020 and 2021, no cases of influenza were detected at the IHU. This could be due to the arrival of the covid-19 pandemic. During this period, many countries implemented sanitary measures against covid-19, such as compulsory masking, physical distancing, confinement and increased use of hand sanitizers (23). At the same time, these measures would have prevented the transmission of influenza viruses during this period. In winter 2021, a high number of

166 influenza cases were detected at the IHU. This could be explained by the lifting of sanitary
167 measures previously established to combat covid-19.

168

BIBLIOGRAPHICAL REFERENCES

1. Taylor KY, Agu I, José I, Mäntynen S, Campbell AJ, Mattson C, et al. Influenza A virus reassortment is strain dependent. PLOS Pathog. 1 mars 2023;19(3):e1011155.

2. Steel J, Lowen AC. Influenza A Virus Reassortment. In: Compans RW, Oldstone MBA, éditeurs. Influenza Pathogenesis and Control - Volume I [Internet]. Cham: Springer International Publishing; 2014 [cité 26 janv 2024]. p. 377-401. (Current Topics in Microbiology and Immunology). Disponible sur: https://doi.org/10.1007/82_2014_395

3. Brady RC. Influenza. Adolesc Med State Art Rev. août 2010;21(2):236-50, viii.

4. Influenza (Seasonal) [Internet]. [cité 25 janv 2024]. Disponible sur: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

5. Institut Pasteur [Internet]. 2015 [cité 25 janv 2024]. Grippe. Disponible sur: https://www.pasteur.fr/fr/centre-medical/fiches-maladies/grippe

6. SPF. Fardeau de la grippe en France métropolitaine, bilan des données de surveillance lors des épidémies 2011-12 à 2021-22 [Internet]. [cité 25 janv 2024]. Disponible sur: https://www.santepubliquefrance.fr/import/fardeau-de-la-grippe-en-france-metropolitaine-bilan-des-donnees-de-surveillance-lors-des-epidemies-2011-12-a-2021-22

7. Naouri D, Pham T, Dres M, Vuagnat A, Beduneau G, Mercat A, et al. Differences in clinical characteristics and outcomes between COVID-19 and influenza in critically ill adult patients: A national database study. J Infect [Internet]. 16 mai 2023 [cité 25 janv 2024]; Disponible sur: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10186846/

8. Takashita E, Watanabe S, Hasegawa H, Kawaoka Y. Are twindemics occurring? Influenza Other Respir Viruses. 2023;17(1):e13090.

192 9. Giraud-Gatineau A, Colson P, Jimeno MT, Zandotti C, Ninove L, Boschi C, et al.
193 Comparison of mortality associated with respiratory viral infections between December 2019
194 and March 2020 with that of the previous year in Southeastern France. Int J Infect Dis. 1 juill
195 2020;96:154-6.

196 10. Influenza-characterisation-report-April-2021.pdf [Internet]. [cité 25 janv 2024].
197 Disponible sur: https://www.ecdc.europa.eu/sites/default/files/documents/Influenza-
198 characterisation-report-April-2021.pdf

199 11. Mitchell PK, Cronk BD, Voorhees IEH, Rothenheber D, Anderson RR, Chan TH, et
200 al. Method comparison of targeted influenza A virus typing and whole-genome sequencing
201 from respiratory specimens of companion animals. J Vet Diagn Investig Off Publ Am Assoc
202 Vet Lab Diagn Inc. mars 2021;33(2):191-201.

203 12. Zhou B, Donnelly ME, Scholes DT, St. George K, Hatta M, Kawaoka Y, et al. Single-
204 Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for
205 Classical and Swine Origin Human Influenza A Viruses. J Virol. oct 2009;83(19):10309-13.

206 13. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
207 Improvements in Performance and Usability. Mol Biol Evol. avr 2013;30(4):772-80.

208 14. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A,
209 et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
210 Genomic Era. Mol Biol Evol. mai 2020;37(5):1530-4.

211 15. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and
212 annotation - PMC [Internet]. [cité 1 févr 2024]. Disponible sur:
213 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8265157/

16. Lohrasbi-Nejad A. Detection of homologous recombination events in SARS-CoV-2. Biotechnol Lett. 2022;44(3):399-414.

17. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 26 mai 2015;1(1):vev003.

18. Nagarajan N, Kingsford C. GiRaF: robust, computational identification of influenza reassortments via graph mining. Nucleic Acids Res. mars 2011;39(6):e34.

19. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. mai 2012;61(3):539-42.

20. Fan S, Kong H, Babujee L, Presler R, Jester P, Burke D, et al. Assessment of the antigenic evolution of a clade 6B.1 human H1N1pdm influenza virus revealed differences between ferret and human convalescent sera. eBioMedicine. 16 févr 2024;101:105013.

21. Kissling E, Pozo F, Martínez-Baz I, Buda S, Vilcu A, Domegan L, et al. Influenza vaccine effectiveness against influenza A subtypes in Europe: Results from the 2021–2022 I-MOVE primary care multicentre study. Influenza Other Respir Viruses. 21 nov 2022;17(1):e13069.

22. Caractérisation du virus grippal - Résumé Europe, septembre 2022 [Internet]. [cité 29 mai 2024]. Disponible sur: https://www.ecdc.europa.eu/en/publications-data/influenza-virus-characterization-summary-europe-september-2022

23. Ng V, Fazil A, Waddell LA, Bancej C, Turgeon P, Otten A, et al. Effets projetés des mesures de santé publique non pharmacologiques visant à prévenir la recrudescence de la transmission du SRAS-CoV-2 au Canada. CMAJ Can Med Assoc J. 30 nov 2020;192(48):E1673-85.

237 **TABLE AND FIGURES**

238 **Table 1: Segment 4 clades and subclades**

Clades and subclades segment 6	Number	Percentage
6B.1A.5a.1	505	**58.9**
6B.1A.5a.2	19	2.2
6B.1A	153	**17.9**
6B.1A.5a	77	**9.1**
6B.1A.5b	70	**8.2**
6B.1A.1	8	0.9
6B.1A.7	18	2.1
6B.1A.6	3	0.3
6B.1A.2	1	0.1
6B.1A.5a.2a.1	3	0.3
Total	85	100

239

240

241

242 **Table 2: Clades and subclades segment 6**

243

Clades and subclades segment 6	Number	percentage
A.1.1	258	**30.1**
B.2.1	48	5.6
B.3	19	2.2
B.2	9	1
B.3.1.2	430	**50.2**
B.3.1	17	2.1
C.5.2	3	0.3
B.1	73	8.5
A.1	0	0
C	0	0
Total	857	100

244

245

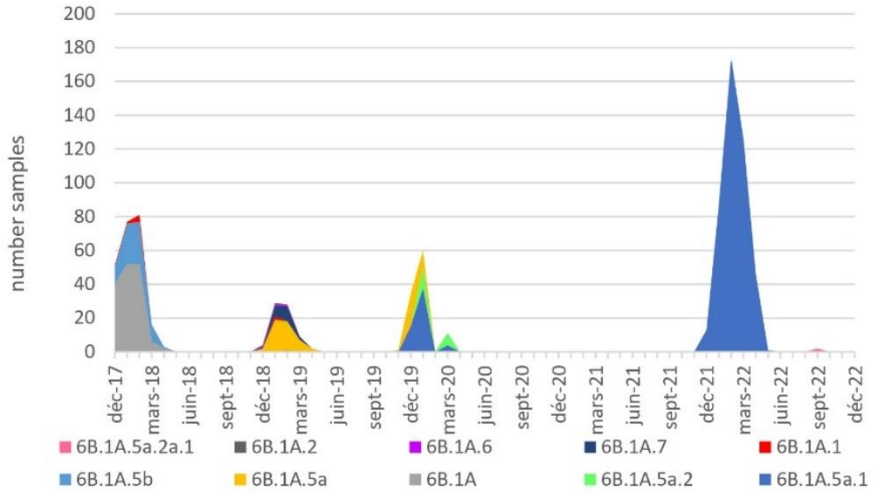


Fig1: Change in segment 4 clades (HA) between 2017 and 2022

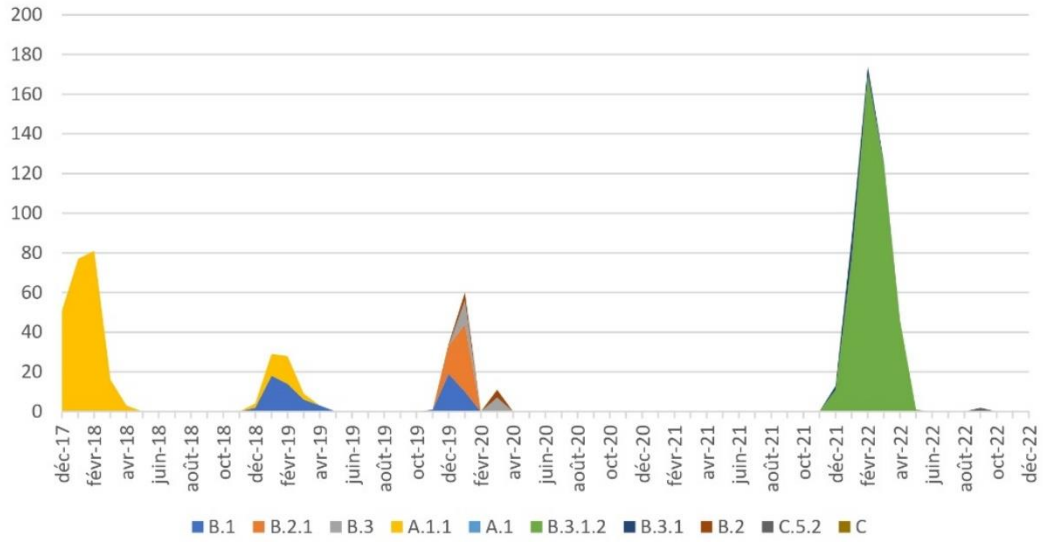


Fig2: Evolution of clades in segment 6 (NA) between 2017 and 2022

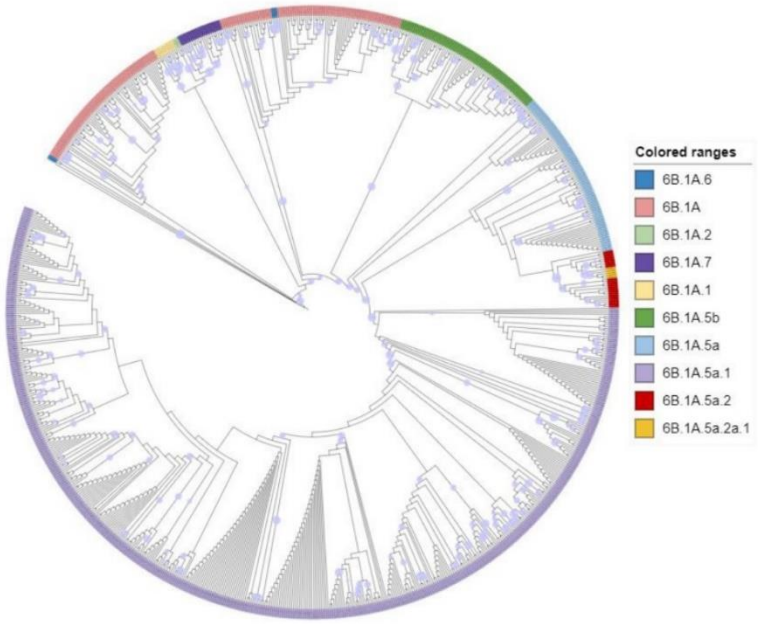


Fig 1 : segment 4

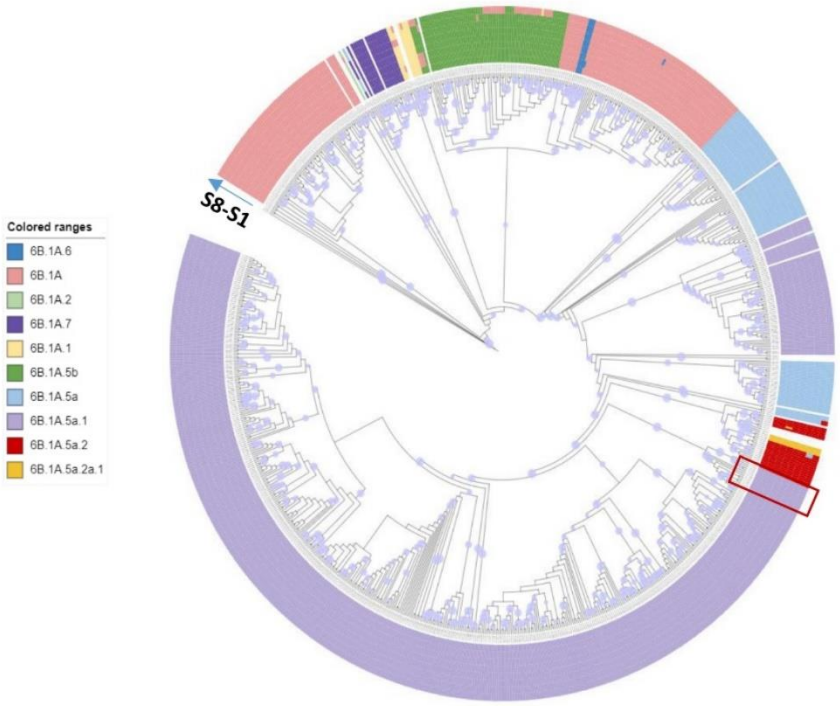


Fig 2 : Arbre concatené

Suplementary data

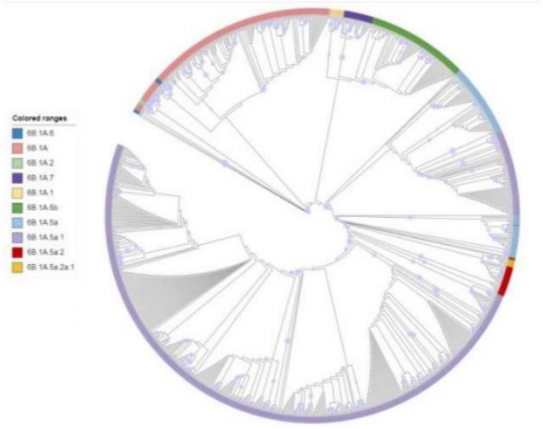


Fig 1: segment 1

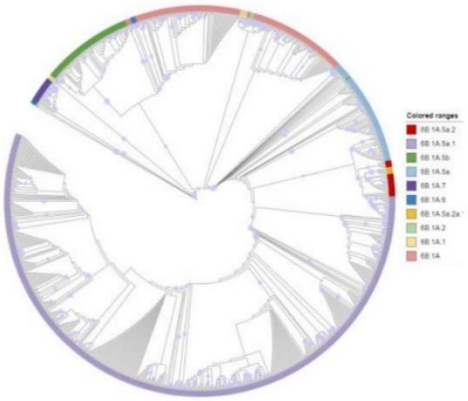


Fig 2: segment 2

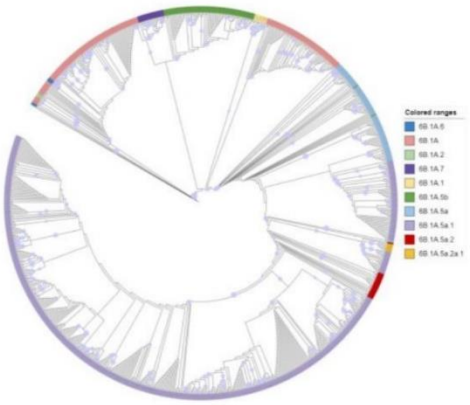


Fig 3 : segment 3

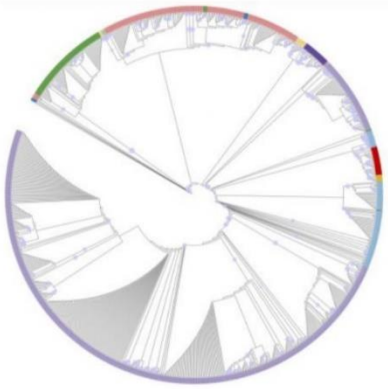
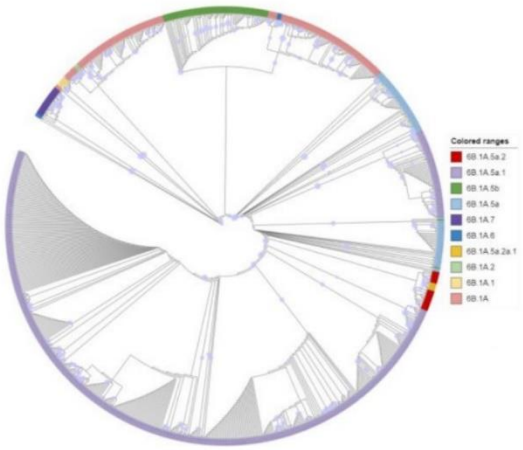


Fig 4 : segment 5



273

274 Fig 5 : segment 6

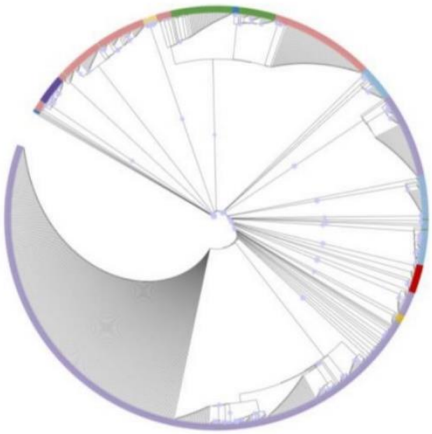
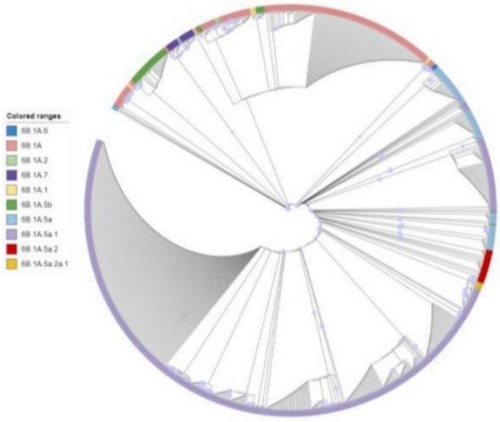


Fig 6 : segment 7

275

276



277

278 Fig 7 : segment 8

279

Chapitre IV : L'importance du séquençage dans la détection (ou non) de faux positif

Avant-propos :

Que l'on parle de séquençage ciblé ou de qPCR spécifique, il est possible que les systèmes d'amplification utilisés voient leur sensibilité diminuée au fur et à mesure du temps en raison des mutations pouvant apparaître sur les séquences cibles. Dans le cadre des qPCR, cette perte de sensibilité peu notamment entraîner des faux négatifs liés à une absence de fixation au génomes. C'est notamment pour cela qu'il est nécessaire de réaliser des suivis réguliers des nouvelles séquences disponibles afin de mettre à jour les systèmes d'amplification utilisés.

Le virus de la grippe mute rapidement. Néanmoins, certains gènes mutent peu et possèdent une faible variation génétique. C'est le cas du gène M1 qui code pour la protéine virale de la matrice et qui est présent sur le segment 7 de l'influenza virus A [61]. C'est pour cela que ce gène est utilisé comme cible dans plusieurs kits de qPCR disponible sur le marché. Mais même si ce gène est plutôt conservé, cela n'empêche pas l'apparition de mutations pouvant être responsable d'une perte de sensibilité des systèmes de qPCR.

En 2022, l'émergence et la circulation de virus grippaux A/H3N2 porteurs de mutations associées à des résultats faussement négatifs de tests qPCR ont été rapportées au Danemark [62]. L'étude de Jørgensen et al. a pu corréler des résultats faussement négatifs de grippe A/H3N2 avec deux mutations présentes dans le Gène M1 du segment 7. Pour cela, ils ont utilisé des échantillons faux négatifs, ils ont pu trouver ces échantillons en obtenant des résultats positifs avec d'autres systèmes qPCR. Leur étude a analysé les génomes de huit échantillons (quatre positifs et quatre négatifs) ainsi que de huit souches circulantes du centre national de référence et du laboratoire de surveillance de la grippe du Danemark.

À la suite de l'étude danoise, l'institut pasteur a lancé une alerte en France sur le risque de faux négatif [63]. Dans le cadre de notre étude sur le suivi de l'évolution des réassortiments de la grippe A nous avions toutes les séquences nécessaires pour vérifier de notre côté si ces mutations circulaient dans les grippes A/H3N2 à Marseille (France) en 2022 et si ces dernières semblaient impliquées dans l'augmentation de faux négatifs.

Nous avons alors pu réaliser une analyse sur 624 séquences du segment 7, provenant d'échantillons prélevés entre septembre 2021 et avril 2022, dont 93,3% présentaient les deux mutations supposées associées aux faux négatifs. Nous avons alors pu tester sur plusieurs échantillons avec une faible charge virale et donc un Ct (cycle threshold)élevé et positifs avec des tests non associés au faux négatif.

Article n°5 :
Matrix-Encoding Gene Diversity of 624 Influenza A/H3N2 Genomes Does Not Show Association with Impaired Viral Detection by Commercialized qPCR Assays

Lorlane Le Targa, Houmadi Hikmat, Céline Boschi, Bernard La Scola, Philippe Colson (Publier dans Viruses)

viruses

MDPI

Brief Report

Matrix-Encoding Gene Diversity of 624 Influenza A/H3N2 Genomes Does Not Show Association with Impaired Viral Detection by Commercialized qPCR Assays

Lorlane Le Targa [1,2,3], Houmadi Hikmat [1,2], Céline Boschi [1,2,4], Bernard La Scola [1,2,4,*] and Philippe Colson [1,2,4,*]

1 IHU Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France
2 Microbes Evolution Phylogeny and Infections (MEPHI), Institut de Recherche pour le Développement (IRD), Aix-Marseille University, 27 Boulevard Jean Moulin, 13005 Marseille, France
3 Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France
4 Assistance Publique-Hôpitaux de Marseille (AP-HM), 264 rue Saint-Pierre, 13005 Marseille, France
* Correspondence: bernard.la-scola@univ-amu.fr (B.L.S.); philippe.colson@univ-amu.fr (P.C.); Tel.: +33-413-732-401 (P.C.)

Abstract: As for the case of SARS-CoV-2, genome sequencing of influenza viruses is of potential interest to raise and address virological issues. Recently, false-negativity of real-time reverse transcription-PCR (qPCR) assays that detect influenza A/H3N2 virus RNA were reported and associated with two mutations (A37T and C161T) in the Matrix-encoding (M1) gene located on viral segment 7. This triggered a national alert in France. The present study sought to assess the association between the presence of these mutations and potential false negative results of influenza A/H3N2 virus RNA detection by commercialized qPCR assays at the clinical virology laboratory of our university hospitals in southern France. This study focused on the genetic diversity in the M1 gene and segment 7 of 624 influenza A/H3N2 virus genomes obtained from respiratory samples having tested qPCR-positive with M1 gene-targeting assays in our clinical virology laboratory. A total of 585 among the 624 influenza A/H3N2 virus genomes (93.7%) were of clade 3C.2a1b.2a.2, and 39 (6.3%) were of clade 3C.2a1b.1a. M1 gene substitutions A37T and C161T were both present in 582 (93.3%) genomes, only of clade 3C.2a1b.2a.2. Substitution A37T was present in 621 (99.5%) genomes. Substitution C161T was present in 585 genomes (93.8%), all of clade 3C.2a1b.2a.2. Moreover, 21 other nucleotide positions were mutated in ≥90% of the genomes. The present study shows that A37T/C and C161T mutations, and other mutations in the M1 gene and segment 7, were widely present in influenza A/H3N2 virus genomes recovered from respiratory samples diagnosed qPCR-positive with commercialized assays.

Keywords: influenza A/H3N2 virus; qPCR; diagnosis; matrix gene; M1 gene; genetic diversity

check for updates

Citation: Le Targa, L.; Hikmat, H.; Boschi, C.; La Scola, B.; Colson, P. Matrix-Encoding Gene Diversity of 624 Influenza A/H3N2 Genomes Does Not Show Association with Impaired Viral Detection by Commercialized qPCR Assays. *Viruses* **2022**, *14*, 2683. https://doi.org/10.3390/v14122683

Academic Editors: Leyi Wang and Ganwu Li

Received: 3 November 2022
Accepted: 25 November 2022
Published: 30 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Genomic surveillance of SARS-CoV-2 has led to a tremendous amount of viral genomes available from multiple countries worldwide during the pandemic due to this virus. It has proved to be powerful in differentiating and classifying viral strains, increasing the understanding of viral evolution, getting a better knowledge of viral genotypic and phenotypic characteristics, allowing to design accurate real-time reverse-transcription PCR (qPCR) systems for the diagnosis of infections and variants, and studying the sensitivity of viral lineages and variants to neutralizing antibodies elicited by past infections or vaccine immunization [1,2]. However, such genome sequencing effort has been far more limited for other respiratory viruses.

The emergence and circulation of influenza A/H3N2 viruses carrying mutations associated with false-negative results of qPCR tests have been recently reported in Denmark [3]. This triggered a national alert in France [4]. These mutations are located in the M1 gene

that encodes the viral matrix protein and is located on genome segment 7, which also harbors the M2 gene that encodes a viroporin [5]. The low genetic diversity identified so far for the M1 gene made it a target of several qPCR commercial assays. Nonetheless, Jørgensen et al. correlated false-negative qPCR results, identified by qPCR-positivity with alternative systems, including some targeting other genomic regions, with the presence of mutations in two M1 gene regions targeted by PCR primers [3]. They analyzed eight M1 gene sequences obtained from clinical samples diagnosed as either PCR-positive (n = 4) or negative (n = 4), plus eight genomes of representatives of circulating strains from the Danish National Reference Center and Surveillance Laboratory for Influenza. Here, we aimed to investigate the presence and prevalence of mutations described by Jørgensen et al. in the M1 gene and segment 7 of influenza A/H3N2 virus genomes obtained from respiratory samples tested by qPCR at our university hospital clinical virology laboratory and the potential association of these mutations with false-negative qPCR results.

2. Materials and Methods

The genetic diversity of the M1 gene and segment 7 was analyzed from 624 influenza A/H3N2 genomes (GenBank accession no. OP546682-OP547305 [6,7]) that had been obtained from respiratory samples collected between September 2021 and April 2022 and were sent for diagnosis of respiratory virus infections to our clinical virology laboratory at university hospitals of Marseille, southeastern France. Influenza A virus infections were diagnosed using the FTD Respiratory pathogens 21 (FTD) (Fast Track Diagnosis, Luxembourg), the GeneXpert Xpert Flu/RSV (GeneXpert) (Cepheid, Sunnyvale, CA, USA), or the BioFire FilmArray Respiratory panel 2 plus (BioFire) (Biomérieux, Marcy-l'Etoile, France) assays, which all target the M1 gene (plus two other genes for the GeneXpert assay).

For influenza A virus genome sequencing, viral RNA was extracted from respiratory samples using the EZ1 Virus Mini Kit v2.0 on an EZ1 Advanced XL instrument (Qiagen, Courtaboeuf, France) or the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, United States), following manufacturer's instructions. Then, viral RNA segments were amplified by a procedure adapted from that described by Zhou et al. [8], and next-generation sequencing (NGS) of the amplicons was performed using Illumina technology on a NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA), as previously described [9]. Viral genomes were assembled by mapping NGS reads on the genome of the influenza A/H3N2 virus strain A/New York/392/2004 (GenBank Accession no. NC_007366.1-NC_007373.1) with the CLC Genomics workbench v.7 software [10].

3. Results

Among the 624 genomes of the influenza A/H3N2 virus, 585 (93.7%) were of clade 3C.2a1b.2a.2 and 39 (6.3%) of clade 3C.2a1b.1a [11]. The two substitutions A37T and C161T described by Jørgensen et al. [3] within the M1 gene were both present in 582 (93.3%) genomes, only in those of clade 3C.2a1b.2a.2 (Figure 1).

Nucleotide A37 was mutated in all genomes, substitution A37T being present in 621 (99.5%) genomes and A37C in three genomes obtained from samples collected in December 2021 (n = 1) and March 2022 (n = 2). Substitution C161T was present in 585 genomes (93.8%), all of clade 3C.2a1b.2a.2. In addition, substitution A76C located in the M1 gene region targeted by the probe of the qPCR system described by Jørgensen et al. [3], was present in two genomes of clade 3C.2a1b.1a. Apart from these mutations, 143 other positions of segment 7 were mutated in $\geq$1 genome, including 21 that were mutated in $\geq$90% of the genomes.

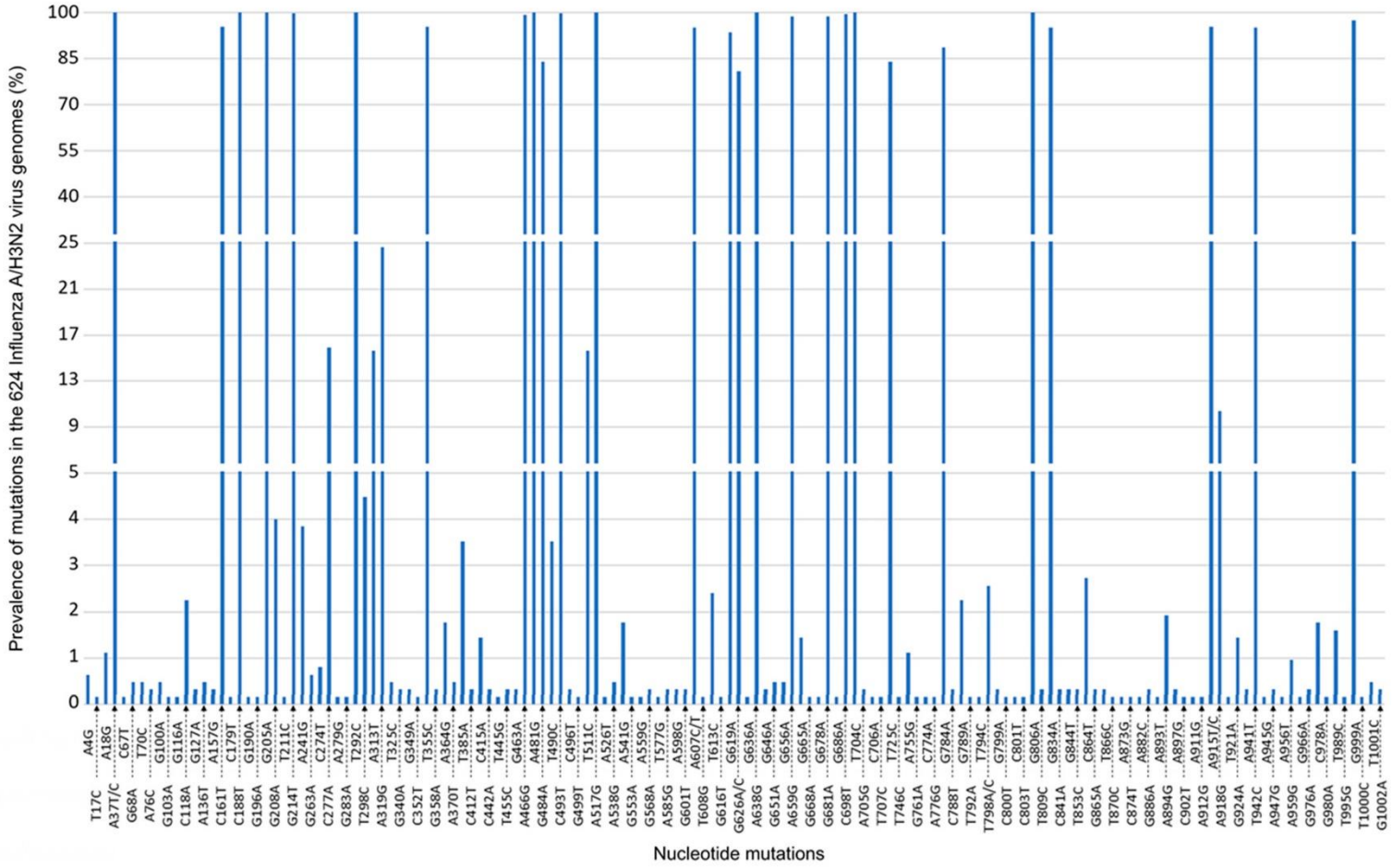


Figure 1. Prevalence of nucleotide mutations in segment 7, including the M1 (matrix) gene of the 624 Influenza A/H3N2 virus genomes analyzed in the present study. The Y-axis is in three parts to better visualize the different levels of prevalence of the mutations.

The 624 influenza A/H3N2 virus genomes had been obtained from respiratory samples diagnosed by qPCR as positive for influenza A virus RNA. In 616 cases (98.7%), the diagnosis of influenza virus infection was made with the FTD assay, which targets the M1 gene of influenza A viruses but whose performance was not assessed by Jørgensen et al. [3]. In seven cases (1.1%), the diagnosis was made with the GeneXpert assay, which was reported by Jørgensen et al. not to cause false negativity for 6 samples [3]. In one case, the diagnosis was made with the BioFire assay, reported by Jørgensen et al. to cause false negativity for two of five samples [3]. As the 624 respiratory samples tested positive with these qPCR assays, this indicates that mutations A37T/C ± C161T in the M1 gene do not seem to be associated with false negative results in our experience. The mean (±standard deviation) qPCR cycle threshold value (Ct) for these samples with the FTD assay was 25.5 ± 6.9 (range, 19.0–34.0). One respiratory sample tested qPCR-positive with the BioFire assay. The viral genome obtained from this sample was of clade 3C.2a1b.2a.2 and harbored both A37T and A161T in the M1 gene, as well as 25 additional mutations in segment 7. In addition, six nasopharyngeal samples among those available in the biobank for storage at −20 °/−80 °C of our laboratory and exhibiting the highest Ct (range, 29–30) as determined by the FTD assay, were retrospectively tested with the BioFire assay. The genome obtained from these samples was of clade 3C.2a1b.2a.2 in four cases and of clade 3C.2a1b.1a in two cases, and all six genomes harbored mutations A37T and A161T. All these six samples tested influenza A/H3N2 RNA-positive with the BioFire assay, indicating that the presence of mutations A37T and A161T does not seem to be associated with false-negative results using this multiplex qPCR assay. Notably, the six viral genomes also harbored between 17 and 27 other mutations in segment 7.

4. Discussion

The present study confirms the presence of both mutations A37T and C161T in the vast majority (99.5%) of the influenza A/H3N2 genomes of clade 3C.2a1b.2a.2 as well as the systematic presence of a mutation at nucleotide 37 for influenza A/H3N2 viruses that circulated since 2021 in Marseille and its geographical area. This study also reports, on a much larger scale than in Jørgensen et al.'s study [3], the majority circulation of clade 3C.2a1b.2a.2 in our geographical region in 2021 and 2022. This viral clade was characterized in 2021 and was the majority influenza A/H3N2 clade during the epidemic that occurred in 2021-2022 in Western developed countries [11]. As the 624 influenza A/H3N2 virus genomes obtained in our laboratory were from respiratory samples diagnosed as influenza A virus-positive by M1 gene targeting-qPCR, mutations A37T and C161T in this gene do not seem to be associated with false negative results in our experience. Nevertheless, the possibility of such qPCR failure or underquantification of viral RNA in the presence of these two mutations, either alone or in combination, or of other mutations in the M1 gene, cannot be excluded. The precise genome regions of the M1 gene targeted by the commercialized qPCR assays are not known, which prevents speculating on the impact of mismatches due to mutations on the performances of detection and quantification of influenza A virus RNA by these assays. Besides, the present study did not directly address whether the recently emerged mutations in the M1 gene might result in the under-quantification of viral RNA; this could rely on the testing of several samples using qPCR assays targeting a single region located either within or outside the segment 7 of the viral genome and harboring or devoid of mutations. To our best knowledge, no country other than France has yet emitted an alert for impaired qPCR performance to detect recently or currently circulating influenza A/H3N2 viruses.

Finally, it is worth noting that the sequencing of the genomes of influenza viruses is currently dramatically underperformed compared to SARS-CoV-2 genomes. According to the GISAID database [12,13], as of 5 October 2022, 141 influenza A/H3N2 virus sequences have been obtained in France from human samples collected since January 2021 and this number is one according to the NCBI Influenza Virus Database sequences [14]. Hence, influenza virus genomic surveillance deserves to be carried out more extensively, in partic-

ular, to ensure the performance of commercialized qPCR tests, but more generally in order to draw from it the information that has shown great interest in the context of the genomic surveillance of SARS-CoV-2 infections [1,2].

Author Contributions: P.C. and B.L.S. designed the study. L.L.T., H.H., C.B. and P.C. provided materials, data, or analysis tools. L.L.T., P.C., C.B. and B.L.S. analyzed the data. P.C. and L.L.T. wrote the first draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR) (Méditerranée-Infection 10-IAHU-03), and by the French Ministry of Higher Education, Research and Innovation (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation) and the French Ministry of Solidarity and Health (Ministère des Solidarités et de la Santé).

Institutional Review Board Statement: The present study has been approved by the ethics committee of the University Hospital Institute (IHU) Méditerranée Infection (N°2022-021) and was registered in the General Data Protection Regulation (GDPR) registry no. ZJ6W5X from Assistance Publique-Hôpitaux de Marseille (AP-HM, Marseille public hospitals) institution.

Informed Consent Statement: The present study has been conducted on influenza A virus genome sequences obtained from respiratory samples sent to our clinical diagnostic laboratory for the purpose of diagnosis of respiratory viruses. No supplementary clinical sample was used that was not sent to our laboratory in this setting. All clinical samples and data have been fully anonymized.

Data Availability Statement: Viral genomes analyzed in the present study have been deposited in the NCBI GenBank sequence database under Accession nos. OP546682-OP547305 [6,7].

Acknowledgments: We are thankful to Raphael Tola, Laurence Thomas, Florine Pelisson, and Claudia Andrieu for their technical help.

Conflicts of Interest: Lorlane Le Targa is an employee of Biosellal (Dardilly, France). The other authors have no conflict of interest to declare relative to the present study.

References

1. Li, J.; Lai, S.; Gao, G.F.; Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **2021**, *600*, 408–418. [CrossRef] [PubMed]
2. Lemey, P.; Ruktanonchai, N.; Hong, S.L.; Colizza, V.; Poletto, C.; Van den Broeck, F.; Gill, M.S.; Ji, X.; Levasseur, A.; Oude Munnink, B.B.; et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* **2021**, *595*, 713–717. [CrossRef] [PubMed]
3. Jørgensen, R.L.; Lerche, C.J.; Pedersen, M.S.; Kirkby, N.S.; Botnen, A.B.; Trebbien, R.; Nilsson-Møller, S.; Pinholt, M.; Nielsen, A.C.; Westh, H.; et al. Emergence of circulating influenza A H3N2 viruses with genetic drift in the matrix gene: Be alert of false-negative test results. *APMIS* **2022**, *130*, 612–617. [CrossRef] [PubMed]
4. Possible Baisse de Sensibilité de Techniques de Biologie Moléculaire Permettant la Détection des Virus Grippaux de Type A Sur certains Virus de Sous-Type H3N2 Ayant Circulé en France lors de la Dernière Epidémie Saisonnière. Société Fr Microbiol 2022. Available online: https://www.sfm-microbiologie.org/2022/09/07/communique-cnr-infections-respiratoires/ (accessed on 21 September 2022).
5. Chauhan, R.P.; Gordon, M.L. An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates. *Virus Genes* **2022**, *58*, 255–269. [CrossRef] [PubMed]
6. GenBank, the genetic sequence database of the National Institutes of Health, United States of America. Available online: https://www.ncbi.nlm.nih.gov/genbank/ (accessed on 27 October 2022).
7. Sayers, E.W.; Cavanaugh, M.; Clark, K.; Pruitt, K.D.; Schoch, C.L.; Sherry, S.T.; Karsch-Mizrachi, I. GenBank. *Nucleic Acids Res.* **2021**, *50*, D161–D164. [CrossRef] [PubMed]
8. Zhou, B.; Donnelly, M.E.; Scholes, D.T.; St George, K.; Hatta, M.; Kawaoka, Y.; Wentworth, D.E. Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses. *J. Virol.* **2009**, *83*, 10309–10313. [CrossRef] [PubMed]
9. Colson, P.; Fournier, P.E.; Chaudet, H.; Delerce, J.; Giraud-Gatineau, A.; Houhamdi, L.; Andrieu, C.; Brechard, L.; Bedotto, M.; Prudent, E.; et al. Analysis of SARS-CoV-2 Variants from 24,181 Patients Exemplifies the Role of Globalization and Zoonosis in Pandemics. *Front. Microbiol.* **2022**, *12*, 786233. [CrossRef] [PubMed]

10. QIAGEN CLC Genomics Workbench. Available online: https://digitalinsights.qiagen.com/ (accessed on 27 October 2022).
11. Melidou, A.; Ködmön, C.; Nahapetyan, K.; Kraus, A.; Alm, E.; Adlhoch, C.; Mooks, P.; Dave, N.; Carvalho, C.; Meslé, M.M.; et al. Influenza returns with a season dominated by clade 3C.2a1b.2a.2 A(H3N2) viruses, WHO European Region, 2021/22. *Eurosurveillance* **2022**, *27*, 2200255. [CrossRef] [PubMed]
12. Global Initiative on Sharing Avian Influenza Data. Available online: https://gisaid.org/ (accessed on 27 October 2022).
13. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall.* **2017**, *1*, 33–46. [CrossRef] [PubMed]
14. The Influenza Virus Database of the National Center for Biotechnology Information, National Institutes of Health, United States of America. Available online: https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database (accessed on 27 October 2022).

Conclusion

La pandémie de la Covid-19 a montré qu'il était nécessaire de faire évoluer la recherche sur les virus respiratoires. Le suivi des épidémies par le biais des eaux usées fait partie des pistes intéressantes pour mieux comprendre et prévenir les futures pandémies. Comme montré dans ce manuscrit, la détection des virus respiratoires dans les eaux usées avait déjà été réalisée avant cette pandémie pour certains virus. Pour les virus influenza, le RSV, les HAdVs, le HBoV et le SARS-CoV, une ou plusieurs études antérieures au SARS-CoV-2 existent [64], [65], [66], [67], [68]. Mais ces études restent rares ou ne visent pas spécifiquement les sous-types respiratoires, comme dans le cas des HAdVs. La détection dans les eaux usées des virus respiratoires n'avait jamais vraiment été mise en place dans le cadre d'une routine de surveillance.

La découverte du SARS-CoV-2 a provoqué une forte augmentation de la détection et du séquençage des virus respiratoires dans les eaux usées. Des études sur la corrélation entre taux de virus dans les eaux usées et le nombre de patients infectés ont pu être réalisées, ainsi que des articles sur l'utilisation des eaux usées pour mieux prédire le début d'une épidémie[36], [69], [70], [71], [72], [73].

Dans notre cas, on a pu voir qu'il était possible de détecter et de séquencer le SARS-CoV-2 dans des eaux usées d'avion, permettant la détection de variants rares en France au moment de l'étude. De plus, nous avons pu montrer que la mesure sanitaire n'était pas 100% efficace contre la propagation des virus respiratoires à travers le monde [25].

Dans le cadre du suivi de l'évolution des virus respiratoires, nous avons aussi mis en avant dans ce manuscrit que des lacunes existaient dans l'analyse des génomes complets de la grippe. Il a été possible, pour la grippe A/H3N2, de voir des événements de réassortiment inter-sous-type sur une période de 2017 à 2022 par l'analyse de plus de 800 génomes, ce qui est une première en France. Pour la suite, il serait intéressant de mettre en place une étude similaire sur la grippe dans les eaux usées, afin de voir si ce type de suivi génomique poussé est applicable sur des génomes potentiellement dégradés ou en faible quantité dans des matrices complexes. De plus, nous avons aussi pu montrer que le séquençage à grande échelle chez les patients pouvait jouer un rôle dans l'étude des potentiels faux négatifs [74].

Entre l'essor du NGS et des systèmes de PCR-multiplexe, il est de plus en plus facile d'obtenir des génomes complets. Mais cela n'est pas suffisant dans le cadre d'une surveillance génomique, puisqu'aujourd'hui encore la majorité des études génomiques sont dirigées vers seulement un ou deux gènes.

Des analyses récentes tendent à montrer que nous ne savons pas tous sur les virus pouvant causer une pandémie. Une étude récente basée sur de la paléo-sérologie suppose par exemple que la pandémie de « grippe russe « de 1889-1890 a pu être causée par un coronavirus saisonnier et non pas par le virus de la grippe [75].

De plus, il est possible que la prochaine pandémie de virus respiratoires soit causée par un virus qui n'existe pas encore ou qui n'a pas encore été découvert. Dans ce sens, il est important de s'intéresser à la surveillance par séquençage métagénomique de potentiels nouveaux virus humains que ce soit chez les patients ou dans les eaux usées.

Il est aussi possible, que la prochaine pandémie provienne d'un virus connu déjà présent chez d'autres espèces. Dans le cadre de la COVID-19, on a notamment pu voir que cette dernière pouvait provenir des chauves-souris ou des pangolins. Dans le cadre de la grippe, les virus de type aviaires sont très surveillés, notamment leur passage chez d'autres espèces augmentant le risque qu'ils finissent par infecter les humains. Mais il est difficile de réaliser des prélèvements chez les animaux sauvages ou de tester tous les animaux d'un élevage. C'est pour cela qu'il est important de nos jours de s'intéresser à la surveillance et au séquençage des virus respiratoires provenant de différentes espèces et l'analyse des eaux usées peut-être un outil important dans cette étude.

Bibliographie

[1] F. Krammer *et al.*, "Influenza," *Nature Reviews. Disease Primers*, vol. 4, no. 1, 2018, doi: 10.1038/s41572-018-0002-y.

[2] R. L. Hodinka, "Respiratory RNA Viruses," *Microbiol Spectr*, vol. 4, no. 4, Aug. 2016, doi: 10.1128/microbiolspec.DMIH2-0028-2016.

[3] "SARS and MERS: recent insights into emerging coronaviruses | Nature Reviews Microbiology." Accessed: Aug. 07, 2024. [Online]. Available: https://www.nature.com/articles/nrmicro.2016.81

[4] N. Zhu *et al.*, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[5] S. E. Jacobs, D. M. Lamson, K. St. George, and T. J. Walsh, "Human Rhinoviruses," *Clin Microbiol Rev*, vol. 26, no. 1, pp. 135–162, Jan. 2013, doi: 10.1128/CMR.00077-12.

[6] "Rhinoviruses and Respiratory Enteroviruses: Not as Simple as ABC - PMC." Accessed: Aug. 07, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728576/

[7] "Human bocavirus: Current knowledge and future challenges - PubMed." Accessed: Aug. 07, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27818586/

[8] "Clinical Overview of Adenovirus | Adenovirus | CDC." Accessed: Aug. 07, 2024. [Online]. Available: https://www.cdc.gov/adenovirus/hcp/clinical-overview/?CDC_AAref_Val=https://www.cdc.gov/adenovirus/hcp/clinical-overview.html

[9] "Economic burden of seasonal influenza in the United States - ScienceDirect." Accessed: Oct. 03, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0264410X18306777?via%3Dihub

[10] A. M. Fendrick, A. S. Monto, B. Nightengale, and M. Sarnes, "The economic burden of non-influenza-related viral respiratory tract infection in the United States," *Arch Intern Med*, vol. 163, no. 4, pp. 487–494, Feb. 2003, doi: 10.1001/archinte.163.4.487.

[11] "WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data." Accessed: Dec. 08, 2022. [Online]. Available: https://covid19.who.int/table

[12] H. Brüssow, "The beginning and ending of a respiratory viral pandemic-lessons from the Spanish flu," *Microb Biotechnol*, vol. 15, no. 5, pp. 1301–1317, May 2022, doi: 10.1111/1751-7915.14053.

[13] C. Peteranderl, S. Herold, and C. Schmoldt, "Human Influenza Virus Infections," *Semin Respir Crit Care Med*, vol. 37, no. 4, pp. 487–500, Aug. 2016, doi: 10.1055/s-0036-1584801.

[14] "European Influenza Surveillance Network (EISN)." Accessed: Oct. 03, 2024. [Online]. Available: https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/eisn

[15] "Survey on the implementation of integrated surveillance of respiratory viruses with pandemic potential." Accessed: Oct. 14, 2024. [Online]. Available:

https://www.ecdc.europa.eu/en/publications-data/survey-implementation-integrated-surveillance-respiratory-viruses-pandemic
[16] réseau S. Université INSERM/Sorbonne, "Réseau Sentinelles > France > Le réseau Sentinelles." Accessed: Oct. 14, 2024. [Online]. Available: https://www.sentiweb.fr/france/fr/?page=presentation
[17] C. Mishra, S. Meena, J. K. Meena, S. Tiwari, and P. Mathur, "Detection of three pandemic causing coronaviruses from non-respiratory samples: systematic review and meta-analysis," *Sci Rep*, vol. 11, no. 1, p. 16131, Aug. 2021, doi: 10.1038/s41598-021-95329-4.
[18] O. J. Kweon, Y. K. Lim, H. R. Kim, T.-H. Kim, and M.-K. Lee, "Fecal Respiratory Viruses in Acute Viral Respiratory Infection and Nasopharyngeal Diarrheal Viruses in Acute Viral Gastroenteritis: Clinical Impact of Ectopic Viruses Is Questionable," *J Microbiol Biotechnol*, vol. 28, no. 3, pp. 465–472, Mar. 2018, doi: 10.4014/jmb.1707.07058.
[19] S. J. Tozer *et al.*, "Detection of human bocavirus in respiratory, fecal, and blood samples by real-time PCR," *Journal of Medical Virology*, vol. 81, no. 3, pp. 488–493, 2009, doi: 10.1002/jmv.21409.
[20] K. S. Cheung *et al.*, "Gastrointestinal Manifestations of SARS-CoV-2 Infection and Virus Load in Fecal Samples From a Hong Kong Cohort: Systematic Review and Meta-analysis," *Gastroenterology*, vol. 159, no. 1, pp. 81–95, Jul. 2020, doi: 10.1053/j.gastro.2020.03.065.
[21] "Viruses in wastewater: occurrence, abundance and detection methods | Elsevier Enhanced Reader." Accessed: Dec. 23, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S0048969720344399?token=407EC9807AAC66AA8ECFDD1FC4FD5F351F10C042EE767AA5F34470972CEE584B6AC61D5C7C2D30031D67230C6BE251C6&originRegion=eu-west-1&originCreation=20221223130026
[22] K. Bibby and J. Peccia, "Identification of viral pathogen diversity in sewage sludge by metagenome analysis," *Environ Sci Technol*, vol. 47, no. 4, pp. 1945–1951, Feb. 2013, doi: 10.1021/es305181x.
[23] J. Gonçalves *et al.*, "Centralized and decentralized wastewater-based epidemiology to infer COVID-19 transmission - A brief review," *One Health*, vol. 15, p. 100405, Dec. 2022, doi: 10.1016/j.onehlt.2022.100405.
[24] A. Albastaki *et al.*, "First confirmed detection of SARS-COV-2 in untreated municipal and aircraft wastewater in Dubai, UAE: The use of wastewater based epidemiology as an early warning tool to monitor the prevalence of COVID-19," *Science of The Total Environment*, vol. 760, p. 143350, Mar. 2021, doi: 10.1016/j.scitotenv.2020.143350.
[25] L. Le Targa *et al.*, "SARS-CoV-2 Testing of Aircraft Wastewater Shows That Mandatory Tests and Vaccination Pass before Boarding Did Not Prevent Massive Importation of Omicron Variant into Europe," *Viruses*, vol. 14, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/v14071511.
[26] "Molecular Detection of Respiratory Viruses - Clinics in Laboratory Medicine." Accessed: Aug. 07, 2024. [Online]. Available: https://www.labmed.theclinics.com/article/S0272-2712(13)00016-4/fulltext
[27] "Annonce des recommandations relatives à la composition des vaccins antigrippaux pour la saison grippale 2024-2025 dans l'hémisphère Nord." Accessed: Oct. 03, 2024. [Online]. Available: https://www.who.int/fr/news/item/23-02-2024-recommendations-announced-for-influenza-vaccine-composition-for-the-2024-2025-northern-hemisphere-influenza-season

[28] A. K. Shetty, E. Treynor, D. W. Hill, K. M. Gutierrez, A. Warford, and E. J. Baron, "Comparison of conventional viral cultures with direct fluorescent antibody stains for diagnosis of community-acquired respiratory virus infections in hospitalized children," *Pediatr Infect Dis J*, vol. 22, no. 9, pp. 789–794, Sep. 2003, doi: 10.1097/01.inf.0000083823.43526.97.
[29] R. Kalendar *et al.*, "Universal whole-genome Oxford nanopore sequencing of SARS-CoV-2 using tiled amplicons," *Sci Rep*, vol. 13, no. 1, p. 10334, Jun. 2023, doi: 10.1038/s41598-023-37588-x.
[30] B. Zhou *et al.*, "Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses," *Journal of Virology*, vol. 83, no. 19, pp. 10309–10313, Oct. 2009, doi: 10.1128/jvi.01109-09.
[31] B. Zhou *et al.*, "Universal influenza B virus genomic amplification facilitates sequencing, diagnostics, and reverse genetics," *J Clin Microbiol*, vol. 52, no. 5, pp. 1330–1337, May 2014, doi: 10.1128/JCM.03265-13.
[32] H. Hikmat *et al.*, "Sequencing and characterization of human bocavirus genomes from patients diagnosed in Southern France between 2017 and 2022," *Journal of Medical Virology*, vol. 96, no. 6, p. e29706, 2024, doi: 10.1002/jmv.29706.
[33] J. Mahony *et al.*, "Development of a respiratory virus panel test for detection of twenty human respiratory viruses by use of multiplex PCR and a fluid microbead-based assay," *J Clin Microbiol*, vol. 45, no. 9, pp. 2965–2970, Sep. 2007, doi: 10.1128/JCM.02436-06.
[34] I. Yasuda *et al.*, "Respiratory virus detection in the upper respiratory tract of asymptomatic, community-dwelling older people," *BMC Infect Dis*, vol. 22, no. 1, p. 411, Apr. 2022, doi: 10.1186/s12879-022-07355-w.
[35] M. L. García-García *et al.*, "Human bocavirus detection in nasopharyngeal aspirates of children without clinical symptoms of respiratory infection," *Pediatr Infect Dis J*, vol. 27, no. 4, pp. 358–360, Apr. 2008, doi: 10.1097/INF.0b013e3181626d2a.
[36] G. A *et al.*, "SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions," *The Science of the total environment*, vol. 804, Jan. 2022, doi: 10.1016/j.scitotenv.2021.150151.
[37] M. Sutton *et al.*, "Detection of SARS-CoV-2 B.1.351 (Beta) Variant through Wastewater Surveillance before Case Detection in a Community, Oregon, USA," *Emerg Infect Dis*, vol. 28, no. 6, pp. 1101–1109, Jun. 2022, doi: 10.3201/eid2806.211821.
[38] S. Karthikeyan *et al.*, "Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission," *Nature*, vol. 609, no. 7925, pp. 101–108, Sep. 2022, doi: 10.1038/s41586-022-05049-6.
[39] C. R. Wells *et al.*, "Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak," *Proc Natl Acad Sci U S A*, vol. 117, no. 13, pp. 7504–7509, Mar. 2020, doi: 10.1073/pnas.2002616117.
[40] M. Chinazzi *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, Apr. 2020, doi: 10.1126/science.aba9757.
[41] A. S. Gonzalez-Reiche *et al.*, "Introductions and early spread of SARS-CoV-2 in the New York City area," *Science*, vol. 369, no. 6501, pp. 297–301, Jul. 2020, doi: 10.1126/science.abc1917.

[42] “Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study - PMC.” Accessed: Oct. 03, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7801817/
[43] D. Tian, Y. Sun, H. Xu, and Q. Ye, “The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant,” *J Med Virol*, vol. 94, no. 6, pp. 2376–2383, Jun. 2022, doi: 10.1002/jmv.27643.
[44] L. Chen, J. Lou, Y. Bai, and M. Wang, “COVID-19 Disease With Positive Fecal and Negative Pharyngeal and Sputum Viral Tests,” *Am J Gastroenterol*, vol. 115, no. 5, p. 790, May 2020, doi: 10.14309/ajg.0000000000000610.
[45] G. Medema, L. Heijnen, G. Elsinga, R. Italiaander, and A. Brouwer, “Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands,” *Environ Sci Technol Lett*, vol. 7, no. 7, pp. 511–516, Jul. 2020, doi: 10.1021/acs.estlett.0c00357.
[46] N. Wurtz *et al.*, “Viral RNA in City Wastewater as a Key Indicator of COVID-19 Recrudescence and Containment Measures Effectiveness,” *Front Microbiol*, vol. 12, p. 664477, 2021, doi: 10.3389/fmicb.2021.664477.
[47] G. La Rosa *et al.*, “SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring,” *Sci Total Environ*, vol. 750, p. 141711, Jan. 2021, doi: 10.1016/j.scitotenv.2020.141711.
[48] N. Wurtz *et al.*, “Monitoring the Circulation of SARS-CoV-2 Variants by Genomic Analysis of Wastewater in Marseille, South-East France,” *Pathogens*, vol. 10, no. 8, p. 1042, Aug. 2021, doi: 10.3390/pathogens10081042.
[49] A. Crits-Christoph *et al.*, “Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants,” *mBio*, vol. 12, no. 1, pp. e02703-20, Jan. 2021, doi: 10.1128/mBio.02703-20.
[50] C. S. Silva *et al.*, “Temporal dynamics of SARS-CoV-2 genome and detection of variants of concern in wastewater influent from two metropolitan areas in Arkansas,” *Sci Total Environ*, vol. 849, p. 157546, Nov. 2022, doi: 10.1016/j.scitotenv.2022.157546.
[51] W. Ahmed *et al.*, “Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: a surveillance tool for assessing the presence of COVID-19 infected travellers,” *J Travel Med*, vol. 27, no. 5, p. taaa116, Aug. 2020, doi: 10.1093/jtm/taaa116.
[52] W. Ahmed *et al.*, “Detection of the Omicron (B.1.1.529) variant of SARS-CoV-2 in aircraft wastewater,” *Sci Total Environ*, vol. 820, p. 153171, May 2022, doi: 10.1016/j.scitotenv.2022.153171.
[53] “Review of Non-bacterial Infections in Respiratory Medicine: Viral Pneumonia - ScienceDirect.” Accessed: Aug. 07, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1579212915003067
[54] “Evolution of Influenza A Virus by Mutation and Re-Assortment.” Accessed: Oct. 03, 2024. [Online]. Available: https://www.mdpi.com/1422-0067/18/8/1650
[55] H. Kim, R. G. Webster, and R. J. Webby, “Influenza Virus: Dealing with a Drifting and Shifting Pathogen,” *Viral Immunol*, vol. 31, no. 2, pp. 174–183, Mar. 2018, doi: 10.1089/vim.2017.0141.
[56] G. J. D. Smith *et al.*, “Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic,” *Nature*, vol. 459, no. 7250, pp. 1122–1125, Jun. 2009, doi: 10.1038/nature08182.
[57] M. Hayati, B. Sobkowiak, J. E. Stockdale, and C. Colijn, “Phylogenetic identification of influenza virus candidates for seasonal vaccines,” *Sci Adv*, vol. 9, no. 44, p. eabp9185, Nov. 2023, doi: 10.1126/sciadv.abp9185.

[58] “Influenza virus characterization - Summary Europe, February 2023.” Accessed: Oct. 03, 2024. [Online]. Available: https://www.ecdc.europa.eu/en/publications-data/influenza-virus-characterization-summary-europe-february-2023
[59] C. Adlhoch and R. Pebody, “What to expect for the influenza season 2020/21 with the ongoing COVID-19 pandemic in the World Health Organization European Region,” *Eurosurveillance*, vol. 25, no. 42, p. 2001816, Oct. 2020, doi: 10.2807/1560-7917.ES.2020.25.42.2001816.
[60] C. Adlhoch *et al.*, “Very little influenza in the WHO European Region during the 2020/21 season, weeks 40 2020 to 8 2021,” *Eurosurveillance*, vol. 26, no. 11, p. 2100221, Mar. 2021, doi: 10.2807/1560-7917.ES.2021.26.11.2100221.
[61] R. P. Chauhan and M. L. Gordon, “An overview of influenza A virus genes, protein functions, and replication cycle highlighting important updates,” *Virus Genes*, vol. 58, no. 4, pp. 255–269, Aug. 2022, doi: 10.1007/s11262-022-01904-w.
[62] R. L. Jørgensen *et al.*, “Emergence of circulating influenza A H3N2 viruses with genetic drift in the matrix gene: be alert of false-negative test results,” *APMIS*, vol. 130, no. 10, pp. 612–617, Oct. 2022, doi: 10.1111/apm.13262.
[63] O. Vong, “Possible baisse de sensibilité de techniques de biologie moléculaire permettant la détection des virus grippaux de type A sur certains virus de sous-type H3N2 ayant circulé en France lors de la dernière épidémie saisonnière,” Société Française de Microbiologie. Accessed: Oct. 03, 2024. [Online]. Available: https://www.sfm-microbiologie.org/2022/09/07/communique-cnr-infections-respiratoires/
[64] H. Ando, W. Ahmed, R. Iwamoto, Y. Ando, S. Okabe, and M. Kitajima, “Impact of the COVID-19 pandemic on the prevalence of influenza A and respiratory syncytial viruses elucidated by wastewater-based epidemiology,” *Sci Total Environ*, vol. 880, p. 162694, Jul. 2023, doi: 10.1016/j.scitotenv.2023.162694.
[65] O. Blinkova *et al.*, “Frequent detection of highly diverse variants of cardiovirus, cosavirus, bocavirus, and circovirus in sewage samples collected in the United States,” *J Clin Microbiol*, vol. 47, no. 11, pp. 3507–3513, Nov. 2009, doi: 10.1128/JCM.01062-09.
[66] M. Iaconelli, M. Divizia, S. Della Libera, P. Di Bonito, and G. La Rosa, “Frequent Detection and Genetic Diversity of Human Bocavirus in Urban Sewage Samples,” *Food Environ Virol*, vol. 8, no. 4, pp. 289–295, Dec. 2016, doi: 10.1007/s12560-016-9251-7.
[67] X.-W. Wang *et al.*, “Excretion and detection of SARS coronavirus and its nucleic acid from digestive system,” *World J Gastroenterol*, vol. 11, no. 28, pp. 4390–4395, Jul. 2005, doi: 10.3748/wjg.v11.i28.4390.
[68] O. Osuolale and A. Okoh, “Incidence of human adenoviruses and Hepatitis A virus in the final effluent of selected wastewater treatment plants in Eastern Cape Province, South Africa,” *Virol J*, vol. 12, p. 98, Jun. 2015, doi: 10.1186/s12985-015-0327-z.
[69] W. Ahmed *et al.*, “Occurrence of multiple respiratory viruses in wastewater in Queensland, Australia: Potential for community disease surveillance,” *Sci Total Environ*, vol. 864, p. 161023, Mar. 2023, doi: 10.1016/j.scitotenv.2022.161023.
[70] T. de Melo, G. Islam, D. B. D. Simmons, J.-P. Desaulniers, and A. E. Kirkwood, “An alternative method for monitoring and interpreting influenza A in communities using wastewater surveillance,” *Front Public Health*, vol. 11, p. 1141136, Jul. 2023, doi: 10.3389/fpubh.2023.1141136.
[71] “Respiratory Syncytial Virus (RSV) RNA in Wastewater Settled Solids Reflects RSV Clinical Positivity Rates | Environmental Science & Technology Letters.” Accessed:

Aug. 07, 2024. [Online]. Available: https://pubs.acs.org/doi/10.1021/acs.estlett.1c00963
[72] M. Koureas *et al.*, "Wastewater Levels of Respiratory Syncytial Virus Associated with Influenza-like Illness Rates in Children—A Case Study in Larissa, Greece (October 2022–January 2023)," *International Journal of Environmental Research and Public Health*, vol. 20, no. 6, Art. no. 6, Jan. 2023, doi: 10.3390/ijerph20065219.
[73] A. Robotto *et al.*, "Wastewater-based SARS-CoV-2 environmental monitoring for Piedmont, Italy," *Environ Res*, vol. 203, p. 111901, Jan. 2022, doi: 10.1016/j.envres.2021.111901.
[74] L. T. L, H. H, B. C, L. S. B, and C. P, "Matrix-Encoding Gene Diversity of 624 Influenza A/H3N2 Genomes Does Not Show Association with Impaired Viral Detection by Commercialized qPCR Assays," *Viruses*, vol. 14, no. 12, Nov. 2022, doi: 10.3390/v14122683.
[75] L. Ramassy *et al.*, "Paleoserology points to Coronavirus as possible causative pathogens of the 'Russian flu,'" *Microb Biotechnol*, vol. 15, no. 7, pp. 1943–1945, Apr. 2022, doi: 10.1111/1751-7915.14058.

Annexes

Article n°6 :
Influenzas at the 2021 Grand Magal of Touba and possible spread to rural villages in South Senegal

Ndiaw Goumballa, Fatou Samba Diouf, Mamadou Beye, Masse Sambou, Hubert Bassène, Mamadou Dieng, Adama Aïdara, **Lorlane Le Targa**, Philippe Colson, Philippe Gautret, Cheikh Sokhna (Publier dans International Journal of Infectious Diseases)

International Journal of Infectious Diseases 141 (2024) 106952

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Short Communication

Influenza at the 2021 Grand Magal of Touba and possible spread to rural villages in South Senegal - a genomic epidemiological study

Ndiaw Goumballa[1,2,3], Fatou Samba Diouf[2,3,4], Mamadou Beye[3], Masse Sambou[2], Hubert Bassène[2], Mamadou Dieng[5], Adama Aïdara[6], Lorlane L.E. Targa[3,4], Philippe Colson[3,4], Philippe Gautret[1,3], Cheikh Sokhna[1,2,3,*]

[1] Aix Marseille Univ, IRD, AP-HM, SSA, VITROME, Marseille, France
[2] VITROME, Campus International IRD-UCAD de l'IRD, Dakar, Senegal
[3] IHU-Méditerranée Infection, Marseille, France
[4] Aix Marseille Univ, IRD, AP-HM, SSA, MEPHI, Marseille, France
[5] Région Médicale de Diourbel, Diourbel, Senegal
[6] Centre de Santé de Mbacké, Touba, Senegal

ARTICLE INFO

Article history:
Received 30 November 2023
Revised 26 January 2024
Accepted 26 January 2024

Keywords:
Mass gathering
Influenza A
Sequencing
Epidemiology

ABSTRACT

Objectives: Influenza is frequent among pilgrims participating in the Grand Magal de Touba (GMT), in Senegal, with a potential to spread to contacts when they return home.
Methods: Ill pilgrims consulting at a health care center in Mbacké city close to Touba during the 2021 GMT, pilgrims returning to Dielmo and Ndiop villages, and patients who did not travel to Touba and consulted at health care centers in these two villages in 2021 were tested for the influenza virus by polymerase chain reaction on nasopharyngeal samples. Next-generation sequencing and comparative and phylogenetic analyses of influenza A virus genomes were performed.
Results: A total of 62 of 685 patients tested positive for influenza A virus, including 34 of 53 who were consulted in Mbacké in late September, six of 129 pilgrims who returned home in early October, and 20 of 42 villagers from October 3 to 29. A total of 27 genomes were obtained. Four clusters were observed based on the phylogenetic analyses, suggesting that Mbacké patients and returned pilgrims may have shared closely related viral strains with patients inhabiting the villages who did not participate in the GMT.
Conclusions: Villagers in Ndiop and Dielmo may have been infected with viral strains originating from the GMT and possibly imported by pilgrims who returned from the GMT.
© 2024 Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

Introduction

The Grand Magal of Touba (GMT) is the largest religious mass gathering in Senegal. An estimated 4-5 million pilgrims participate in the event each year.

Early works showed that respiratory tract infection (RTI) symptoms are among the most common reasons for ill pilgrims to present to health care centers during the GMT [1,2]. In addition, cohort studies conducted in pilgrims from two rural villages in South Senegal evidenced rates of 41% for RTI symptoms and acquisition of respiratory viruses after participating in this event [3–5]. Finally, studies conducted in pilgrims consulting health care centers for RTI during the GMT allowed evidencing outbreaks of influenza related to this mass gathering, with up to 64% of individuals testing positive for the influenza A virus [6]. Therefore, the present study aimed to test the hypothesis of the introduction of the influenza virus into two villages in South Senegal by pilgrims returning from the September 26, 2021 GMT, with a subsequent spread of the virus.

Materials and methods

Patients and samples

Data were collected in (i) the Mbacké health care center (near Touba) from 53 patients with cough from September 24 to 28 2021, (ii) Dielmo and Ndiop health care centers located in the south of

* Corresponding author: Tel: +33 4 13 73 24 36.
E-mail address: cheikh.sokhna@ird.fr (C. Sokhna).

https://doi.org/10.1016/j.ijid.2024.01.019
1201-9712/© 2024 Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

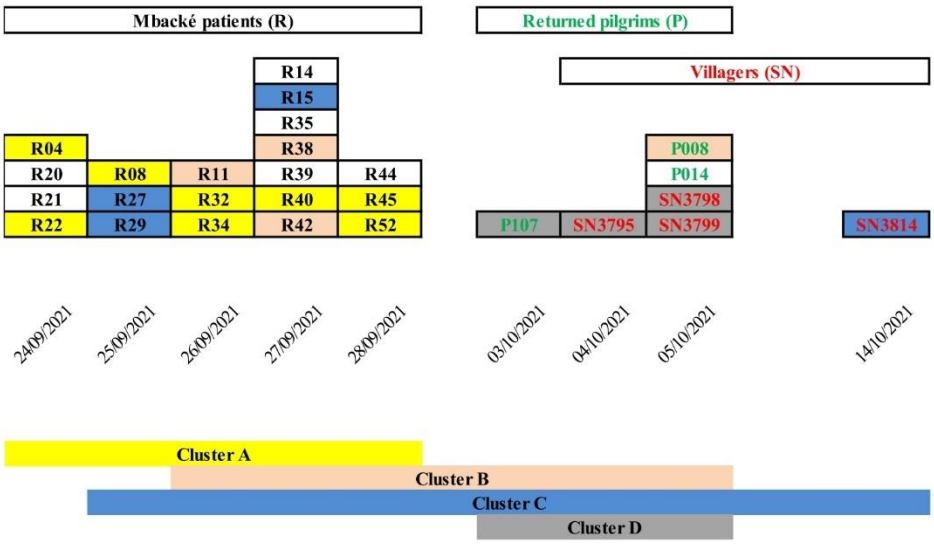


Figure 1. Clusters of influenza cases based on phylogenetic analysis of genomes.

Table 1
Characteristics of the participants.

	Patients with respiratory tract infection symptoms at Mbacké health care center during the GMT N = 53	Pilgrims from villages before the GMT N = 129	Pilgrims from villages after the GMT N = 129	Patients with fever in the villages (January-December 2021) N = 374
Median age (years)	13	23	23	NA
Male gender	37 (69.8)	61 (47.3)	61 (47.3)	NA
Respiratory symptoms	53 (100.0)	17 (13.2)	51 (39.5)	NA
Cough	53 (100.0)	11 (8.5)	36 (27.9)	NA
Rhinitis	44 (83.0)	6 (4.7)	21 (16.3)	NA
Fever	30 (56.6)	0 (0)	13 (10.1	374 (100.0)
Positive for influenza A by PCR	34 (64.2)	0 (0)	8 (6.2)	20 (5.3)
Positive for influenza B by PCR	0 (0)	0 (0)	0 (0)	0 (0)
Number of positive samples that underwent virus genome sequencing	20/34	0/0	3/8	4/20

GMT, Grand Magal de Touba; PCR, polymerase chain reaction.

Senegal from January 1 to December 19, 2021 for 374 patients who presented with fever ≥37.5°C to detect the first cases of influenza in the villages; and (iii) 129 GMT pilgrims from these two villages who were systematically sampled before leaving for Touba (September 14 to 18, 2021) and upon return (October 5 to 8, 2021).

Virological analyses (Supplementary material)

The samples were tested for influenza A and B viruses by polymerase chain reaction (PCR), and next-generation sequencing of the influenza A virus genomes was performed on RNA extracts, with a PCR cycle threshold value below 32.

Results

Prevalence of respiratory symptoms and influenza virus infections

In the patients with cough who presented to the Mbacké health care center, the median age was 13 years (range 0-75 years) and 70% were men (Table 1). Apart from cough (100%), the most frequent symptoms were rhinitis (83%) and fever (57%). A total of 34 of 53 (62%) patients were positive for the influenza A virus and none for the influenza B virus. In pilgrims returning from the GMT to the villages, the median age was 23 years (range 0-76 years) and 47% were men. A total of 51 of 129 (40%) patients reported respiratory symptoms, including cough (36 [28%]) and rhinitis (21 [16%]). Eight of 129 (6%) pilgrims were positive for the influenza A virus and none for the influenza B virus. Finally, in patients with fever who presented to the health care center in the villages, 20 of 374 (5%) patients tested positive for the influenza A virus and none for the influenza B virus.

Chronology of positivity of samples for influenza A virus by polymerase chain reaction

From January 1 October 1, 2021, no influenza case was detected in ill people in the villages of Dielmo and Ndiop. In the 34 of 53 patients positive for influenza A at the Mbacké health care center, 10 of 16, three of six, seven of nine, nine of 13, and five of nine were diagnosed on September 24, 25, 26, 27, and 28, respectively. The first case of influenza A in the village was detected on October 2, 2021 in a pilgrim returning from the GMT (Supplementary Figure 1). Further cases were observed in other returned pilgrims on October 2, 3, and 4 and in villagers who did not travel to Touba from October 3-29. Of note, one case was also observed in a returned pilgrim on October 28.

Analysis of influenza A virus genomes

A total of 27 full-length or near full-length genomes with the eight segments were obtained. They were generated from 20 patients who participated in the GMT and were diagnosed with influenza A virus at the Mbacké health care center, from three patients who participated in the GMT and were pilgrims diagnosed

with influenza A virus when returning in Ndiop and Dielmo villages, and from four patients from the villages who did not participate in the GMT. All genomes were classified in the same influenza A/H1N1 virus lineage 6B.1A.5a.1 (Supplementary results). Four clusters (A, B, C, and D) could be identified based on phylogenetic analyses (Figure 1), suggesting that patients who were infected in Touba may have shared closely related viral strains with patients inhabiting the Ndiop village who did not participate in the GMT.

Discussion

National or international spread of infectious diseases triggered by mass gathering events has been exemplified by the globalization of cholera or meningococcal meningitis after epidemics occurring at the Hajj pilgrimage in Saudi Arabia [7]. Amplification of SARS-CoV-2 circulation at mass gatherings has also been a concern during the COVID-19 pandemic [8]. Surveillance data on the incidence of influenza-related hospitalizations have suggested that the carnival may have played a role in the spread of influenza virus in the Netherlands in 2018 [9]. However, to the best of our knowledge, no study based on virus genomic analysis has been conducted to challenge the hypothesis that a mass gathering could play a role in the spread of the influenza virus.

The present study identified influenza A/H1N1 virus of clade 6B.1A.5a.1 as responsible for the influenza cases diagnosed during the GMT, as well as in two Senegalese villages over a 1-month period. This lineage was present since at least August 2020 in various countries, then considerably decreased (or almost vanished) at the end of 2022. Interestingly, the analyses of viral genomes allowed the identification of four clusters of the influenza A virus, indicating that people who participated in the GMT and others who did not participate in the GMT and inhabit the Ndiop and Dielmo villages may have shared closely genetically related viral strains. In addition, they could indicate that people in the villages may have been infected with viral strains originating from the GMT and possibly imported from villagers in Ndiop and Dielmo when returning from the GMT. These hypotheses are also supported by the chronology of influenza A virus cases diagnosed by PCR because the incubation for influenza is no longer than 2 days and the duration of symptoms is most often no longer than 7 days. Nevertheless, interpretation of phylogeny reconstructions should be cautious. Indeed, the genome sequences analyzed here were obtained from samples collected during the same and short period of time, in the same country, and even the same epidemiological settings for some of them. Hence, they displayed small numbers of nucleotide differences between each other, which hampered providing robustly supported genetic clusters in phylogeny reconstructions. Topologies of phylogenetic trees at least partly differed according to the segment, and sequences from the present study were also clustered with some sequences from the Global Initiative on Sharing All Influenza Data database obtained from patients sampled in Senegal in 2021, for which detailed epidemiological data were not available (Supplementary material, Results, Tables 1 and 2, and Figures 2-9). Thus, the strains detected here may only be representative of those that circulated during the study period and that could have been transmitted from one place to another, including from Touba to the Ndiop and Dielmo villages.

Before the COVID-19 pandemic, influenza epidemiology in Senegal was characterized by a continuous circulation of influenza virus throughout the year, with two distinct seasonal peaks, the first around March and the second during the rainy season, which is around August. This epidemiological pattern was also observed in 2000; however, in 2021, the March seasonal peak was completely absent, whereas the August seasonal peak started between the September 13 and 19 and lasted until end of November [10]. Interestingly, although the influenza virus was circulating in Senegal since mid-September 2021, the first cases in Ndiop and Dielmo were diagnosed in early October in pilgrims returning from the GMT and then in villagers who did not travel to Touba.

Taken together, present data suggest that mass gatherings, such as the GMT, may play a role in the spread of influenza viruses at the country level through infected participants returning to areas where the virus was not yet circulating. Genomic surveillance of communicable diseases could help identifying the geographical spread of respiratory viruses in mass gathering settings.

Author contributions

CS and PG designed the study. NG, FSD, MB, HB, MS, LT and PC analyzed the data. NG, PG, CS and PC wrote the manuscript. MD, AD and CS supervised the data collection. All authors read and approved the final manuscript.

Declarations of competing interest

The authors have no competing interests to declare.

Funding

This study was supported by the French Government under the “Investments for the Future” program managed by the National Agency for Research with reference Méditerranée-Infection 10-IAHU-03.

Ethical approval

The project was approved by the National Ethics Committee for Health Research in Senegal, authorization number N°. 00087 MSAS/ DGS/ DS/ CNERS for the surveillance among villagers and N°. SEN17/62 for the surveillance among pilgrims. It was performed in accordance with the good clinical practices recommended by the Declaration of Helsinki and its amendments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2024.01.019.

References

[1] Sokhna C, Mboup BM, Sow PG, Camara G, Dieng M, Sylla M, et al. Communicable and non-communicable disease risks at the Grand Magal of Touba: the largest mass gathering in Senegal. *Travel Med Infect Dis* 2017;**19**:56–60. doi:10.1016/j.tmaid.2017.08.005.

[2] Sokhna C, Goumballa N, Hoang VT, Mboup BM, Dieng M, Sylla AB, et al. Senegal's Grand magal of Touba: syndromic surveillance during the 2016 mass gathering. *Am J Trop Med Hyg* 2020;**102**:476–82. doi:10.4269/ajtmh.19-0240.

[3] Hoang VT, Goumballa N, Dao TL, Ly TDA, Ninove L, Ranque S, et al. Respiratory and gastrointestinal infections at the 2017 Grand Magal de Touba, Senegal: a prospective cohort survey. *Travel Med Infect Dis* 2019;**32**:101410. doi:10.1016/j.tmaid.2019.04.010.

[4] Goumballa N, Hoang VT, Perières L, Anh Ly TD, Gaye PM, Diouf FS, et al. Respiratory infections among pilgrims at the Grand Magal of Touba: a comparative cohort controlled survey. *Travel Med Infect Dis* 2021;**43**:102104. doi:10.1016/j.tmaid.2021.102104.

[5] Goumballa N, Hoang VT, Diouf FS, Mbaye B, Parola P, Sokhna C, et al. Risk factors for symptoms of infection and the acquisition of pathogens among pilgrims at the Grand Magal of Touba, 2017–2021. *Travel Med Infect Dis* 2022;**49**:102418. doi:10.1016/j.tmaid.2022.102418.

[6] Goumballa N, Sambou M, Bassene H, Dieng M, Aidara A, Fenollar F, et al. High influenza A prevalence but no SARS-CoV-2 among 2021 Grand Magal pilgrims in Touba, Senegal. *Travel Med Infect Dis* 2021;**44**:102189. doi:10.1016/j.tmaid.2021.102189.

[7] Ahmed QA, Arabi YM, Memish ZA. Health risks at the Hajj. *Lancet* 2006;**367**:1008–15. doi:10.1016/S0140-6736(06)68429-8.

[8] Hui DSC, Zumla A. Mass gathering events and transmission of respiratory tract infections: updates pre and post COVID-19 lockdown. *Curr Opin Pulm Med* 2023;**29**:133–7. doi:10.1097/MCP.0000000000000956.

[9] Chen Q, Toorop MMA, de Boer MGJ, Rosendaal FR, Lijfering WMLUMC-COVID-19 Research Group. Why crowding matters in the time of COVID-19 pandemic? – a lesson from the carnival effect on the 2017/2018 influenza epidemic in the Netherlands. *BMC Public Health* 2020;**20**:1516. doi:10.1186/s12889-020-09612-6.

[10] Lampros A, Talla C, Diarra M, Tall B, Sagne S, Diallo MK, et al. Shifting patterns of influenza circulation during the COVID-19 pandemic. *Senegal. Emerg Infect Dis* 2023;**29**:1808–17. doi:10.3201/eid2909.230307.

Article n°7 :
Sequencing and characterization of human bocavirus genomes from patients diagnosed in Southern France between 2017 and 2022

Houmadi Hikmat, **Lorlane Le Targa**, Celine Boschi, Justine Py, Marielle Bedotto, Aurélie Morand, Nadim Cassir, Sarah Aherfi, Bernard La Scola, Philippe Colson (Publier dans Journal of Medical Virology)

Received: 7 February 2024 | Accepted: 13 May 2024

DOI: 10.1002/jmv.29706

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Sequencing and characterization of human bocavirus genomes from patients diagnosed in Southern France between 2017 and 2022

Houmadi Hikmat[1,2] | Lorlane Le Targa[1,2,3] | Celine Boschi[1,2,4] | Justine Py[1,2] | Marielle Bedotto[4] | Aurélie Morand[5,6] | Nadim Cassir[1,2,4] | Sarah Aherfi[1,2,4] | Bernard La Scola[1,2,4] | Philippe Colson[1,2,4]

[1]Microbes Evolution Phylogeny and Infection (MEPHI), Aix-Marseille Universite, Marseille, France

[2]IHU Méditerranée Infection, Marseille, France

[3]Biosellal, Lyon, France

[4]Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France

[5]Service d'Accueil des Urgences Pédiatriques, Hôpital Nord, AP-HM, Marseille, France

[6]Service de Pédiatrie Générale, Hôpital Timone, AP-HM, Marseille, France

Correspondence

Bernard La Scola and Philippe Colson, IHU Méditerranée Infection, 19-21 Blvd Jean Moulin, 13005 Marseille, France.
Email: bernard.la-scola@univ-amu.fr and philippe.colson@univ-amu.fr

Funding information

Agence Nationale de la Recherche

Abstract

The diversity and evolution of the genomes of human bocavirus (HBoV), which causes respiratory diseases, have been scarcely studied. Here, we aimed to obtain and characterize HBoV genomes from patients's nasopharyngeal samples collected between 2017 and 2022 period (5 years and 7 months). Next-generation sequencing (NGS) used Illumina technology after having implemented using GEMI an in-house multiplex PCR amplification strategy. Genomes were assembled and analyzed with CLC Genomics, Mafft, BioEdit, MeV, Nextclade, MEGA, and iTol. A total of 213 genomes were obtained. Phylogeny classified them all as of Bocavirus 1 (HBoV1) species. Five HBoV1 genotypic clusters determined by hierarchical clustering analysis of 27 variable genome positions were scattered over the study period although with differences in yearly prevalence. A total of 167 amino acid substitutions were detected. Besides, coinfection was observed for 52% of the samples, rhinoviruses then adenoviruses (HAdVs) being the most common viruses. Principal component analysis showed that HBoV1 genotypic cluster α tended to be correlated with HAdV co-infection. Subsequent HAdV typing for HBoV1-positive samples and negative controls demonstrated that HAdVC species predominated but HAdVB was that significantly HBoV1-associated. Overall, we described here the first HBoV1 genomes sequenced for France. HBoV1 and HAdVB association deserves further investigation.

KEYWORDS

adenovirus, coinfections, evolution, genome, human bocavirus

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Journal of Medical Virology* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Human bocaviruses (HBoV) belong to family *Parvoviridae*[1,2] and have been classfied in genus *Bocaparvovirus* and in species *Bocaparvovirus primate 1* according to the current (2022) taxonomy release of the International Committee on Taxonomy of Viruses[3] (https://ictv.global/taxonomy). They are nonenveloped single-stranded DNA viruses. Their genome is approximately 5.3 kb in length and comprises three open reading frames including two (the most conserved) that encode the nonstructural proteins, and one (the most variable) that encodes the capsid protein.[4–6] There are four genotypes, namely, HBoV1, HBoV2, HBoV3 and HBoV4.

HBoV are responsible for respiratory tract infections, particularly HBoV1 in children under 5 years of age.[7–9] HBoV2, 3 and 4 are less frequently detected in respiratory infections. HBoV1-4 are also detected in stool samples of children, but the association between gastroenteritis and these viruses has not been clearly established yet.[10–14] Respiratory infections caused by HBoV1 are mostly acute, but chronic infection has also been reported, as for the case of a 29-year-old immunocompetent patient.[15] HBoV was also detected in some children with Kawasaki disease.[16,17] HBoV was commonly found in co-infection with other respiratory viruses that co-circulate with them, such as respiratory syncytial virus, rhinoviruses, adenoviruses (HAdVs), human parainfluenza viruses, influenza A virus, human coronavirus OC43, or SARS-CoV-2.[18,19] As a matter of fact, the pathogenicity of HBoV has long been debated due to their high rate of co-infection with other respiratory viruses. Nonetheless, some studies showed that they can cause respiratory infection in absence of an association with any other known respiratory virus or bacteria.[20,21]

HBoV infections are distributed worldwide, with a seasonal distribution that changes according to the country. In temperate countries, they are mainly detected during winter and spring. Their prevalence was estimated to be 6.3% globally, and was reported to range between 6.1% and 8.5% in France.[1,21,22] As observed with the SARS-CoV-2 epidemics, understanding viral epidemics linked to different genotypes requires the availability of thousands of genome sequences to characterize the most comprehensively mutations and genetic evolution. However, there were only 309 complete HBoV genomes available in the NCBI Genbank nucleotide sequence database (https://www.ncbi.nlm.nih.gov/genbank/)[23] as of 05/06/2023; this is mighty limited for example in comparison with over 8 million SARS-CoV-2 genomes available in GenBank (and over 15 million available in the GISAID database (https://gisaid.org/).

Here we aimed to perform retrospectively the next-generation sequencing (NGS) and characterization of HBoV1 genomes from respiratory samples of patients diagnosed between 2017 and 2022 at university and public hospitals of Marseille, Southern France.

2 | MATERIALS AND METHODS

2.1 | Clinical samples

Clinical samples analyzed were nasopharyngeal samples sent to the clinical microbiology and virology laboratory of university and public hospitals of Marseille, southeastern France, for the purpose of routine clinical diagnosis of respiratory viruses. These samples were collected between February 2017 and August 2022 (67 months). DNA was extracted from nasopharyngeal samples using the KingFisher Flex system (Thermo Fisher Scientific), following the manufacturer's instructions. HBoV diagnosis was performed by qPCR with the FTD Respiratory pathogens 21 (FTD) assay (Fast Track Diagnosis) that targets HBoV1 DNA.

2.2 | NGS

2.2.1 | Primer design for PCR amplification of overlapping regions covering the whole viral genome

As viral genome NGS from clinical samples most often requires enrichment of viral nucleic acid, we implemented an "Artic-like" strategy, first used for Zika virus[24] then used at a large scale for SARS-CoV-2 genome NGS (https://artic.network/). All HBoV1 genomes available in GenBank as of 05/06/2023 were retrieved. Then, a sequence set clean-up was carried out by removing very short genomes and those with gaps, while keeping those with a minimal length of 5200 nucleotides, corresponding to ≥94% of the length of reference genome GenBank Accession no. NC_075120.1. Genome alignment was thereafter performed using the Mafft software.[25] PCR primers were designed on the basis of the alignment using the GEMI software[26] that detects the most conserved genome regions and propose primer sequences. In addition, this tool allows selecting parameters such as amplicon sizes (here, the 300–1700 nucleotide range was selected), primer sizes (here, the 19–22 nucleotide range was selected), primer hybridization temperatures (here, 58°C was selected), or numbers of degenerated nucleotides in the primers to handle intergenome diversity (here, up to 2 degenerated primers were allowed). Primers' sequences are available in Supplementary Data.

2.2.2 | PCR amplification of overlapping regions covering the whole genomes and NGS

The previously designed PCR primers were mixed into two pools using the mutliplex PCR "Artic-like" principle. PCR amplification using the two previously pools of designed PCR primers was performed separately for each pool, using the SuperScript III One-Step RT-PCR System with Platinum Taq High Fidelity (SS-III-HF kit; Invitrogen, Life

19969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Technologies). Primer concentrations in the pools were between 10 and 12.5 pmol/μL (Supplementary Table 1). Per PCR reaction, a total of 12.5 μL of 2X mix was added to 8.05 μL of water, 0.7 μL of enzyme, 0.75 μL of each PCR primer pool, and 3 μL of nucleic acid extract. The PCR program included an initial denaturation at 95°C for 5 min, then 39 cycles that included a denaturation step at 95°C for 15 s, an hybridization step at 58°C for 45 s, and an elongation step at 70°C for 90 s. These 39 cycles were followed by a final elongation step at 70°C for 5 min. Purification of the PCR products was carried out on a NucleoFast 96-well plate (Macherey Nagel ref 743100.50) followed by an elution of the PCR amplicons with 40 μL of pure water. Subsequently, the two pools of PCR products were mixed at equal volume.

To set up our system, we used the Oxford Nanopore Technology with the Ligation Sequencing Kit (SQK-LSK109) library that was sequenced on a SpotON flow cell Mk I, R9.4.1 on a GridION instrument, according to the manufacturer's instructions (Oxford Nanopore Technologies). After our "Artic-like" system was developed to be used on the studied nasopharyngeal samples, HBoV genomes were sequenced using the Illumina technology on a NovaSeq 6000 instrument or on a MiSeq instrument using the COVIDSeq protocol (Illumina Inc.) and replacing the COVID-19 ARTIC PCR primers with the PCR primers designed in the present work. The NovaSeq. 6000 was loaded using the classic Illumina loading procedure in a SP flow cell according to the NovaSeq-XP workflow with a reading of 2 × 50, according to a previously described procedure.[27] For the Miseq instrument, procedures of loading and sequencing with a reading of 2 × 250 were performed as recommended in the MiSeq sequencing guide (Illumina Inc.) and following a previously described procedure.[27] Genomes obtained were submitted to the GenBank sequence database (https://www.ncbi.nlm.nih.gov/genbank/; submission no. 2793582; GenBank Accession no PP296703 - PP296915).

2.3 | Bioinformatic analyses of viral genomes

A pipeline was set up using the CLC Genomics Workbench v.7 software, consisting of NGS reads trimming and subsequent mapping using as reference the genome Accession no. LC651179.1 available in GenBank. Trimming was carried out with a minimum number of 15 and a maximum of 1000 nucleotides in the reads. For mapping, the parameters used were a minimum coverage of sequencing reads on the reference genome of 90% with a minimum identity of 90%. The number of mismatches, deletions, and insertions allowed was 2, 3, and 3, respectively. Ambiguity symbols "N" were in the parts where there was ambiguity. After read mapping, the consensus genome sequences with a minimum of 3 reads as NGS depth were recovered in fasta format. Sequence alignments were then performed using the Mafft software[25] with default parameters and attempting to align regions with gaps for further analysis. The BioEdit software (https://bioedit.software.informer.com/) was used to visualize the alignments and create the sequence identity matrices. The MEGA v.11[28] and iTOL[29] softwares were used to reconstruct and visualize phylogenetic trees. The phylogenetic tree based on HBoV genomes was constructed using the Neighbor-Joining method, the Kimura 2-parameter model with 1000 replicates and a number of 100 threads. Other parameters used were inclusion of transition and tranversion as substitutions, uniform rates among sites, pairwise deletion for gaps/missing date, and selection of all codon positions and of noncoding sites.

The Nextclade tool (https://clades.nextstrain.org/)[30] was adapted to HBoV1 and allowed the report of nucleotide and amino acid mutations harbored by HBoV genomes. We chose as reference the genome GenBank Accession no. NC_075120.1, which was obtained from a sample collected in 2005 and which was updated in March 2023. Finally, to delineate HBoV1 genotypic clusters, a hierarchical clustering was performed based on presence/absence patterns identified using the BioEdit software within a set of 27 mutations each present in at least 10% of the whole genome set, using the Multiple Experiment Viewer (MeV) software,[31,32] with default parameters. A principal component analysis (PCA) was then carried out using the Xlstat Microsoft software to see if there was any correlation between a HBoV1 genotypic cluster and a particular co-infection.

2.4 | Viral coinfection analysis and HAdV genotyping

By performing a hierarchical clustering analysis with Mev, we found a correlation between one of the genotypic clusters and the HAdV. We hence investigated if this correlation was linked to a particular HAdV species or not. For nasopharyngeal samples for which a coinfection with HAdV was diagnosed, HAdV typing was performed using a previously described in-house qPCR system specific to each type of HAdV[33] with the Light Cycler 480 Probes Master reagent on a Light Cycler 480 thermal cycler (Roche), according to the manufacturer's protocol. Specific qPCR for the different HAdV species C serotypes (1, 2, 5 and 6) were performed using previously described primers and probe systems.[34]

2.5 | Statistical analyses

Statistical analyses were performed using the Social Science Statistics web application (https://www.socscistatistics.com/tests/fisher/default2.aspx). $p < 0.10$ was considered statistically significant.

3 | RESULTS

3.1 | Primer design for PCR amplification of overlapping regions covering the whole genomes

We retrieved 100 HBoV1 genomes from GenBank. Their alignment with the GEMI software allowed obtaining 80 proposed primer pairs as output, of which 16 were selected as their PCR products allowed covering the whole HBoV1 genomes with overlaps between

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

contiguous amplicons. These 16 primer pairs were tested, and 8 were selected for pooling as they provided spots on agarose gels with sufficient intensity and at the expected nucleotide size. These eight primers were optimized by increasing their concentration when the bands on the agarose gel were of low intensity. Selected concentrations varied between 10 and 12.5 pmol/µL. Nanopore technology on a Gridion instrument was used to carry out the NGS tests. All samples were then sequenced using Illumina technology.

3.2 | NGS

For the NGS step, 9% of our samples were sequenced on a Miseq instrument and 91% on a Novaseq. 6000 instrument. A total of 601 nucleic acid extracts obtained from HBoV1 DNA-positive nasopharyngeal samples collected between February 2017 and August 2022 were tested. Sequencing of these samples was attempted using our "artic-like" strategy, and 213 genome sequences were obtained with at least 90% coverage of the reference genome GenBank Accession no. LC651179.1 and a depth of NGS reads of at least 3X. These genomes were obtained from nasopharyngeal samples collected between February 06, 2017 and July 07, 2022 (67 months). The total number of reads per sample mapped to the reference genome ranged from 5675 to 9 917 033, with an average of 3 569 789 ± 2 943 862 reads. These 213 selected genomes exhibited a coverage ranging from 91% to 99% of the reference genome, with an average of 98% ± 1%. The cycle threshold values (Ct) of the qPCR used for diagnostic of HBoV1 infection for the 213 samples analyzed ranged from 8 to 34 with an average of 22 ± 5, while that for the samples from which the obtained genomes displayed a coverage <90%, and which were not retained, ranged from 17 to 38 with an average of 33.1 ± 2.8.

The HBoV1 sequences obtained in this study have been deposited on GenBank (GenBank accession no PP296703 - PP296915).

3.3 | Bocavirus mutational diversity

Within the 213 genomes, the total number of nucleotide substitutions compared with reference GenBank Accession no. NC_075120.1 ranged from 7 to 44, with an average of 23 ± 5. These mutations tended to be scattered over the whole genome length. Only 7 of the 213 genomes exhibited nucleotide deletions; 4 exhibited one deletion and 3 exhibited 2. These nucleotide mutations were mostly synonymous mutations. Regarding amino acid mutations, the total number of substitutions ranged from 5 to 18, with an average of 9 ± 3. One genome harbored an amino acid deletion while 2 harbored two or one amino acid insertions. A total of 167 different amino acid substitutions were recorded in the set of 213 genomes, and only 33 of these were present in at least three genomes. The most predominant amino acid mutations were NS1:K649E, which also corresponds to UP1:K28E and was present in all of our genomes, followed by VP1:K613E corresponding to VP2:K524E and VP3:K484E, both present in 211 of the 213 genomes (Figure 1).

Stop codons were present in four genomes. Two of these genomes harbored the same stop codons, namely VP1:Q418*, corresponding to VP2:Q329*, and VP3:Q289*, therefore involved the VP1, VP2 and VP3 overlapping genes that encode the minor capsid proteins. These stop codons are located approximately in the middle of these genes. Another genome harbored ORFx:L93* and another VP2:G81* corresponding to VP3:G41*. ORFx:L93* is located near the end of the protein, while VP2:G81* corresponding to VP3:G41* is found at the beginning of the gene.

As a comparison, we were also interested in recording mutations present in the HBoV1 complete genomes available in GenBank, which were between 4800 and 5600 nucleotide in size. A total of 308 genomes were included in this analysis. Four exhibited no mutation compared with reference NC_075120.1. Regarding the 304 remaining genomes, the total number of nucleotide substitutions ranged from 10 to 68, with an average of 23 ± 8 substitutions, and the total number of amino acid substitutions ranged from 5 to 58 with an average of 10 ± 6 substitutions. Seven genomes exhibited between 1 and 6 amino acid deletions. No amino acid insertions were found in these samples. Overall, 596 different amino acid mutations were harbored by any of these 308 genomes, 80 being present in at least 3 genomes. Of these 596 amino acid substitutions, 69 were in common with those found in the genomes obtained in the present study. Amino acid substitutions, either in common between our data and those of GenBank or not in common between these two sets of genomes, affected all genes. We plotted in Figure 2 all 165 amino acid mutations present in at least two genomes (as mutations harbored by only one genome was too large for a legible representation) (Figure 2).

3.4 | HBoV1 classification into genotypic clusters

The 213 HBoV genomes obtained in the present study were from samples whose collection dates were scattered over the whole 2017–2022 period (Figure 3). All obtained genomes belonged to the HBoV1 species as determined by phylogeny. The mean nucleotide diversity between these genomes was 99.61 ± 0.18% and ranged between 98.70% and 100.00%, corresponding to a mean number of 13 ± 7 nucleotides (range, 0–49). Congruently with this overall low intergenome genetic diversity, bootstrap values in the phylogenetic tree were very low, most often <60%, which prevented a classification into lineages within the HBoV1 genotypic cluster. Besides, no genome clustering was observed according to the date of collection of the samples.

To get a better insight in the genetic evolution of these HBoV1 genomes over time, we performed a hierarchical clustering analysis based on a list of 27 variable positions exhibiting mutations present in at least 10% of the 213 genomes. This allowed assigning genomes in five genotypic clusters we named cluster α, β, γ, δ, and ε. Some genomes could not be unambiguously assigned and were assigned to

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

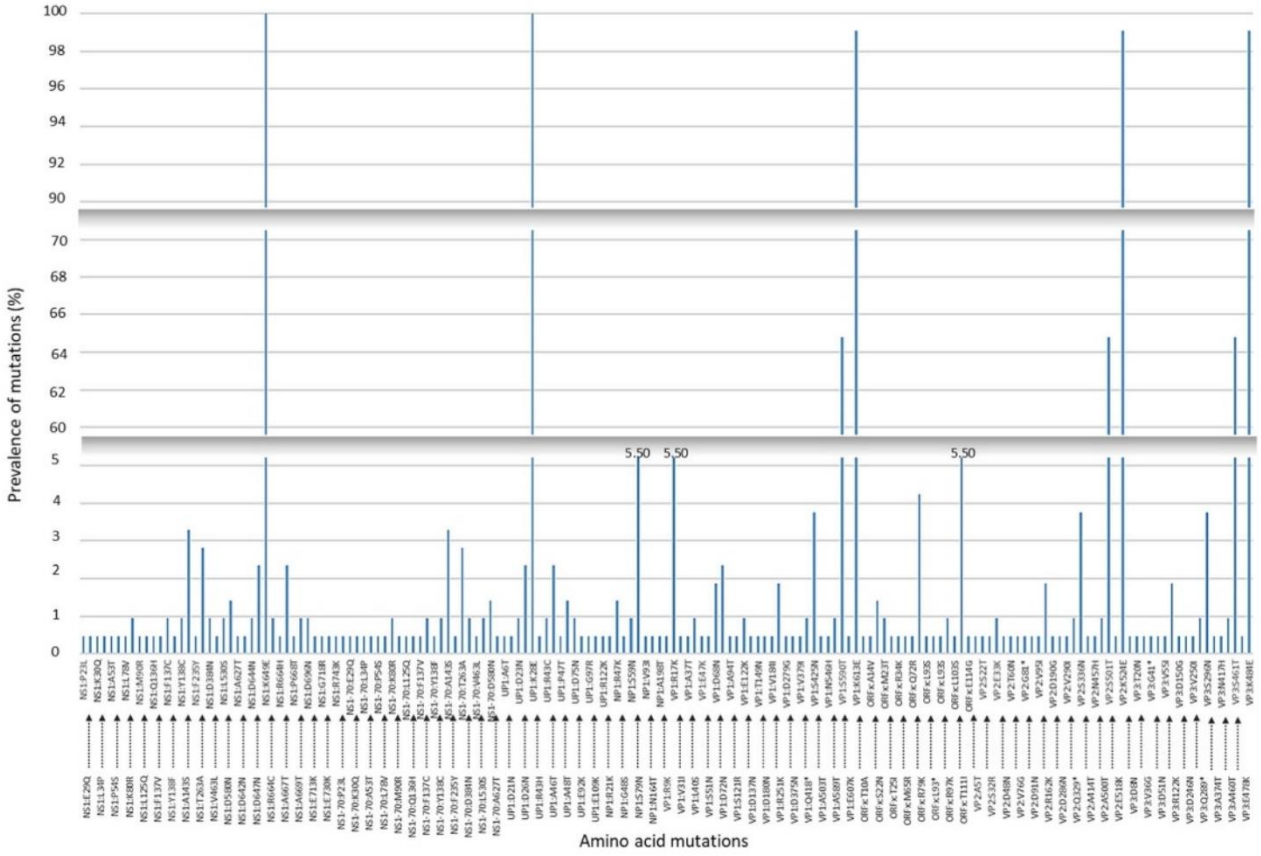


FIGURE 1 Representation of amino acid substitutions in the 213 genomes obtained in the present study.

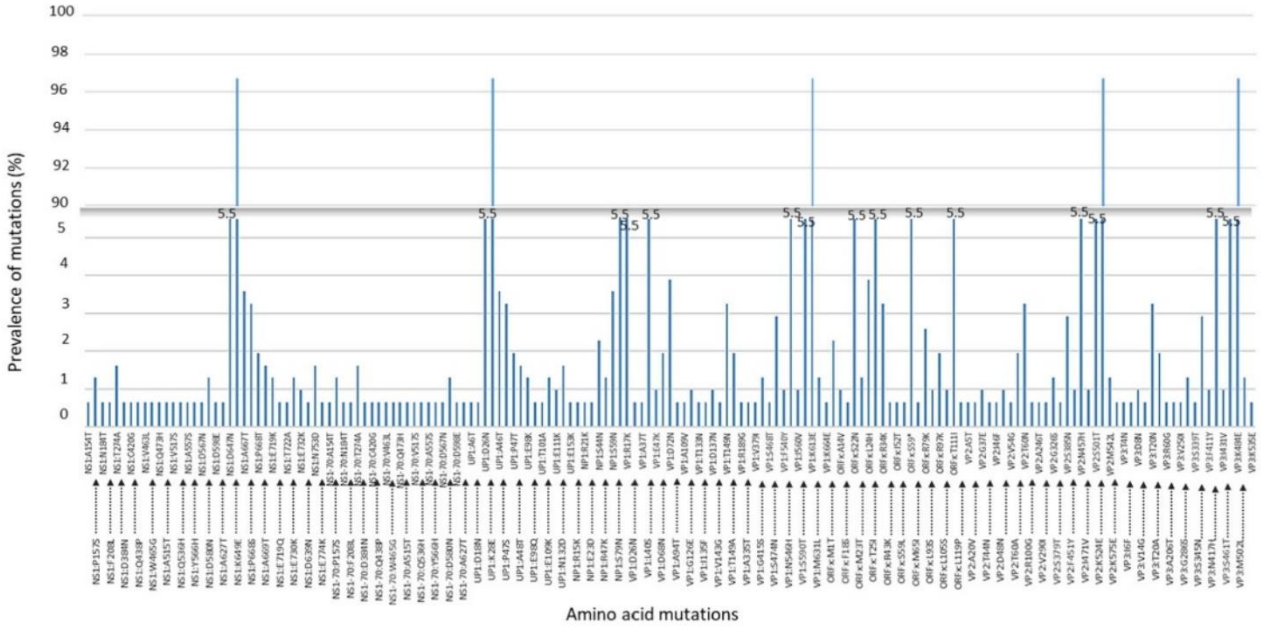


FIGURE 2 Representation of amino acid substitutions in NCBI genomes.

an "unclassified" cluster (Figure 4). We looked at how the circulation of these clusters occurred over time and observed they were scattered over the study period, being each detected in different years although with different prevalence (Figure 5). As HBoV1 was diagnosed for some samples in coinfection with other respiratory viruses, we were interested in analyzing the distribution of these viral coinfections according to the HBoV1 cluster. For 52% of the respiratory samples from which the 213 genomes were obtained,

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

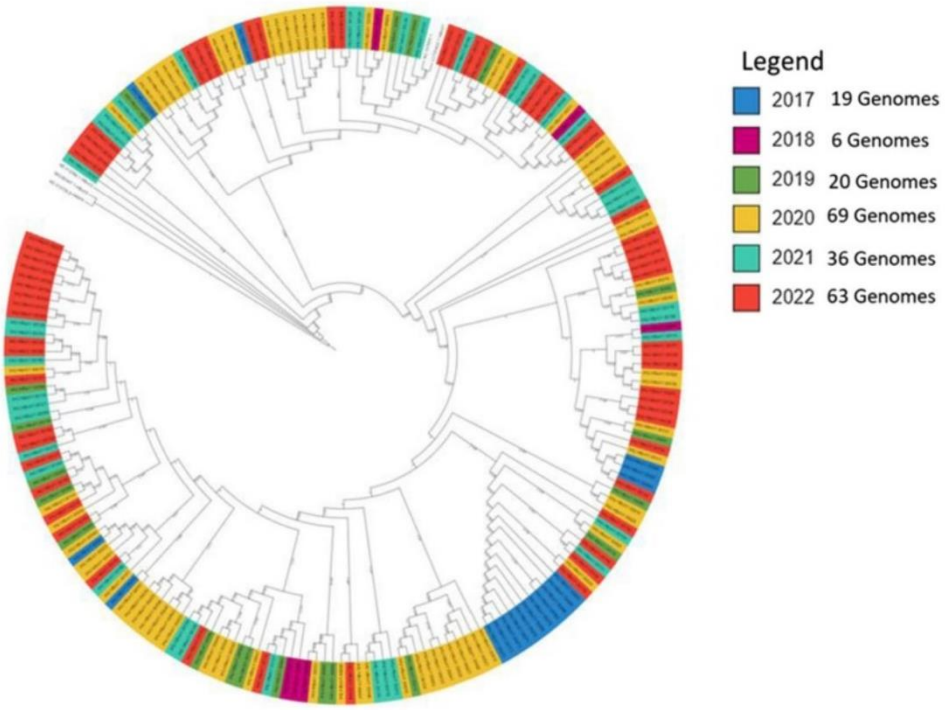


FIGURE 3 HBoV1 distribution by year.

between one or up to four additional respiratory viruses were simultaneously diagnosed, rhinoviruses and HAdVs being those the most frequent, in 32% and 46% of the cases, respectively. We observed that before Year 2020 (before the SARS-CoV-2 pandemic), epidemic peaks of HBoV1 were observed just after those of HAdVs and rhinoviruses. Then, during and after 2020, we observed that HBoV1, HAdVs, and rhinoviruses exhibited similar seasonal prevalence curves (Figure 6).

We then tested using PCA if the prevalence of HBoV1 genotypic clusters that we previously defined (α, β, γ, δ, ε, and "unclassified") was correlated with those of the co-diagnosed respiratory viruses. This allowed revealing a tendency toward a correlation between the prevalence of HAdVs and that of the HBoV1 cluster α (Figure 7). Of the 18 nasopharyngeal samples for which this association between HBoV1 cluster α and HAdV was observed, a sufficient volume of sample was available for 15 to perform the classification of HAdVs into genotypes A-G. These 15 samples were tested concurrently with 15 other nasopharyngeal samples collected during the same year and week and positive by PCR for HAdV but negative for HBoV1. Four HBoV1-positive samples and one HBoV1-negative sample were co-infected with HAdVs. Three HAdV genotypes including B, C, and F were detected. There was no significant association between the prevalence of HBoV1 and that of HAdV C and F, but we found a significant association with the prevalence of HAdV B (Table 1). Indeed, 6/15 HBoV1 genotypic cluster α-positive samples showed co-diagnosis of HAdV B infection, compared with only 1/15 HBoV1-negative samples ($p = 0.08$; Fisher exact test).

Finally, to investigate further the association HBoV1 cluster α with HAdV C, we determined the HAdV C cluster and observed that HAdV C2 predominated in both HBoV1 cluster α-positive and HBoV1-negative samples (Table 2), and that co-infection with both HAdV C1 and C2 serotypes was observed in both groups without significant differences in prevalence.

4 | DISCUSSION

Here, we were able to obtain and characterize 213 HBoV1 genomes, which span the most recent 6 years period from 2017 to 2022. This represents approximately two-thirds of the 309 HBoV1 genomes available worldwide in GenBank at the beginning of our study, and therefore expands considerably the global set of genomes available. For these 309 genomes deposited in GenBank, no information regarding the date of sample collection was available in 32 cases. For the remaining 277 genomes, only 55 sequences were obtained from samples collected between 2017 and 2021. In addition, no HBoV1 genome from samples collected in 2022 was deposited in GenBank. The 213 genomes obtained in the present study were obtained from nasopharyngeal samples collected between 2017 and 2022, 64 being from samples collected in 2022. Moreover, none of the HBoV1 genomes available in GenBank as of June 05, 2023 originated from France. The present work therefore provides and reports on the first genomes from this country.

All genomes obtained in the present study belong to the *Bocaparvovirus primate 1* species, which is in line with previously

09690071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

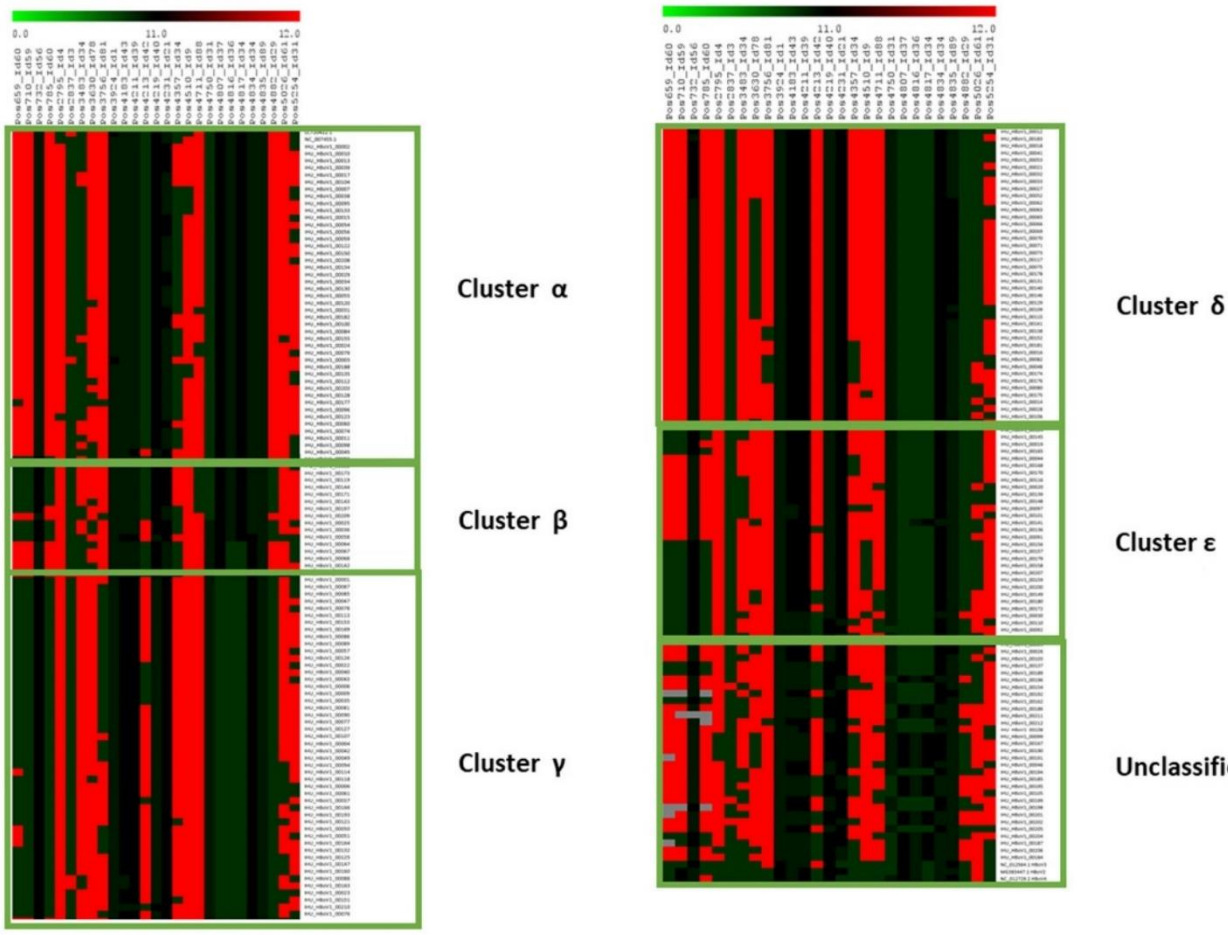


FIGURE 4 Hierarchical clustering of HBoV1.

published studies showing that HBoV1 predominates in human respiratory infections, while HBoV2, HBoV3, and HBoV4 predominate in gastroenteritis.[35,36] The evolution of HBoV, which have relatively stable genomes, differs from that of other epidemic viruses such as SARS-CoV-2, whose temporal evolution is associated with the successive circulation of a few viral variants, which replace those that circulated during the previous months[37] (https://covariants.org/). This low level of genetic variability for HBoV can be at least partly explained by the fact that they are DNA viruses, which are known to be more genetically stable than RNA viruses.[38] Compared with a reference HBoV genome obtained from a patient sampled between December 2003 and March 2004, some of the amino acid substitutions identified in all the genomes in our study were also hosted by viral genomes available in GenBank. Four of these mutations were present in all genomes deposited in GenBank since 2005. The diversity observed in this study may have an impact on the contagiousness and severity of HBoV1 infections. This could be linked to mutations in the capsid proteins that are involved in attachment of the virus to its host cell, as has been observed for other viruses such as SARS-CoV-2 for which mutation in the spike glycoprotein was associated with a high contagiousness[39] or the respiratory syncytial virus for which capacities of immune response were reported to vary according to the viral genotype.[40] Further studies are needed to gain a better knowledge about these points and the impact of given mutations. Intriguingly, stop codons were observed in three genomes from the present study and also in genomes from GenBank at different genome positions. The significance of these stop codons is unclear. It was for instance observed in SARS-CoV-2 variants of concern, indicating that some accessory viral genes are dispensable in some settings.[41] Nonetheless, here stop codons are in overlapping genes that encode the minor capsid proteins which are essential for viral replication. Therefore, there could be mechanisms such as translational readthrough reported in some viruses.[42]

Other respiratory viruses were concurrently detected with HBoV1 in more than half of the nasopharyngeal samples studied here, which is congruent with other studies conducted on HBoV that showed a relatively high proportion of coinfections ranging from 45% to 72% of cases.[43,44] Here, the most predominant coinfecting viruses were rhinoviruses followed by HAdVs, which is also congruent with previous studies.[21] We observed that prevalence rates of these different viruses varied between before, during, and after Year 2020. As a matter of fact, the epidemiological features of co-circulation of respiratory viruses

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

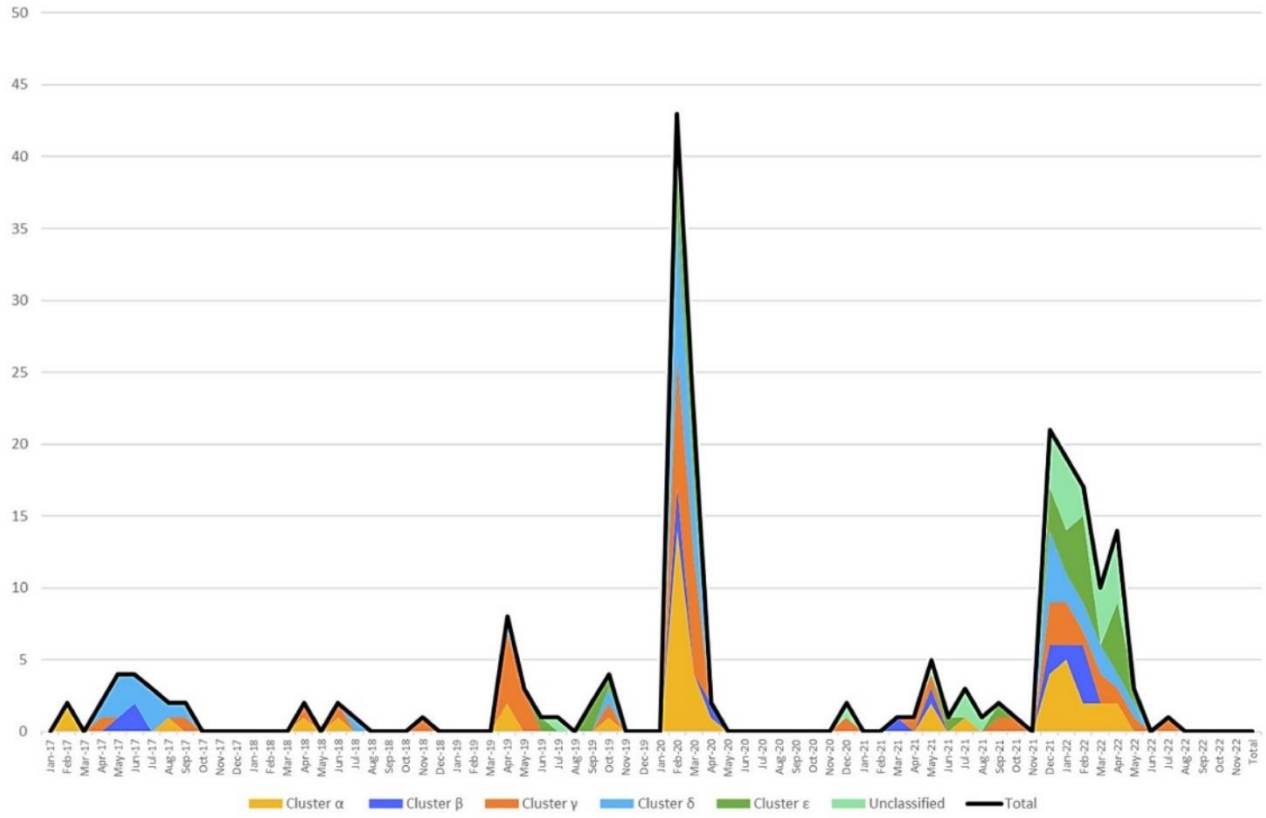


FIGURE 5 Distribution of HBoV1 genotypes over time.

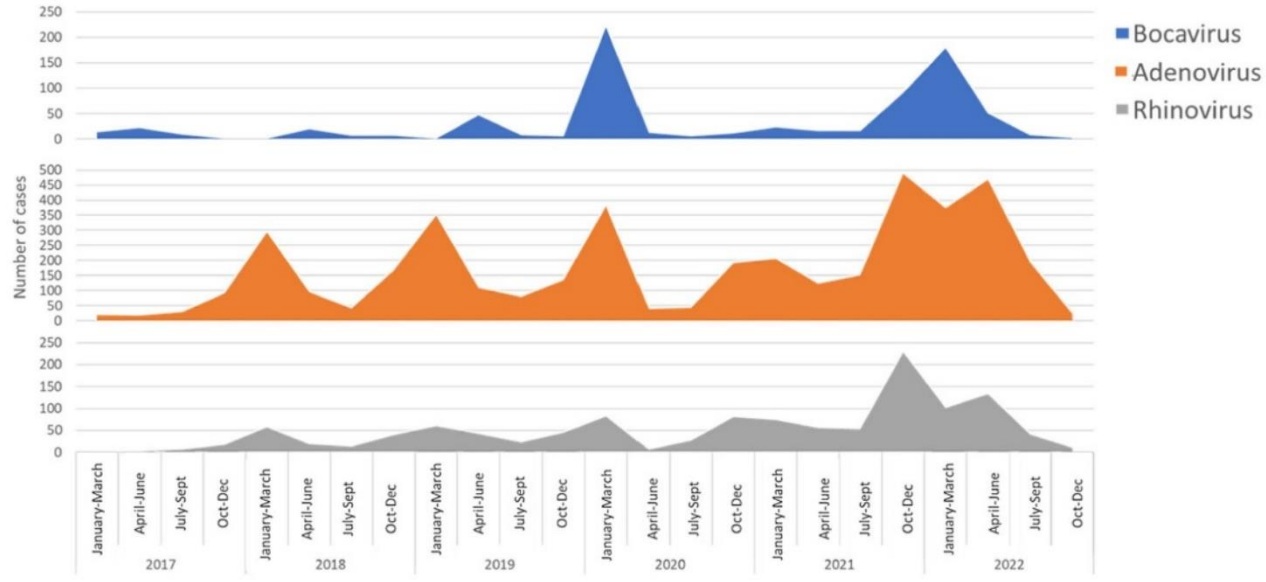


FIGURE 6 Distribution per year and season of the numbers of diagnoses of HBoV1, HAdV and rhinovirus infections.

and the processes of possible interference between these viruses appear to be complex issues. These have been for instance reported and investigated for respiratory syncytial virus, influenza virus, and rhinovirus during and after the SARS-CoV-2 pandemic.[45–48]

Hierarchical cluster analysis carried out on 27 variable positions present in at least 10% of the HBoV1 genomes delineated five genotypic clusters, and this enabled us identifying an association between HAdV and HBoV1 genotypic cluster α in coinfection. HAdV genotyping showed that HAdVC was the predominant HAdV during

969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the study period, which is in agreement with previous studies.[49,50] We further found that among HAdV-positive and HBoV1-negative samples, HAdVC and HAdVB genotypes were those the most frequently detected, which has been also reported previously in respiratory infections.[51] Furthermore, we found that HAdVC2 predominated in both HBoV1 cluster α-positive and HBoV1-negative control samples. We also found that five nasopharyngeal samples were coinfected with two different HAdV genotypes. Such coinfections between HAdVs were already reported.[52,53] Most interestingly, we observed a significant association of HBoV1 cluster α with HAdVB. This questions if some HBoV1 might need an auxiliary virus to replicate, as for the case of Adeno-Associated Virus[54] that belongs to the same viral family. This observation requires further study. It should be considered that the different HBoV1 genotypic clusters identified here were detected during different years and at different prevalence. In addition, it was reported that monoinfection with HBoV1 could be associated with severe acute lower respiratory infections in some cases, that viral load could be a factor associated to clinical severity, and that other respiratory viruses were particularly found in codetection with HBoV1, including respiratory syncytial virus in children.[55-58] Molecular monitoring of these viruses in mono- and co-detection by complete sequencing would help to clarify this question in future work. Taken together, previous findings suggest that in respiratory infections HAdVB may be linked specifically to HBoV1 genotypic cluster α, but further analyses with larger cohorts are required to investigate this point.

The present study has some limitations. Thus, there is a bias linked to the sampling as it involved samples collected in the setting of university public hospitals, and from a single geographical area, South of France; in addition, our sampling is limited in time (2017–2022). Therefore, we acknowledge that results could have been different if the study would have involved respiratory samples from patients who received cares outside hospitals or in other European or world areas, notably due to different indications or policies for the collection of respiratory samples, or due to differences in the panel of respiratory viruses tested. Nevertheless, the results found in this study can be expected to be similar to those from other geographical areas since we analyzed the diversity of all HBoV1 genomes available in GenBank regardless of the countries they originated and the time periods during which the samples were collected, and we found genotypic patterns comparable to those based on the sequences we obtained in the present study (Figure 2). However, more in-depth analyses of genetic changes over time and in different countries would be necessary to verify these results, and our findings would hopefully contribute generating interest from other laboratories in several countries to start implementing global

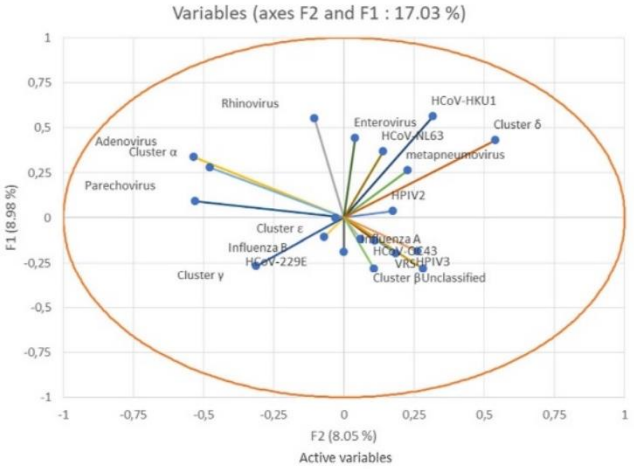


FIGURE 7 Principal component analysis of respiratory virus coinfections according to the HBoV1 genotypes.

TABLE 1 Prevalence of HAdV genotypes according to HBoV1 genotypic cluster α positivity or HBoV1 negativity.

	HAdVB*[c]	HAdVC[d]	HAdVF	Not typed HAdV
HBoV1 cluster α-positive ($n = 15$)[a]	6	12	1	1
HBoV1-negative ($n = 15$)[b]	1	11	0	4

[a]Five samples were coinfected by two HAdVs.
[b]One sample was coinfected with two HAdVs.
[c]p value = 0.08 (Fisher exact test).
[d]p Value = 1 (Fisher exact test).

TABLE 2 Genotyping of HAdV C for HBoV1 genotypic cluster α-positive and HBoV1-negative samples.

	HAdVC1[a]	HAdVC2[b]	HAdVC5	HAdVC6	Not typed HAdVC
HBoV1 cluster α-positive ($n = 12$)	2	10	0	0	1
HBoV1-negative ($n = 11$)	2	7	2	1	1

[a]p value = 1 (Fisher exact test).
[b]p value = 0.37 (Fisher exact test).

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

HBoV genomic surveillance. Also, we acknowledge that data availability for a longer study period could have provided here a more comprehensive picture of the genetic evolution and dynamics of HBoVs.

Overall, the present study expands considerably the set of HBoV1 genomes available worldwide and provides the first genomes for France. It highlights limited genetic variability at the genomic level over a 5-year recent period of time, and unexpectedly, reveals an association between an HBoV1 genotypic cluster and HAdV in coinfection, which prompts conducting further studies, particularly considering the knowledge of auxiliary viruses in the same viral family. The clinical significance of these coinfections were not investigated in the present work, which only consisted in genome sequencing and analyses. However, the greater severity is a much-discussed factor which has been the subject of contradictory results, which may suggest a lack of severity[56] but in fact very probably depends on the viruses considered and on the age and clinical background of the patients.[57,59] Another issue is whether these coinfections are concurrent acute infections with the two viruses or alternatively an acute, superinfection with one virus in a patient protracted carrier of a second virus. This latter case is suspected when one of the two viruses is detected at a very low viral load, which is particularly common in respiratory virus co-infections involving HBoV.[60] Beyond, here, the association of Adenovirus C and HBoV1 caught our attention because HBoVs (genus *Bocaparvovirus*) are members of the subfamily *Parvovirinae*, as the Adeno-associated viruses (genus *Dependoparvovirus*) that require coinfection with a helper virus to replicate efficiently.[61] Therefore, this questions if HBoVs may have a similar role as helper. In fact, the present work opens the question of the interaction of certain HBoV genotypes with certain adenovirus genotypes, even if many other factors could account for this nonfortuitous association. Beyond, present data warrant performing genomic surveillance for HBoVs as already performed for influenza viruses, respiratory syncytial virus, and SARS-CoV-2.[62]

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Bernard La Scola, Philippe Colson. *Contributed materials/analysis tools*: Houmadi Hikmat, Lorlane Le Targa, Celine Boschi, Justine Py, Marielle Bedotto, Aurélie Morand, Nadim Cassir, Sarah Aherfi, Bernard La Scola, Philippe Colson. *Analyzed the data*: Houmadi Hikmat, Celine Boschi, Bernard La Scola, Philippe Colson. *Wrote the paper*: Houmadi Hikmat, Bernard La Scola, and Philippe Colson.

ACKNOWLEDGMENTS

We are grateful to Claudia Andrieu from the next-generation sequencing platform of IHU Méditerranée Infection, Marseille, France, for her technical help. This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03; and by the French Ministry of Higher Education, Research and Innovation (Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation) and the French Ministry of Solidarity and Health (Ministère des Solidarités et de la Santé).

CONFLICT OF INTEREST STATEMENT

Lorlane Le Targa is an employee of the Biosellal company. The other authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

HBoV genomes analyzed here have been submitted to the GenBank sequence database (https://www.ncbi.nlm.nih.gov/genbank/)[22]; submission no. 2793582 (GenBank Accession no PP296703 - PP296915).

ETHICS STATEMENT

The present study has been approved and registered on the Health Data Access Portal of Assistance Publique-Hôpitaux de Marseille - (AP-HM) under No. PADS23-265 and has been approved under No. CSE23-39 by the Ethics and Scientific Committee of AP-HM.

ORCID

Bernard La Scola http://orcid.org/0000-0001-8006-7704
Philippe Colson http://orcid.org/0000-0001-6285-0308

REFERENCES

1. Guido M, Tumolo MR, Verri T, et al. Human bocavirus: current knowledge and future challenges. *World J Gastroenterol*. 2016;22(39):8684-8697. doi:10.3748/wjg.v22.i39.8684
2. Thanh TT, Van HMT, Hong NTT, et al. The first genome sequences of human bocaviruses from Vietnam. *Wellcome Open Res*. 2016;1:16. doi:10.12688/wellcomeopenres.10042.2
3. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res*. 2018;46(D1):D708-D717. doi:10.1093/nar/gkx932
4. Zou W, Cheng F, Shen W, Engelhardt JF, Yan Z, Qiu J. Nonstructural protein NP1 of human bocavirus 1 plays a critical role in the expression of viral capsid proteins. *J Virol*. 2016;90(9):4658-4669. doi:10.1128/JVI.02964-15
5. Shen W *Human Bocavirus 1 Genome Organization and Replication Mechanism*. Dissertation. University of Kansas; 2016. Accessed October 7, 2022. https://kuscholarworks.ku.edu/handle/1808/24814
6. Pogka V, Moutousi A, Kossyvakis A, et al. Genetic variability of human metapneumo- and bocaviruses in children with respiratory tract infections. *Influenza Other Respir Viruses*. 2014;8(1):107-115. doi:10.1111/irv.12185
7. Koseki N, Teramoto S, Kaiho M, et al. Detection of human bocaviruses 1 to 4 from nasopharyngeal swab samples collected from patients with respiratory tract infections. *J Clin Microbiol*. 2012;50(6):2118-2121. doi:10.1128/JCM.00098-12
8. Han TH, Chung JY, Hwang ES. Human bocavirus 2 in children, South Korea. *Emerging Infect Dis*. 2009;15(10):1698-1700. doi:10.3201/eid1510.090337
9. Risku M, Kätkä M, Lappalainen S, Räsänen S, Vesikari T. Human bocavirus types 1, 2 and 3 in acute gastroenteritis of childhood. *Acta Paediatr*. 2012;101(9):e405. doi:10.1111/j.1651-2227.2012.02727.x
10. Silva PE, Figueiredo CA, Luchs A, et al. Human bocavirus in hospitalized children under 5 years with acute respiratory infection,

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

São Paulo, Brazil, 2010. *Arch Virol*. 2018;163(5):1325-1330. doi:10.1007/s00705-017-3694-5
11. Nantachit N, Kochjan P, Khamrin P, Kumthip K, Maneekarn N. Human bocavirus genotypes 1, 2, and 3 circulating in pediatric patients with acute gastroenteritis in Chiang Mai, Thailand, 2012-2018. *J Infect Public Heal*. 2021;14(2):179-186. doi:10.1016/j.jiph.2020.12.007
12. Rizk NM, Abd-Elmaksoud S, Farid TM, Abohashish MMA, Al-Herrawy AZ, Hamza IA. Etiology of diarrheal disease among children under 5 years in Egypt: a high incidence of human bocavirus. *J Egypt Public Health Assoc*. 2021;96(1):24. doi:10.1186/s42506-021-00084-z
13. Bergallo M, Daprà V, Rassu M, Calvi C, Montanari P, Galliano I. Human bocavirus in children with acute gastroenteritis in Piedmont, Italy. *Minerva Pediatrics*. 2023;75(4):528-535. doi:10.23736/S2724-5276.18.05365-3
14. Jin Y, Cheng W, Xu Z, et al. High prevalence of human bocavirus 2 and its role in childhood acute gastroenteritis in China. *J Clin Virol*. 2011;52(3):251-253. doi:10.1016/j.jcv.2011.07.012
15. Windisch W, Pieper M, Ziemele I, et al. Latent infection of human bocavirus accompanied by flare of chronic cough, fatigue and episodes of viral replication in an immunocompetent adult patient, Cologne, Germany. *JMM Case Rep*. 2016;3:e005052. doi:10.1099/jmmcr.0.005052
16. Bajolle F, Meritet JF, Rozenberg F, et al. Markers of a recent bocavirus infection in children with Kawasaki disease: "A year prospective study.". *Pathol Biol*. 2014;62(6):365-368. doi:10.1016/j.patbio.2014.06.002
17. Catalano-Pons C, Giraud C, Rozenberg F, Meritet JF, Lebon P, Gendrel D. Detection of human bocavirus in children with Kawasaki disease. *Clin Microbiol Infect*. 2007;13(12):1220-1222. doi:10.1111/j.1469-0691.2007.01827.x
18. Wang Y, Li Y, Liu J, et al. Genetic characterization of human bocavirus among children with severe acute respiratory infection in China. *J Infect*. 2016;73(2):155-163. doi:10.1016/j.jinf.2016.05.014
19. Boschi C, Hoang VT, Giraud-Gatineau A, et al. Coinfections with SARS-CoV-2 and other respiratory viruses in Southeastern France: a matter of sampling time. *J Med Virol*. 2021;93(4):1878-1881. doi:10.1002/jmv.26692
20. Moesker FM, van Kampen JJA, van der Eijk AA, et al. Human bocavirus infection as a cause of severe acute respiratory tract infection in children. *Clin Microbiol Infect*. 2015;21(10):964.e1-964.e8. doi:10.1016/j.cmi.2015.06.014
21. Verbeke V, Reynders M, Floré K, et al. Human bocavirus infection in Belgian children with respiratory tract disease. *Arch Virol*. 2019;164(12):2919-2930. doi:10.1007/s00705-019-04396-6
22. Polo D, Lema A, Gándara E, Romalde JL. Prevalence of human bocavirus infections in Europe. A systematic review and meta-analysis. *Transbound Emerg Dis*. 2022;69(5):2451-2461. doi:10.1111/tbed.14233
23. Sayers EW, Cavanaugh M, Clark K, et al. GenBank 2023 update. *Nucleic Acids Res*. 2022;51(D1):D141-D144. doi:10.1093/nar/gkac1012
24. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20(1):8. doi:10.1186/s13059-018-1618-7
25. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772-780. doi:10.1093/molbev/mst010
26. Sobhy H, Colson P. Gemi: PCR primers prediction from multiple alignments. *Comp Funct Genomics*. 2012;2012:783138. doi:10.1155/2012/783138
27. Papa Mze N, Kacel I, Beye M, et al. High throughput SARS-CoV-2 genome sequencing from 384 respiratory samples using the Illumina COVIDSeq protocol. *Genes*. 2023;14(3):681. doi:10.3390/genes14030681
28. Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics*. 1994;10(2):189-191. doi:10.1093/bioinformatics/10.2.189
29. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23(1):127-128. doi:10.1093/bioinformatics/btl529
30. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. *Nextclade: clade assignment, mutation calling and quality control for viral genomes*. October 28 2021. doi:10.5281/zenodo.5607694
31. Saeed AI, Sharov V, White J, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. 2003;34(2):374-378. doi:10.2144/03342mt01
32. Saeed AI, Bhagabati NK, Braisted JC, et al TM4 microarray software suite. *Methods Enzymol*. 2006;411:134-193. doi:10.1016/S0076-6879(06)11009-5
33. Watzinger F, Suda M, Preuner S, et al. Real-time quantitative PCR assays for detection and monitoring of pathogenic human viruses in immunosuppressed pediatric patients. *J Clin Microbiol*. 2004;42(11):5189-5198. doi:10.1128/JCM.42.11.5189-5198.2004
34. Lu X, Erdman DD. Quantitative real-time PCR assays for detection and type-specific identification of the endemic species C human adenoviruses. *J Virol Methods*. 2016;237:174-178. doi:10.1016/j.jviromet.2016.05.020
35. Guo L, Wang Y, Zhou H, et al. Differential seroprevalence of human bocavirus species 1-4 in Beijing, China. *PLoS One*. 2012;7(6):e39644. doi:10.1371/journal.pone.0039644
36. Mietzsch M, Kailasan S, Garrison J, et al. Structural insights into human bocaparvoviruses. *J Virol*. 2017;91(11):10-1128. doi:10.1128/jvi.00261-17
37. Hodcroft EB. *CoVariants: SARS-CoV-2 Mutations and Variants of Interest*. Accessed January 31, 2024. https://covariants.org/
38. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci*. 2016;73(23):4433-4448. doi:10.1007/s00018-016-2299-6
39. Carabelli AM, Peacock TP, Thorne LG et al. 2023 SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol*. doi:10.1038/s41579-022-00841-7
40. Zhou X, Jiang M, Wang F et al. 2023 Immune escaping of the novel genotypes of human respiratory syncytial virus based on gene sequence variation. *Front Immunol*, 13. doi:10.3389/fimmu.2022.1084139
41. Colson P, Delerce J, Fantini J, Pontarotti P, La Scola B, Raoult D. The return of the "Mistigri" (virus adaptative gain by gene loss) through the SARS-CoV-2 XBB.1.5 chimera that predominated in 2023. *J Med Virol*. 2023;95(10):e29146. doi:10.1002/jmv.29146
42. Kramarski L, Arbely E. Translational read-through promotes aggregation and shapes stop codon identity. *Nucleic Acids Res*. 2020;48(7):3747-3760. doi:10.1093/nar/gkaa136
43. Hao R, Ni K, Xia Q, et al. Correlation between nucleotide mutation and viral loads of human bocavirus 1 in hospitalized children with respiratory tract infection. *J Gen Virol*. 2013;94(Pt 5):1079-1085. doi:10.1099/vir.0.047472-0
44. Liu WK, Liu Q, Chen DH, et al. Epidemiology of HBoV1 infection and relationship with meteorological conditions in hospitalized pediatric patients with acute respiratory illness: a 7-year study in a subtropical region. *BMC Infect Dis*. 2018;18:329. doi:10.1186/s12879-018-3225-3
45. Hernández-Rivas L, Pedraz T, Calvo C, San Juan I, Mellado MJ, Robustillo A. Respiratory syncytial virus outbreak during the COVID-19 pandemic. How has it changed? *Enferm Infecc Microbiol Clin*. 2023;41(6):352-355. doi:10.1016/j.eimce.2021.12.018
46. Chow EJ, Uyeki TM, Chu HY. The effects of the COVID-19 pandemic on community respiratory virus activity. *Nat Rev Microbiol*. 2022;21(3):195-210. doi:10.1038/s41579-022-00807-9

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

47. Sabastin M, Mone K, Manivel A, Cherayi Padinakarai A, Krishnasamy K. Epidemic profile of common respiratory viruses in association SARS CoV-2 among SARI and ARI-two year study. *Mol Biol Rep*. 2024;51(1):156. doi:10.1007/s11033-023-09084-y
48. Matera L, Manti S, Petrarca L, et al. An overview on viral interference during SARS-CoV-2 pandemic. *Front Pediatr*. 2023;11:1308105. doi:10.3389/fped.2023.1308105
49. Rayne F, Wittkop L, Bader C, et al. Rapid Adenovirus typing method for species identification. *J Virol Methods*. 2017;249:156-160. doi:10.1016/j.jviromet.2017.09.009
50. Dhingra A, Hage E, Ganzenmueller T, et al. Molecular evolution of human adenovirus (HAdV) species C. *Sci Rep*. 2019;9(1):1039. doi:10.1038/s41598-018-37249-4
51. Wu X, Lu X, Schneider E, et al. Reassessment of high prevalence human adenovirus detections among residents of two refugee centers in Kenya under surveillance for acute respiratory infections. *J Med Virol*. 2019;91(3):385-391. doi:10.1002/jmv.25320
52. Metzgar D, Osuna M, Yingst S, et al. PCR analysis of Egyptian respiratory adenovirus isolates, including identification of species, serotypes, and coinfections. *J Clin Microbiol*. 2005;43(11):5743-5752. doi:10.1128/JCM.43.11.5743-5752.2005
53. Echavarria M, Maldonado D, Elbert G, Videla C, Rappaport R, Carballal G. Use of PCR to demonstrate presence of adenovirus species B, C, or F as well as coinfection with two adenovirus species in children with flu-like symptoms. *J Clin Microbiol*. 2006;44(2):625-627. doi:10.1128/JCM.44.2.625-627.2006
54. Sant'Anna TB, Araujo NM. Adeno-associated virus infection and its impact in human health: an overview. *Virol J*. 2022;19(1):173. doi:10.1186/s12985-022-01900-4
55. Trapani S, Caporizzi A, Ricci S, Indolfi G. Human bocavirus in childhood: a true respiratory pathogen or a "Passenger" virus? A comprehensive review. *Microorganisms*. 2023;11(5):1243. doi:10.3390/microorganisms11051243
56. Asner SA, Science ME, Tran D, Smieja M, Merglen A, Mertz D. Clinical disease severity of respiratory viral co-infection versus single viral infection: a systematic review and meta-analysis. *PLoS One*. 2014;9(6):e99392. doi:10.1371/journal.pone.0099392
57. Lin GL, Drysdale SB, Snape MD, et al. Targeted metagenomics reveals association between severity and pathogen co-detection in infants with respiratory syncytial virus. *Nat Commun*. 2024;15(1):2379. doi:10.1038/s41467-024-46648-3
58. Oldhoff E, Bennet R, Eriksson M, Allander T. Human bocavirus 1 epidemiology in children in relation to virus load and codetection. *Acta Paediatr*. 2023;112(5):1049-1055. doi:10.1111/apa.16699
59. Li Y, Pillai P, Miyake F, Nair H. The role of viral co-infections in the severity of acute respiratory infections among children infected with respiratory syncytial virus (RSV): a systematic review and meta-analysis. *J Glob Health*. 2020;10(1):010426. doi:10.7189/jogh.10.010426
60. Jiang W, Yin F, Zhou W, Yan Y, Ji W. Clinical significance of different virus load of human bocavirus in patients with lower respiratory tract infection. *Sci Rep*. 2016;6(1):20246. doi:10.1038/srep20246
61. Flotte TR, Berns KI. Adeno-associated virus: a ubiquitous commensal of mammals. *Hum Gene Ther*. 2005;16(4):401-407. doi:10.1089/hum.2005.16.401
62. Chen Z, Azman AS, Chen X, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nature Genet*. 2022;54(4):499-507. doi:10.1038/s41588-022-01033-y

How to cite this article: Hikmat H, Le Targa L, Boschi C, et al. Sequencing and characterization of human bocavirus genomes from patients diagnosed in Southern France between 2017 and 2022. *J Med Virol*. 2024;96:e29706. doi:10.1002/jmv.29706

10969071, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jmv.29706 by Cochrane France, Wiley Online Library on [27/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Article n°8 :
Five-year (2017-2022) evolutionary dynamics of human coronavirus HKU1 in southern France with emergence of viruses harboring spike H512R substitution

Houmadi Hikmat, **Lorlane Le Targa**, Céline BOSCHI, Justine PY, Aurélie Morand, Jean-Christophe Lagier, Sarah Aherfi, Jacques Fantini, Bernard La Scola, Philippe Colson (En cours de review dans Journal of Medical Virology)

TITLE PAGE

Type of article: Full-length article

Five-year (2017-2022) evolutionary dynamics of human coronavirus HKU1 in southern France with emergence of viruses harboring spike H512R substitution

Short title (for the running head):

Genomic evolution of human coronavirus HKU1 in southern France

Author list: Houmadi HIKMAT[1,2], Lorlane LE TARGA[1,2,3], Céline BOSCHI[1,2,4], Justine PY[1,2], Aurélie MORAND[4,5,6], Jean-Christophe LAGIER[1,2,4], Sarah AHERFI[1,2,4], Jacques FANTINI[7], Bernard LA SCOLA[1,2,4*], Philippe COLSON[1,2,4*]

Affiliations:

[1] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Marseille, France; [2] IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [3] Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France; [4] Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France; [5] Service d'accueil des Urgences Pédiatriques, hôpital Nord, AP-HM, chemin des Bourrely, 13015 Marseille, France; [6] Service de Pédiatrie générale, hôpital Timone, AP-HM, 264 rue Saint-Pierre, 13005 Marseille, France; [7] INSERM UMR UA 16, Aix-Marseille Université, Marseille, France

*** Corresponding authors:** Bernard La Scola and Philippe Colson, IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: bernard.la-scola@univ-amu.fr, philippe.colson@univ-amu.fr

Word counts: abstract, 200; text, 5,333

Figures: 7; **Tables:** 1; **References:** 47

Supplementary material: Figure: 1; Tables: 5

Keywords

Coronavirus; HKU1; respiratory virus; genome; evolution; molecular epidemiology, TMPRSS2

ABSTRACT

Objectives

HCoV-HKU1 diversity and evolution was scarcely studied. We performed next-generation sequencing (NGS) and analysis of HCoV-HKU1 genomes over five years.

Methods

NGS used Illumina technology on NovaSeq 6000 following whole genome PCR amplification by a in-house set of primers designed using Gemi and PrimalScheme. Genome assembly and analyses used CLC Genomics, Mafft, BioEdit, Nextstrain, Nextclade, MEGA and iTol bioinformatic tools. Spike molecular modelling and dynamics simulations used Molegro Molecular Viewer/Hyperchem.

Results

Twenty-eight PCR systems allowed obtaining 158 HCoV-HKU1 genomes including 69 and 89 of genotype A and B, respectively. Both genotypes co-circulated during the study period but one predominated each year. 1,683 amino acid substitutions including 80 in ≥10 genomes were detected in genotype A relatively to a 2004 reference. H512R in spike, first detected in 2009 and reported as involved in antibody neutralization, was found in all genotype A, almost always with V387I and K478N, and was predicted here to significantly improve cellular TMPRSS2 protein binding. 1,802 amino acid substitutions including 64 in ≥10 genomes were detected in genotype B relatively to a 2005 reference.

Conclusion

This study substantially expands the global set of HCoV-HKU1 genomes. Genomics with protein structural analyses contributed to our understanding of HCoV-HKU1 evolution.

HIGHLIGHTS

- 158 HCoV-HKU1 genomes were obtained for the 2017-2022 period in southern France
- Expansion by around 1.5 times of the global set of HCoV-HKU1 genomes
- Obtaining of all but one genomes available for France
- Spike H512R was harbored by all genotype A genomes, combined with V387I and K478N
- H512R was predicted here to significantly improve binding to cellular TMPRSS2

TEXT

INTRODUCTION

Seven coronaviruses are currently known to cause respiratory infections in humans, of which four (human coronaviruses (HCoV)-OC43, 229E, NL63, and HKU1) are endemic [1,2]. Coronaviruses have a wide host spectrum including humans and non-human animals, and are capable of crossing the species barriers, such as for the case of HCoV-HKU1 that originated from a murine host [3]. HCoV-HKU1 was discovered in China in 2004 in a 71-year-old patient with comorbidities who exhibited cough and fever [4]. This virus belongs to the *Betacoronavirus* genus. The viral genome is approximately 30 kilobase–long and its two-thirds consist in ORF1a and 1b, which encode the set of non-structural proteins (NSP) 1 to 16. The remaining third of the genome encode the structural proteins, namely the spike, the matrix, the viral envelope and the nucleocapsid. One of the characteristics of HCoV-HKU1 is the presence of a gene encoding an hemagglutinin esterase (HE), as for the case of some betacoronaviruses among which HCoV-OC43, but which is absent in alphacoronavirus genomes. Three genotypes of HCoV-HKU1, named A, B and C, have been described [4,5]. The protein receptor for HCoV-HKU1 is still unknown, but the viral spike binds to 9-O-acetylated sialic acid of the host cell , and HKU1 is thought to have TMPRSS2 as a receptor involved in fusion [6,7].

HCoV-HKU1 is responsible for infections of the upper respiratory tract in most cases, but can also affect the lower respiratory tract [8,9]. Symptomatology is variable and comparable to those of other respiratory viruses, with such symptoms as fever, cough, rhinitis, and dyspnea [4,10]. The seasonality of this virus varies depending on the geographical area. In temperate countries such as in France, the virus is mainly detected during winter, and spring [11–13].

Unlike for other viruses such as SARS-CoV-2 and influenza viruses, there are few evolutionary studies reported for HCoV-HKU1. We believe that understanding the dynamics of HCoV-HKU1 infections requires an understanding of the genetic evolution of these viruses, but as of 04/12/2023, there were only 106 complete genomes (between 29,367 and 30,144 nucleotides in size) of HCoV-HKU1 in the NCBI GenBank nucleotide sequence database [14] (https://www.ncbi.nlm.nih.gov/genbank/), which is very few with regard to genomic and genotypic surveillance, and dramatically low compared to the numbers of SARS-CoV-2 complete genomes in GenBank and in GISAID which are more than 8 and 15 millions, respectively [15,16] . Therefore, here we aimed to carry out retrospectively the sequencing and analysis of HCoV-HKU1 genomes from samples of patients diagnosed as HCoV-HKU1 RNA-positive by real-time PCR (qPCR) during the period 2017-2022 at the university and public hospitals of Marseille, southern France.

MATERIALS AND METHODS

Respiratory samples

Nasopharyngeal samples analyzed retrospectively were sent to the clinical microbiology-virology laboratory of university and public hospitals of Marseille, southeastern France, for routine clinical diagnosis of respiratory virus infections. They were collected over a 67 month-period that extends from February 2017 to August 2022. RNA was extracted from nasopharyngeal samples using the KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's instructions. HCoV-HKU1 was diagnosed by real-time reverse transcription-PCR (qPCR) using either the FTD Respiratory pathogens 21 (FTD) assay (Fast Track Diagnosis, Luxembourg) or the BioFire FilmArray Respiratory panel 2 plus (BioFire) assay (Biomérieux, Marcy-l'Etoile, France). The present study has been registered on the Health Data Access Portal of Marseille public and university hospitals

(Assistance Publique-Hôpitaux de Marseille (AP-HM)) with No. PADS24-190 and was approved by the Ethics and Scientific Committee of AP-HM under No. CSE24-4.

Next-generation sequencing

Primer design for PCR amplification of overlaping regions covering the whole viral genome

All HCoV-HKU1 complete genomes available in Genbank (https://www.ncbi.nlm.nih.gov/genbank/) were collected in April 2022. They were aligned using the Mafft software [17]. Primers targeting the regions the most conserved within the obtained alignment were designed using the GEMI software, which allowed us to choose amplicon size (447-1587 nucleotides), primer size (19-23 nucleotides), temperature of the PCR hybridization step (58-60°C), and maximum number of degenerated nucleotide positions (1-4) to handle genetic diversity in the set of genomes [18]. The PrimalScheme software (https://primalscheme.com/), which previously allowed designing the mutiplex PCR systems for SARS-CoV-2 ('ARTIC', https://artic.network/ncov-2019) and Zika virus [19], was also used here. Unlike GEMI, PrimalScheme only enables the choice of amplicon size.

PCR amplification of overlaping regions covering the whole genomes and next-generation sequencing

The KingFisher Flex system (Thermo Fisher Scientific, Waltham, MA, USA) was used to extract viral RNA from the residual nasopharyngeal samples, following the manufacturer's instructions. Two pools were produced by mixing the previously designed PCR primers following an 'Artic-like' methodology priorly implemented for Zika virus [20] and SARS-CoV-2 (https://artic.network/). Reverse transcription-PCR amplification was performed using the SuperScript III One-Step with High-Fidelity Platinum Taq DNA Polymerase kit in a reaction containing 12.5 µL of 2X mix, 0.7 µL of enzyme, 0.75 µL of each PCR pool, 3 µL of nucleic acid extract, and 8.05 µL of water. The PCR programme used included a reverse transcription step at 50°C for 25 minutes, then an initial denaturation step

146 at 95°C for 2 minutes, followed by 39 PCR cycles comprising a denaturation step at 95°C for
147 15 sec, a primer hybridization step at 58°C for 45 sec, and an elongation step at 70°C for 2
148 min, before a final elongation step at 70°C for 5 min. Purification of the amplicons was
149 performed on a NucleoFast 96-well plate (Macherey Nagel ref 743100.50, Hoerdt, France);
150 elution was carried out with 40 µL of pure water. Once purified, the amplicons generated
151 from the two pools of PCR systems were mixed together.
152 For the setting up of our system, NGS was performed with the Oxford Nanopore
153 Technology using the Ligation Sequencing Kit (SQK-LSK109) library on a SpotON flow cell
154 Mk I, R9.4.1 and a GridION instrument, following the manufacturer's instructions (Oxford
155 Nanopore Technologies, Oxford, UK). Thereafter, at the step of using our ready to use 'Artic-
156 like' system on the studied nasopharyngeal samples, NGS was performed with the Illumina
157 technology on a NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA) using the
158 COVIDSeq protocol but replacing the COVID-19 ARTIC PCR primers by the PCR primers
159 designed here. Loading of the NovaSeq 6000 instrument was performed on a SP flow cell
160 according to the NovaSeq-XP workflow with a reading of 2x50, following a previously
161 described procedure [21]. The HCoV-HKU1 genomes obtained were submitted to the GenBank
162 database (https://www.ncbi.nlm.nih.gov/genbank/; GenBank Accession no: PP356751 -
163 PP356908).
164 **Bioinformatic analyses of viral genomes**
165 The CLC Genomics Workbench v.7 software was used for the bioinformatics analysis
166 of our sequences. A trimming step was first performed followed by the mapping of next-
167 generation sequencing (NGS) reads against the HCoV-HKU1 reference genome GenBank
168 Accession no. NC_006577.2, that was performed with as minimum thresholds a coverage of
169 80% of reads on the reference genome and a nucleotide identity of 90%. All obtained
170 consensus genomes with at least 80% coverage of the reference genome and a sequencing

depth of at least 3 were recovered. The Mafft software [17] was used for sequence alignments and the MEGA v.11 software [22] was used to perform phylogeny reconstruction using the Neighbor-Joining and Maximum composite likelihood parameter methods with 1,000 replicates. Phylogenetic analysis was performed by including, in our dataset of sequences, genomes available in GenBank that belong to the three HCoV-HKU1 genotypes A, B and C. iTOL [23] was used to visualize phylogenetic trees. The Nextclade tool (https://clades.nextstrain.org/) [24] was adapted to be able identifying the mutations and the different lineages to which the genomes we obtained belong. Nucleotide and amino acid diversities as well as mutational patterns were determined in obtained genomes according to the viral genotype, relatively to two references genomes of genotypes A and B. For each genotype, an historical reference HCoV-HKU1 genome, NC_006577.2 for genotype A and DQ415911.1 for genotype B, dating back to 2004 and 2005, respectively, was used. In addition, more recent reference genomes, LC315650.2 for genotype A and LC315651.2 for genotype B, dating back to 2014 and 2016, respectively, were used. We used the same methodology than in a previous study [25] to classify genes into "structural", "informational", "other non-structural" and "accessory" genes (Supplementary methods).

Molecular modelling of the HCoV-HKU1 spike

The coordinates of HCoV-HKU1 spike proteins (both monomers and trimers) were retrieved from pdb 8ohn (open conformation) and 8opn (closed conformation) [26]. The structure of human TMRPSS2 was retrieved from pdb 7meq. The structure of the TMPRSS2-HKU1 complex was retrieved from pdb 8vgt. The models were visualized with the Molegro Molecular Viewer (http://molexus.io/molegro-molecular-viewer), as previously described [27]. Energy minimized models were generated with the Polak-Ribiere algorithm with the Bio-CHARMM force field in Hyperchem using a maximum of 3×10^5 steps and a root mean square (RMS) gradient of 0.01 kcal/mol.Å as the convergence condition [28]. Mutations were

196 introduced with Deep View/Swiss-Pdb viewer [29], followed by several rounds of energy
197 minimization as described previously [30,31]. Molecular dynamics simulations (MDS) were
198 performed on a Dell workstation with the HyperChem program
199 (http://www.hypercubeusa.com), as previously described [32]. The systems were equilibrated at
200 constant temperature (310 K) and constant pressure (1 atm) [28].
201
202 **RESULTS**
203 **Primer design for PCR amplification of overlaping regions covering the whole HCoV-**
204 **HKU1 genomes**
205 For primer design, the Gemi software generated 156 PCR primer pairs. Of these 156,
206 36 were tested and 25 selected as they provided PCR amplicons in sufficient amount and of
207 appropriate size as assessed by gel agarose electrophoresis. Three primer pairs designed using
208 the PrimalScheme software were added. Finally, 28 PCR systems covering the genome were
209 selected and used (Supplementary Table S1). Those showing weak bands on a 1.5% agarose
210 gel were optimised by progressively increasing the primer concentration. These latter ranged
211 between 10 and 15 pmol/µL. Two pools were then made with these 28 primer systems so as
212 not to have unplanned hybridization during PCR. The Oxford Nanopore technology and a
213 Gridion instrument were used for testing and optimizing the PCR amplifications, then all
214 processed samples were sequenced using the Illumina technology on a NovaSeq 6000
215 instrument.
216 **Next-generation sequencing**
217 The whole HCoV-HKU1 nucleic acid enrichment and NGS strategy was applied to
218 RNA extracts recovered from HCoV-HKU1 RNA-positive nasopharyngeal samples. A total
219 of 803 nasopharyngeal samples, which were tested in the diagnosis setting in our institute
220 between January 2017 and December 2022 by PCR assays targeting respiratory viruses other

than SARS-CoV-2, were positive for HCoV-HKU1. In 2020 and 2022, the HCoV-HKU1 epidemic peaks occurred between January and March, while no HCoV-KHU1 infection was diagnosed between July 2020 and June 2021 (Figure 1). Among the 803 HCoV-HKU1-positive samples, we could retrieve 411 from our freezers (-80°C storage) and all their RNA extracts were submited to the ARTIC-like amplification with our PCR primers' pools. The cycle threshold (Ct) of the qPCR used to diagnose HCoV-HKU1 infection was known for 293 of these 411 samples and ranged between 13 and 36 with a mean (± standard deviation) value of 24.8±5.2. For these 411 samples, a coverage of at least 80% of genome GenBank Accession no. LC315650.2 with a NGS depth of at least 3X was obtained in 158 cases (38%). The qPCR Ct was known for 128 of these 158 samples and ranged between 10 and 32 with a mean value of 20.8 ± 4.2. These 158 obtained genomes were generated from nasopharyngeal samples collected between October 2017 and May 2022.The number of NGS reads ranged between 15,028 and 10,388,629 with a mean number of 2,086,518 reads. Genome coverage ranged between 81.05% and 99.90% with a mean value of 96.09 ± 4.65%. The remaining 253 samples not included in our analyses had a genome coverage lower than 80% compared with the reference. The Ct of these 253 non-retained samples were known for 165 samples and ranged between 20 and 36 with an average of 27.9 ± 3.7.

Phylogeny of HCoV-HKU1 genomes

Of the 158 HCoV-HKU1 selected genomes with a coverage ≥80%, 69 were clustered with genotype A genomes and 89 were clustered with genotype B genomes (Figure 2). None of our sequences belonged to HCoV-HKU1 genotype C. These results were also confirmed using the Nextclade tool [24]. Pairwise nucleotide diversity between our genotype A genomes varied between 70.0% and 99.9% with a mean of 92.2 ± 6.1%. Regarding our B genotype genomes, this diversity varied between 72.5% and 99.9% with a mean of 93.2 ± 5.8%. When we examined the distribution of viral genotypes by year, we observed that both genotypes A

and B circulated at the same time, but with a predominance of one the two genotypes during a given year. Thus, genotype A predominated in 2020 while genotype B predominated in 2022 (Figure 3).

HCoV-HKU1 mutational patterns

Genotype A sequences

A total of 5,686 nucleotide substitutions, 449 deletions and 848 insertions were detected in genotype A sequences obtained here compared with the reference genome NC_006577.2 dating back to 2004. This corresponds to a substitution rate of $6.4\times10^{-4}\pm3.9\text{x}10^{-4}$ per nucleotide per year. To gain a better understanding of these mutations and their putative impact, we considered gene categories encompassing "structural", "informational", "other non-structural" and "accessory" genes. In the informational genes category, nsp12 that encodes the RNA-dependent RNA polymerase [33] was the gene the most affected with 339 substitutions per 1,000 nucleotides compared to the reference, considering that only 22 substitutions per 1,000 nucleotides are present in at least 10 genomes; and the least affected gene was nsp15 that encodes an endoribonuclease with 49 substitutions per 1,000 nucleotides. Regarding structural genes, the gene encoding hemagglutinin esterase was the one the most affected by nucleotide substitutions, with 299 substitutions per 1,000 nucleotides, followed by the spike gene with 183 substitutions per 1,000 nucleotides, and the gene the least affected was the one encoding the envelope protein, with 12 substitutions per 1,000 nucleotides. The accessory genes exhibited 57 substitutions per 1,000 nucleotides (Supplementary Table S2).

For these genomes belonging to genotype A, a total of 1,683 amino acid substitutions were reported relatively to reference genome NC_006577.2. This corresponds to a substitution rate of $3.0\times10^{-4}\pm1.9\times10^{-4}$ per nucleotide per year. The number of amino acid substitutions in the genomes ranged from 58 to 635 with a mean value of 143.1±93.6. The number of deletions ranged from 0 to 120 with a mean value of 22.1±21.5 and the number of

271 insertions ranged from 0 to 115 with a mean value of 8.8±16.4. In the informational gene
272 category, nsp12 had the highest diversity with 303 amino acid substitutions per 1,000 amino
273 acids, while nsp7 had none. In the structural gene category, the gene with the highest diversity
274 was HE followed by the spike gene, with 298 and 146 substitutions per 1,000 amino acids,
275 respectively (Supplementary Table S2). Of the 1,683 amino acid substitutions, only 80 were
276 present in at least 10 genomes. The genes encoding the greatest number of these 80 amino
277 acid mutations were nsp3, which encodes a papain-like protease (n= 22), the spike gene (n=
278 16), and the nucleocapsid gene (n= 14). Nsp5, nsp6, nsp10, nsp11, nsp13, nsp14, nsp16, E
279 and ORF4 genes harbored none of the amino acid substitutions present in at least 10 genomes
280 (Figure 4A).
281 The H512R mutation observed relatively to the 2004 reference genome
282 (NC_006577.2) in the spike gene of genotype A genomes obtained here was reported to be
283 involved in antibody neutralization [34]. All HCoV-HKU1 genotype A spikes covering this
284 position were retrieved from Genbank and analyzed with those obtained here. According to
285 GenBank data, the H512R mutation appeared in 2009 and became the majority mutation
286 among genomes over time (Figure 5A). For year 2009, only one sequence (11%) was detected
287 as carrying this mutation. Between 2010 and 2016, 20/33 (60%) of the sequences carried this
288 mutation. No genotype A sequences were detected for year 2017 either in Genbank or in the
289 present study. For years 2018 and 2019, all 14 sequences analysed both in GenBank and the
290 present study harbored this mutation. For year 2020, only one sequence was found in
291 GenBank and it did not carry this mutation, whereas all 45 sequences from the present study
292 carried this mutation. Finally, for 2021 and 2022, all 10 sequences obtained here carried the
293 mutation (Table 1). The first sequence carrying the H512R mutation detected in 2009
294 originated from the United States, while overall genomes deposited in GenBank and carrying
295 this mutation originated from the USA, Thailand, China, Japan and Russia; and the present

296 study further show the presence of this mutation in genotype A sequences circulating in the
297 south of France.
298 Another amino acid mutation in the spike, V387I, was observed in the sequences
299 obtained here. Hence, we collected from GenBank all HCoV-HKU1 spike sequences that
300 covered this amino acid position. We observed that this mutation appeared in 2009, being
301 present in a single sequence out of 13 from GenBank (Figure 5B). Thereafter, between 2010
302 and 2016, this mutation was observed in 3 (8%) of 38 sequences from GenBank. For years
303 2018, 2019 and 2020, all 63 sequences (from GenBank and the present study) carried this
304 mutation. For 2021, the two sequences from the present study and 4 of the 5 sequences from
305 GenBank carried this mutation. Finally, for 2022, the 4 sequences from GenBank and from
306 the present study all carried this mutation (Table 1). Regarding its geographical distribution,
307 this mutation was observed for the first time in the USA then was also observed in China,
308 Japan, and Russia, and in France as shown in the present study. K478N was also present in
309 the spike of genotype A sequences obtained here. It was first detected in sequences available
310 from GenBank in 2018 and continued to circulate since then (Figure 5C). All 64 sequences
311 obtained for years 2018, 2019, 2020 and 2022 carried this mutation, and it was the case for
312 2021 for 4 of the 5 sequences available from GenBank and the two genotype A sequences
313 from the present study (Table 1). Overall, this mutation was observed in China, Japan, the
314 USA, and France.
315 We then looked at whether the H512R mutation was linked to V387I and/or K478N
316 mutations. We first determined the proportion of sequences with the H512R and V387I
317 mutations (Table 1). In 2009, the only sequence with the H512R mutation also carried V387I.
318 For sequences obtained from samples collected between 2010 and 2016, only three of 29
319 sequences carried these two mutations combined. Then, all 69 sequences between 2018 and
320 2022 available from GenBank or from the present study carried these two mutations

321 combined. Regarding mutations H512R and K478N, all 69 sequences carried these two
322 mutations combined (Table 1). Regarding mutations V387I and K478N, they were combined
323 in all 69 sequences except in 1 of the 5 genomes available from GenBank for 2021;
324 unfortunately, these latter genomes did not cover spike amino acid 512 (Table 1). Finally, to
325 investigate fruther how these three mutations (H512R, V387I and K478N) were linked, we
326 recovered all the sequences that covered their positions, and determined that all 69 such
327 sequences harbored these three mutations combined (Table 1). Additional comparative
328 genomics results can be found in Supplementary results.***Genotype B sequences***
329 For the HCoV-HKU1 genotype B genomes, 6,129 nucleotide substitutions, 428
330 deletions and 801 insertions relatively to the reference genome DQ415911.1 dating back to
331 2005 were observed. This corresponds to a substitution rate of $4.4\times10^{-4}\pm4.0\times10^{-4}$ per
332 nucleotide per year. Among informational genes, the gene with the greatest nucleotide
333 diversity was nsp12 with 290 substitutions per 1,000 nucleotides. In the non-structural gene
334 category, the nsp9 gene showed the greatest diversity (424 substitutions per 1,000
335 nucleotides) and nsp1 showed the lowest diversity (54 substitutions per 1,000 nucleotides).
336 Among structural genes, the HE gene showed the greatest diversity with 475 substitutions per
337 1,000 nucleotides, followed by the spike gene with 211 substitutions per 1,000 nucleotides.
338 Accessory genes exhibited 160 substitutions per 1,000 nucleotides overall (Supplementary
339 Table S4).
340 Compared with this reference genome DQ415911.1, 1,802 amino acid substitutions
341 were detected within genome sequences obtained here. This corresponds to a substitution rate
342 of 1. $1\times10^{-4}\pm0.8\times10^{-4}$ per nucleotide per year. The number of amino acid substitutions ranged
343 from 22 to 245 with a mean value of 60.3±44.4;.the number of deletions ranged from 0 to 46
344 with a mean value of 8.3±9.6; and the number of insertions ranged from 0 to 36 with an
345 average of 2.8±5.3. Among informational proteins, nsp14 was that with the greatest number

346 of substitutions per 1,000 amino acids (n= 305) and nsp15 was that with the lowest number of
347 mutations (n= 117). Among structural proteins, HE showed the greatest diversity followed by
348 the envelope, with 382 and 292 substitutions per 1,000 amino acids, respectively. The spike
349 harbored 103 substitutions per 1,000 amino acids. The accessory protein coding for ORF4
350 exhibited 183 substitutions per 1,000 amino acids (Supplementary Table S4). Of all the 1,802
351 amino acid substitutions, 64 were found in at least 10 genomes. Proteins with no amino acid
352 substitution in at least 10 genomes were nsp1, nsp5, nsp6, nsp8, nsp10, nsp11, nsp13, nsp15,
353 E and ORF4 (Figure 4B). Additional comparative genomics results can be found in
354 Supplementary results.

355 **Molecular modelling of HCoV-HKU1 spikes**

356 We then carried out an *in silico* molecular modelling of the HCoV-spike protein with
357 the H512R mutation and the two other mutations, V387I and K478N, observed in genotype A
358 genomes obtained here. In the closed state, the receptor binding motif (RBM) of the spike
359 protein is initially partially masked and definitely not accessible to the TMPRSS2 receptor
360 (Figure 6A). It has been assigned to a region globally encompassing amino acids 505-532,
361 with the addition of residues 488 and 554 [35]. The H512R mutation is close to critical amino
362 acids that control binding to the receptor, such as H488 and W515 (Figure 6B). Since H512 is
363 exposed on the surface of the spike protein, its substitution by a cationic amino acid (arginine)
364 is perfectly tolerated by the protein trimer. Indeed, there is very little difference between the
365 3D structures of wild-type and H512R mutant trimers in the closed conformation. Once bound
366 to sialic acids, the trimer undergoes a dramatic conformational rearrangement that propels the
367 receptor binding domain (RBD) towards the TMPRSS2 receptor (Figure 6C). As expected,
368 both the wild-type H512 (Figure 6C) and mutant H512R (Figure 6D) are accessible for the
369 receptor in the open conformation of the trimeric spike.

370 The H512R mutation has a profound impact on the electrostatic surface potential of

371 the RBM (Figure 7 A). Compared with the wild-type trimer, the H512R mutant is
372 characterized by an increased electropositive area at the tip of the RBM. This increase in
373 electropositivity is then partially compensed by secondary mutations V387I and K478N, but
374 not at the tip of the RBD whose electrostatic surface potential remains highly positive. We
375 used molecular modeling approaches to predict the impact of these mutations on the
376 TMPRSS2-HKU1 complex (Figure 7 B-E). Only H512R appeared to be involved in the
377 formation of this complex (Figure 7 B-C), since both K478N and V387I were clearly outside
378 the interaction site. These observations confirmed that K478N and V387I are secondary
379 mutations that finely tune the architecture of the RBM already bearing H512R.
380 During the binding reaction, we detected the formation of an electrostatic bond
381 between the gamma-carboxylate group of E505 (HKU1 RBM) and the cationic guanidinium
382 group of R470 (TMPRSS2), in full agreement with recently published cryoelecromicroscopy
383 data [35]. We also observed a similar electrostatic bond between the beta-carboxylate group of
384 D511 and the cationic group of H512R (Figure 7 D-E). This intramolecular bond prevents the
385 electrostatic repulsion between D417 (TMPRSS2) and D511 (HKU1 RBM) and stabilizes the
386 underneath loop displaying H488 (see comparison of Figures 7D and 7E). Since H488
387 interacts with D417 in the TMPRSS2-HKU1 complex [35], a conformational adjustment is
388 required to bring both amino acids in close contact. Indeed, the cationic group of H512R
389 initiates a rearrangement of the binding interface, resulting in the formation of a network of
390 electrostatic interactions involving H512R, D511, H488 (HKU1) and D417 (TMPRSS2). This
391 is predicted to induce a slight conformational change that allows the formation of an
392 electrostatic interaction between E554 (HKU1) and R413 (TMPRSS2). Overall, the H512R
393 mutation is predicted to affect both the kinetics and the affinity of the TMPRSS2-HKU1
394 complex, explaining the selection of this mutant over the wild-type virus.
395

DISCUSSION

Coronaviruses can cross the species barrier into humans from other animals, making them candidates for emergence events, as observed with the recent SARS-CoV-2 pandemic [1,36,37]. Here we analyzed 158 genomes of HCoV-HKU1, which is one of the four coronaviruses that were endemic before the SARS-CoV-2 emergence. This set of genomes represents about 1.5 times the number of genomes previously available for this virus, in the GenBank database, which was, and remains a dramatically small set in comparison with more than 16 million SARS-CoV-2 genomes available. In addition, in GenBank, most HCoV-HKU1 genomes are from samples originating from the USA and China, and for the 2017-2022 period, there was only 55 HCoV-HKU1 genomes available in GenBank. Moreover, a single genome was available from France, which dates back to 2005. Thus, our study carried out on nasopharyngeal samples collected between 2017 and 2022 substantially expands the number of genomes available for this virus at the global scale and is the one that reports the largest number of genomic sequences.

Most of the viral genomes described here were obtained from samples collected in 2020 and 2022. This can be explained by the SARS-CoV-2 pandemic that led to test massively people with symptoms of respiratory viruses. Most of the respiratory samples from which genomes were obtained were collected between January and March months, with peaks between January and February, which is in line with studies showing that in temperate countries HCoV-HKU1 incidence peaks occur during these two latter months [11–13]. No infection with HCoV-HKU1 was diagnosed in our clinical microbiology laboratory between July 2020 and June 2021, which may be explained in part by the public health measures implemented during this period such as confinement and mask wearing that might have resulted in a reduction of cases, as previously reported in our area [38].

We identified HCoV-HKU1 of genotypes A and B. Co-circulation of these two

421 genotypes has been reported in many studies [5,8,9,39,40]. Even if there was a co-circulation of the
422 two genotypes every year in previous studies and the present study, there was a predominance
423 of one genotype over the other according to the year. Besides, no C genotype has been
424 detected since 2005 [40]. Substitution rates were generally comparable to those found in the
425 literature for betacoronaviruses, which is of the order of 1.8×10^{-4} substitutions per site per
426 year [41]. Regarding nucleotide substitutions, whatever the genotype studied, at the level of
427 informational genes, the Nsp12 gene that encodes the RNA-dependent RNA polymerase is
428 the one that exhibited the greatest nucleotide diversity, which was not reported to be the case
429 for SARS-CoV-2 in a previous study [25]. Nevertheless, most of the nucleotide mutations
430 detected in Nsp12 were only found in a few genomes. In the category of structural genes, the
431 envelope encoding gene showed the lowest diversity among genotype A genomes. In a
432 previous study that compared genomes of genotypes A to those of genotypes B and C, the
433 genes exhibiting the greatest diversity were those encoding the envelope, the spike, and the
434 HE, while the other genes were relatively conserved [40]. The present data are consistent with
435 those of previous studies showing that the spike protein is encoded by the most variable gene
436 within this virus [42]. For genotype A sequences, some amino acid substitutions (L28K, S70A,
437 V387I, K478N) in the spike we observed here were previously reported [40] and are located in
438 the S1 subunit. Besides, spike amino acids V509, L510, D511, H512, W515 or R517 were
439 reported to have an impact on the sensibility to immune responses based on structural
440 prediction [34].
441 In the present study, at one of these positions, amino acid substitution H512R was
442 detected, which was previously reported [43]. After analyzing all the HCoV-HKU1 spike
443 sequences available in GenBank as well as the sequences obtained here, it appeared that this
444 mutation was present for the first time in early 2009, and that 512R thereafter become over
445 time the majority amino acid, and almost the only one in circulating A genotype genomes.

446 This H512R mutation has been detected in the USA, China, Thailand, Japan and in France as
447 shown in this study. This mutation was first observed in association with another mutation,
448 V387I, but we determined that these two mutations were not always associated. Mutation
449 K487N was present in genomes obtained from more recent samples, in 2018. Then, since
450 2018, the majority of genotype A sequences harbored these 3 mutations combined. The
451 present study points out that H512R can occur alone but tend to be accompanied by mutations
452 V387I and K478N. Although, the low number of sequences available in GenBank and
453 through our study may lead to an inaccurate representation of the global HCoV-HKU1
454 genome diversity, we may consider H512R, and to a lesser extent mutations V387I and
455 K478N, as "hyperfertile", similar to a previous study conducted on SARS-CoV-2 [25].
456 Structural studies of the three mutations H512R, V387I and K478N revealed new
457 information on the evolution of the HKU1 coronavirus. H512R, is well tolerated by the
458 trimeric spike. The appearance of a cationic charge located at the tip of the RBM increases the
459 electrostatic surface potential of the trimeric spike. This effect has been previously observed
460 with other RNA viruses such as HIV-1 and SARS-CoV-2, which use lipid rafts as primary
461 adhesion sites on host cells [44]. The recent demonstration that sialic acid binding triggers
462 HKU1 spike opening [45] confirmed our model of a dual glycolipid-protein receptor for virus
463 entry mechanisms [46]. Since lipid rafts are enriched in gangliosides, which contain sialic acids,
464 they are negatively charged and attract more rapidly mutant virus with increased
465 electropositivity. This effect explains how the evolution of those viruses is driven by host cell
466 membranes [47]. However, compensatory mechanisms often occur to balance the global
467 electrostatic surface potential of mutant viruses, as remarkably shown for the Omicron lineage
468 of SARS-CoV-2. In this case, the electrostatic potential gradually increases in the RBD and
469 simultaneously decreases in the N-terminal domain (NTD) of the spike protein [47]. Yet the
470 compensation does not result in a return to the initial state since as the variants come one after

471 the other, the electrostatic potential becomes stronger and stronger, as observed from Alpha to
472 Delta SARS-CoV-2 variants [27]and then in the Omicron lineage [47]. In this respect, it is likely
473 that V387I and K478N, which appeared after the selection of H512R, are compensatory
474 mutations which finely tune the electrostatic surface potential of the trimeric spike. The
475 second important impact of H512R concerns the binding of TMPRSS2. Molecular dynamics
476 simulations confirmed the global geometry of the TMPRSS2-HKU1 complex and gave some
477 insights on the potential effects of the H512R mutation. We propose that H512R forms an
478 intramolecular electrostatic bond with D511, which minimizes the electrostatic repulsion
479 between D511 (HKU1) and D417 (TMPRSS2) (Figure 7 D-E). This mechanism is expected
480 to improve the initial binding steps between the spike trimer and the TMPRSS2 receptor,
481 which may then involve subtle conformational rearrangements. This mechanism i) facilitates
482 the interaction between H488 (HKU1) and D417 (TMPRSS2) and ii) reinforces the complex
483 through an electrostatic network of interactions involving H512R with D417. Overall, our
484 molecular modeling studies suggest that H512R can significantly improve TMPRSS2 binding
485 to mutant viruses.

486 The present study is not necessarily representative of what is happening worldwide. It
487 was carried out for a given location and a given period of time, and the number of data
488 available worldwide has been revealed here to be limited. Notwithstanding, it contributes to
489 our understanding of the evolution of HCoV-HKU1 and also considerably enrich the HCoV-
490 HKU1 database. We show here that genomic monitoring directly from respiratory samples
491 combined with protein structural predictions analysis can detect and explore the emergence of
492 HCoV-HKU1 lineages. Further works on a large number of samples would be necessary to be
493 able estimating whether or not the different HCoV-HKU1 genotypes may be associated with
494 particular clinical presentations. Also, future studies conducted in our institute or by other

teams should isolate several HCoV-HKU1 strains of various lineages for their phenotypic characterization.

Acknowledgments

We are very grateful to the technical team of the next-generation sequencing platform of IHU Méditerranée Infection.

Data Availability

HCoV-HKU1 genomes analysed here have been submitted to the GenBank sequence database [14] (https://www.ncbi.nlm.nih.gov/genbank/) (GenBank Accession no: PP356751 - PP356908).

Author contributions

Conceived and designed the experiments: BLS, PC. Contributed materials, analysis tools: HH, LLT, CB, JP, AM, JCL, SA, JF, BLS, PC. Analyzed the data: HH, CB, JF, BLS, PC. Writing—original draft preparation: HH, BLS, JF, BLS, and PC. Writing—review and editing: All authors. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Lorlane Le Targa works for Biosellal, a company located in Dardilly, France. The other authors have no conflicts of interest to declare. Funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Funding

This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03; and by the French Ministry of Higher Education, Research and Innovation

(Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation) and the French Ministry of Solidarity and Health (Ministère des Solidarités et de la Santé).

Ethics

The present study has been registered on the Health Data Access Portal of Marseille public and university hospitals (Assistance Publique-Hôpitaux de Marseille (AP-HM)) with No. PADS24-190 and was approved by the Ethics and Scientific Committee of AP-HM under No. CSE24-4.

REFERENCES

1 Corman Victor M., Muth Doreen, Niemeyer Daniela, Drosten Christian. Hosts and Sources of Endemic Human Coronaviruses. *Adv Virus Res* 2018;**100**:163–88. Doi: 10.1016/bs.aivir.2018.01.001.

2 Kakuya Fujio, Terao Ryuta, Onoda Hikaru, Okubo Hitoshi, Fujiyasu Hiroaki, Inyaku Fumie, et al. Epidemiology of endemic human coronavirus infection during the COVID-19 pandemic. *J Infect Chemother* 2024;**30**(5):400–5. Doi: 10.1016/j.jiac.2023.11.012.

3 Otieno James R., Cherry Joshua L., Spiro David J., Nelson Martha I., Trovão Nídia S. Origins and Evolution of Seasonal Human Coronaviruses. *Viruses* 2022;**14**(7):1551. Doi: 10.3390/v14071551.

4 Woo Patrick C. Y., Lau Susanna K. P., Chu Chung-ming, Chan Kwok-hung, Tsoi Hoi-wah, Huang Yi, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 2005;**79**(2):884–95. Doi: 10.1128/JVI.79.2.884-895.2005.

5 Woo Patrick C. Y., Lau Susanna K. P., Yip Cyril C. Y., Huang Yi, Tsoi Hoi-Wah, Chan Kwok-Hung, et al. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J Virol* 2006;**80**(14):7136–45. Doi: 10.1128/JVI.00509-06.

6 Hulswit Ruben J. G., Lang Yifei, Bakkers Mark J. G., Li Wentao, Li Zeshi, Schouten Arie, et al. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci* 2019;**116**(7):2681–90. Doi: 10.1073/pnas.1809667116.

7 Saunders Nell, Fernandez Ignacio, Planchais Cyril, Michel Vincent, Rajah Maaran Michael, Baquero Salazar Eduard, et al. TMPRSS2 is a functional receptor for human coronavirus HKU1. *Nature* 2023;**624**(7990):207–14. Doi: 10.1038/s41586-023-06761-7.

8 Jin Yu, Song Jing-Rong, Xie Zhi-Ping, Gao Han-Chun, Yuan Xin-Hui, Xu Zi-Qian, et al. Prevalence and clinical characteristics of human CoV-HKU1 in children with acute respiratory tract infections in China. *J Clin Virol Off Publ Pan Am Soc Clin Virol* 2010;**49**(2):126–30. Doi: 10.1016/j.jcv.2010.07.002.

9 Gerna Giuseppe, Percivalle Elena, Sarasini Antonella, Campanini Giulia, Piralla Antonio, Rovida Francesca, et al. Human respiratory coronavirus HKU1 versus other coronavirus infections in Italian hospitalised patients. *J Clin Virol* 2007;**38**(3):244–50. Doi: 10.1016/j.jcv.2006.12.008.

10 Su Shuo, Wong Gary, Shi Weifeng, Liu Jun, Lai Alexander C. K., Zhou Jiyong, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol* 2016;**24**(6):490–502. Doi: 10.1016/j.tim.2016.03.003.

11 Liu Peilin, Shi Lei, Zhang Wei, He Jianan, Liu Chunxiao, Zhao Chunzhong, et al. Prevalence and genetic diversity analysis of human coronaviruses among cross-border children. *Virol J* 2017;**14**:230. Doi: 10.1186/s12985-017-0896-0.

12 Gaunt ER, Hardie A., Claas ECJ, Simmonds P., Templeton KE. Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years Using a Novel Multiplex Real-Time PCR Method. *J Clin Microbiol* 2010;**48**(8):2940–7. Doi: 10.1128/JCM.00636-10.

13 Cui Li-Jin, Chen Zhang, Ting Zhang, Rou-Jian Lu, Zheng-De Xie. Human Coronaviruses HCoV-NL63 and HCoV-HKU1 in Hospitalized Children with Acute Respiratory Infections in Beijing, China. Available at https://www.hindawi.com/journals/av/2011/129134/. Accessed December 26, 2022, 2011.

14 Sayers Eric W., Cavanaugh Mark, Clark Karen, Pruitt Kim D., Sherry Stephen T., Yankie Linda, et al. GenBank 2023 update. *Nucleic Acids Res* 2023;**51**(D1):D141–4. Doi: 10.1093/nar/gkac1012.

15 Khare Shruti, Gurry Céline, Freitas Lucas, Schultz Mark B, Bach Gunter, Diallo Amadou, et al. GISAID's Role in Pandemic Response. *China CDC Wkly* 2021;**3**(49):1049–51. Doi: 10.46234/ccdcw2021.255.

16 Burki Talha. First shared SARS-CoV-2 genome: GISAID vs virological.org. *Lancet Microbe* 2023;**4**(6):e395. Doi: 10.1016/S2666-5247(23)00133-7.

17 Katoh Kazutaka, Standley Daron M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**(4):772–80. Doi: 10.1093/molbev/mst010.

18 Sobhy Haitham, Colson Philippe. Gemi: PCR Primers Prediction from Multiple Alignments. *Comp Funct Genomics* 2012;**2012**:783138. Doi: 10.1155/2012/783138.

19 Quick Joshua, Grubaugh Nathan D., Pullan Steven T., Claro Ingra M., Smith Andrew D., Gangavarapu Karthik, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;**12**(6):1261–76. Doi: 10.1038/nprot.2017.066.

20 Grubaugh Nathan D., Gangavarapu Karthik, Quick Joshua, Matteson Nathaniel L., De Jesus Jaqueline Goes, Main Bradley J., et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019;**20**(1):8. Doi: 10.1186/s13059-018-1618-7.

21 Papa Mze Nasserdine, Kacel Idir, Beye Mamadou, Tola Raphael, Sarr Mariéma, Basco Leonardo, et al. High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory Samples Using the Illumina COVIDSeq Protocol. *Genes* 2023;**14**(3):681. Doi: 10.3390/genes14030681.

22 Kumar Sudhir, Tamura Koichiro, Nei Masatoshi. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Bioinformatics* 1994;**10**(2):189–91. Doi: 10.1093/bioinformatics/10.2.189.

23 Letunic Ivica, Bork Peer. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;**23**(1):127–8. Doi: 10.1093/bioinformatics/btl529.

24 Aksamentov Ivan, Roemer Cornelius, Hodcroft Emma B., Neher Richard A. Nextclade: clade assignment, mutation calling and quality control for viral genomes 2021. Doi: 10.5281/zenodo.5607694.

25 Colson Philippe, Chaudet Hervé, Delerce Jérémy, Pontarotti Pierre, Levasseur Anthony, Fantini Jacques, et al. Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic. *J Infect* 2024;**88**(5):106150. Doi: 10.1016/j.jinf.2024.106150.

26 Berman Helen M., Westbrook John, Feng Zukang, Gilliland Gary, Bhat T. N., Weissig Helge, et al. The Protein Data Bank. *Nucleic Acids Res* 2000;**28**(1):235–42. Doi: 10.1093/nar/28.1.235.

27 Fantini Jacques, Yahi Nouara, Azzaz Fodil, Chahinian Henri. Structural dynamics of SARS-CoV-2 variants: A health monitoring strategy for anticipating Covid-19 outbreaks. *J Infect* 2021;**83**(2):197–206. Doi: 10.1016/j.jinf.2021.06.001.

28 Azzaz Fodil, Hilaire Didier, Fantini Jacques. Structural basis of botulinum neurotoxin serotype A1 binding to human SV2A or SV2C receptors. *Chem Biol Interact* 2023;**373**:110384. Doi: 10.1016/j.cbi.2023.110384.

29 Johansson Maria U., Zoete Vincent, Michielin Olivier, Guex Nicolas. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* 2012;**13**(1):173. Doi: 10.1186/1471-2105-13-173.

30 Colson Philippe, Fournier Pierre-Edouard, Delerce Jeremy, Million Matthieu, Bedotto Marielle, Houhamdi Linda, et al. Culture and identification of a "Deltamicron" SARS-CoV-2 in a three cases cluster in southern France. *J Med Virol* 2022;**94**(8):3739–49. Doi: 10.1002/jmv.27789.

31 Colson Philippe, La Scola Bernard, Beye Mamadou, Delerce Jérémy, Raoult Didier, Fantini Jacques. Emergence of a second SARS-CoV-2 variant with a tremendous genetic leap from its ancestors. *J Med Virol* 2023;**95**(10):e29124. Doi: 10.1002/jmv.29124.

32 Fantini Jacques, Carlus Denis, Yahi Nouara. The fusogenic tilted peptide (67-78) of α-synuclein is a cholesterol binding domain. *Biochim Biophys Acta* 2011;**1808**(10):2343–51. Doi: 10.1016/j.bbamem.2011.06.017.

33 V'kovski Philip, Kratzel Annika, Steiner Silvio, Stalder Hanspeter, Thiel Volker. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 2021;**19**(3):155–70. Doi: 10.1038/s41579-020-00468-6.

34 Ou Xiuyuan, Guan Hongxin, Qin Bo, Mu Zhixia, Wojdyla Justyna A., Wang Meitian, et al. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat Commun* 2017;**8**(1):15216. Doi: 10.1038/ncomms15216.

35 McCallum Matthew, Park Young-Jun, Stewart Cameron, Sprouse Kaitlin R., Brown Jack, Tortorici M. Alejandra, et al. Human coronavirus HKU1 recognition of the TMPRSS2 host receptor. *BioRxiv Prepr Serv Biol* 2024:2024.01.09.574565. Doi: 10.1101/2024.01.09.574565.

36 Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019 | NEJM. Available at https://www.nejm.org/doi/full/10.1056/nejmoa2001017. Accessed December 12, 2022, 2020.

37 Li Qian, Shah Taif, Wang Binghui, Qu Linyu, Wang Rui, Hou Yutong, et al. Cross-species transmission, evolution and zoonotic potential of coronaviruses. *Front Cell Infect Microbiol* 2023;**12**. Doi: 10.3389/fcimb.2022.1081370.

38 Giraud-Gatineau, C Boschi, C Devaux, Jp Casalta, P Gautret, H Chaudet, et al. Control of common viral epidemics but not of SARS-CoV-2 through the application of hygiene and distancing measures. *J Clin Virol Off Publ Pan Am Soc Clin Virol* 2022;**150–151**. Doi: 10.1016/j.jcv.2022.105163.

39 Sloots Theo P., McErlean Peter, Speicher David J., Arden Katherine E., Nissen Michael D., Mackay Ian M. Evidence of human coronavirus HKU1 and human bocavirus in Australian children. *J Clin Virol Off Publ Pan Am Soc Clin Virol* 2006;**35**(1):99–102. Doi: 10.1016/j.jcv.2005.09.008.

40 Chen Xiangpeng, Zhu Yun, Li Qi, Lu Gen, Li Changchong, Jin Rong, et al. Genetic characteristics of human coronavirus HKU1 in mainland China during 2018. *Arch Virol* 2022;**167**(11):2173–80. Doi: 10.1007/s00705-022-05541-4.

41 Forni Diego, Cagliani Rachele, Clerici Mario, Sironi Manuela. Molecular Evolution of Human Coronavirus Genomes. *Trends Microbiol* 2017;**25**(1):35–48. Doi: 10.1016/j.tim.2016.09.001.

42 Dominguez Samuel R., Shrivastava Susmita, Berglund Andrew, Qian Zhaohui, Góes Luiz Gustavo Bentim, Halpin Rebecca A., et al. Isolation, propagation, genome analysis and epidemiology of HKU1 betacoronaviruses. *J Gen Virol* 2014;**95**(4):836–48. Doi: 10.1099/vir.0.059832-0.

43 Shao Nan, Zhang Chi, Dong Jie, Sun Lilian, Chen Xiangpeng, Xie Zhengde, et al. Molecular evolution of human coronavirus-NL63, -229E, -HKU1 and -OC43 in hospitalized children in China. *Front Microbiol* 2022;**13**:1023847. Doi: 10.3389/fmicb.2022.1023847.

44 Fantini Jacques, Chahinian Henri, Yahi Nouara. Convergent Evolution Dynamics of SARS-CoV-2 and HIV Surface Envelope Glycoproteins Driven by Host Cell Surface

Receptors and Lipid Rafts: Lessons for the Future. *Int J Mol Sci* 2023;**24**(3):1923. Doi: 10.3390/ijms24031923.

45 Pronker Matti F., Creutznacher Robert, Drulyte Ieva, Hulswit Ruben J. G., Li Zeshi, van Kuppeveld Frank J. M., et al. Sialoglycan binding triggers spike opening in a human coronavirus. *Nature* 2023;**624**(7990):201–6. Doi: 10.1038/s41586-023-06599-z.

46 Hammache D., Yahi N., Maresca M., Piéroni G., Fantini J. Human erythrocyte glycosphingolipids as alternative cofactors for human immunodeficiency virus type 1 (HIV-1) entry: evidence for CD4-induced interactions between HIV-1 gp120 and reconstituted membrane microdomains of glycosphingolipids (Gb3 and GM3). *J Virol* 1999;**73**(6):5244–8. Doi: 10.1128/JVI.73.6.5244-5248.1999.

47 Matveeva Mélanie, Lefebvre Marine, Chahinian Henri, Yahi Nouara, Fantini Jacques. Host Membranes as Drivers of Virus Evolution. *Viruses* 2023;**15**(9):1854. Doi: 10.3390/v15091854.

FIGURE LEGENDS

Figure 1: Temporal distribution of HCoV-HKU1 samples used for next-generation sequencing between 2017 and 2022

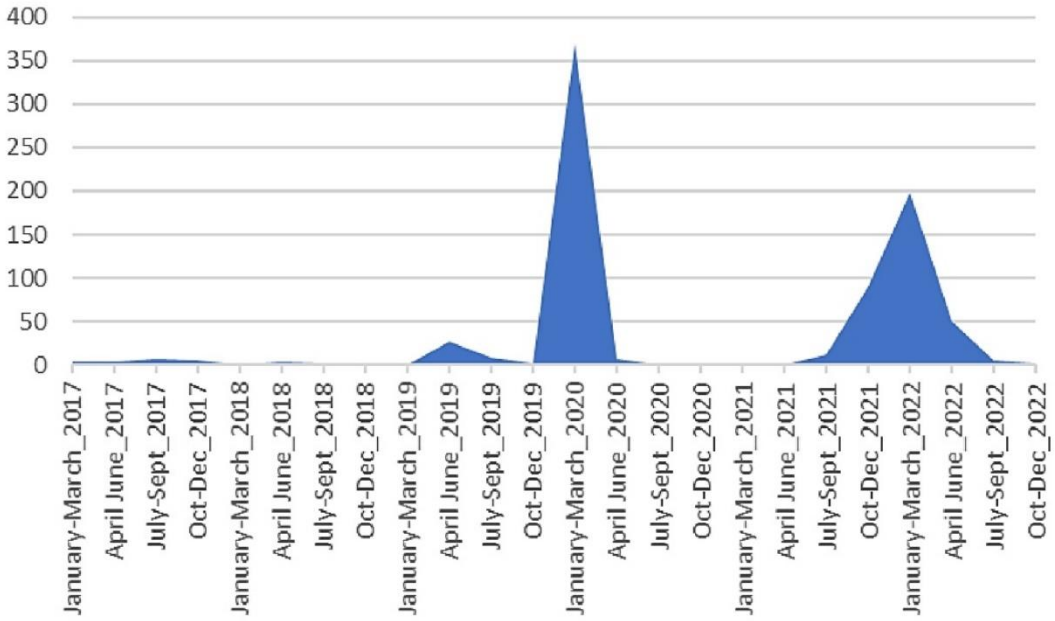
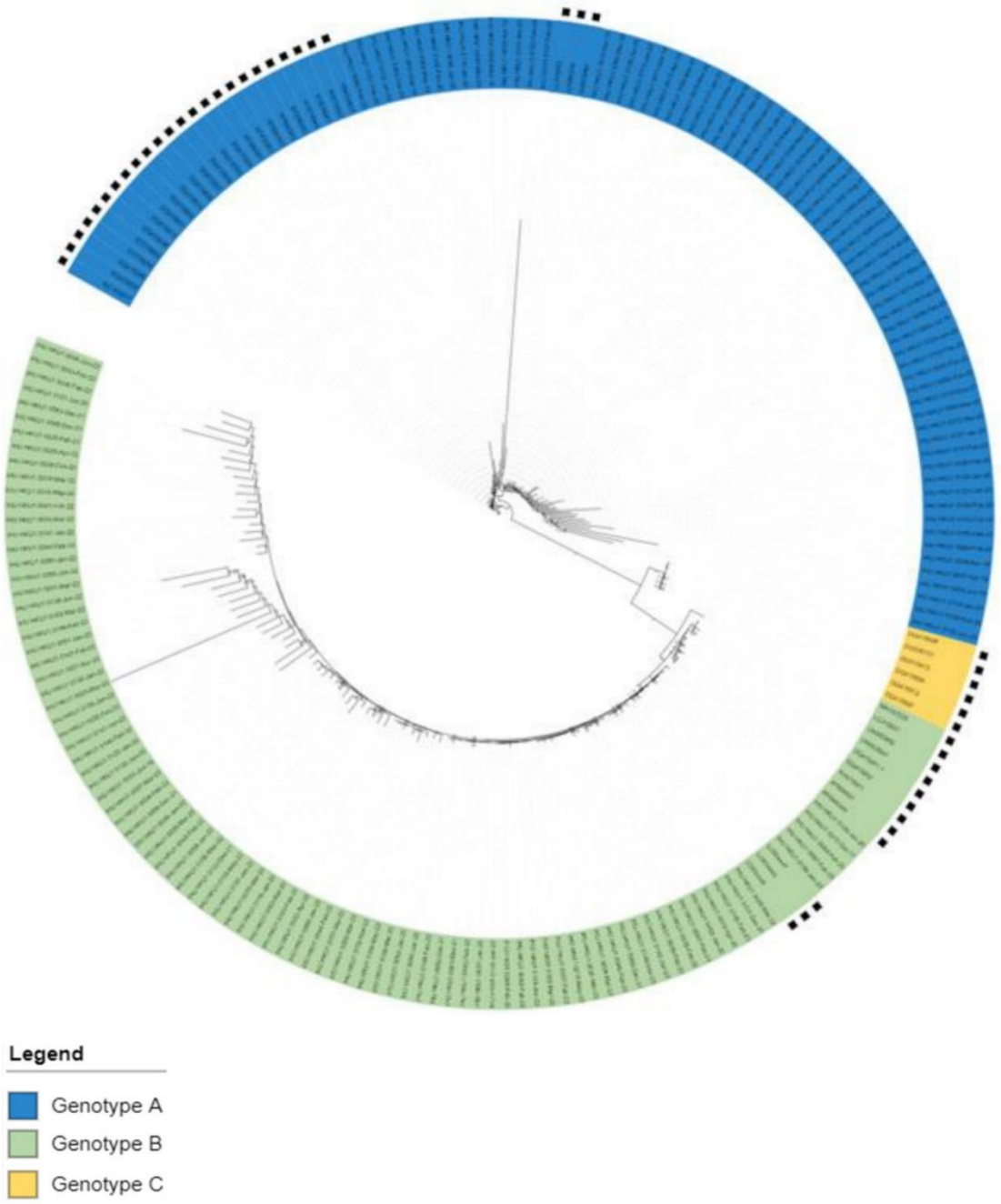
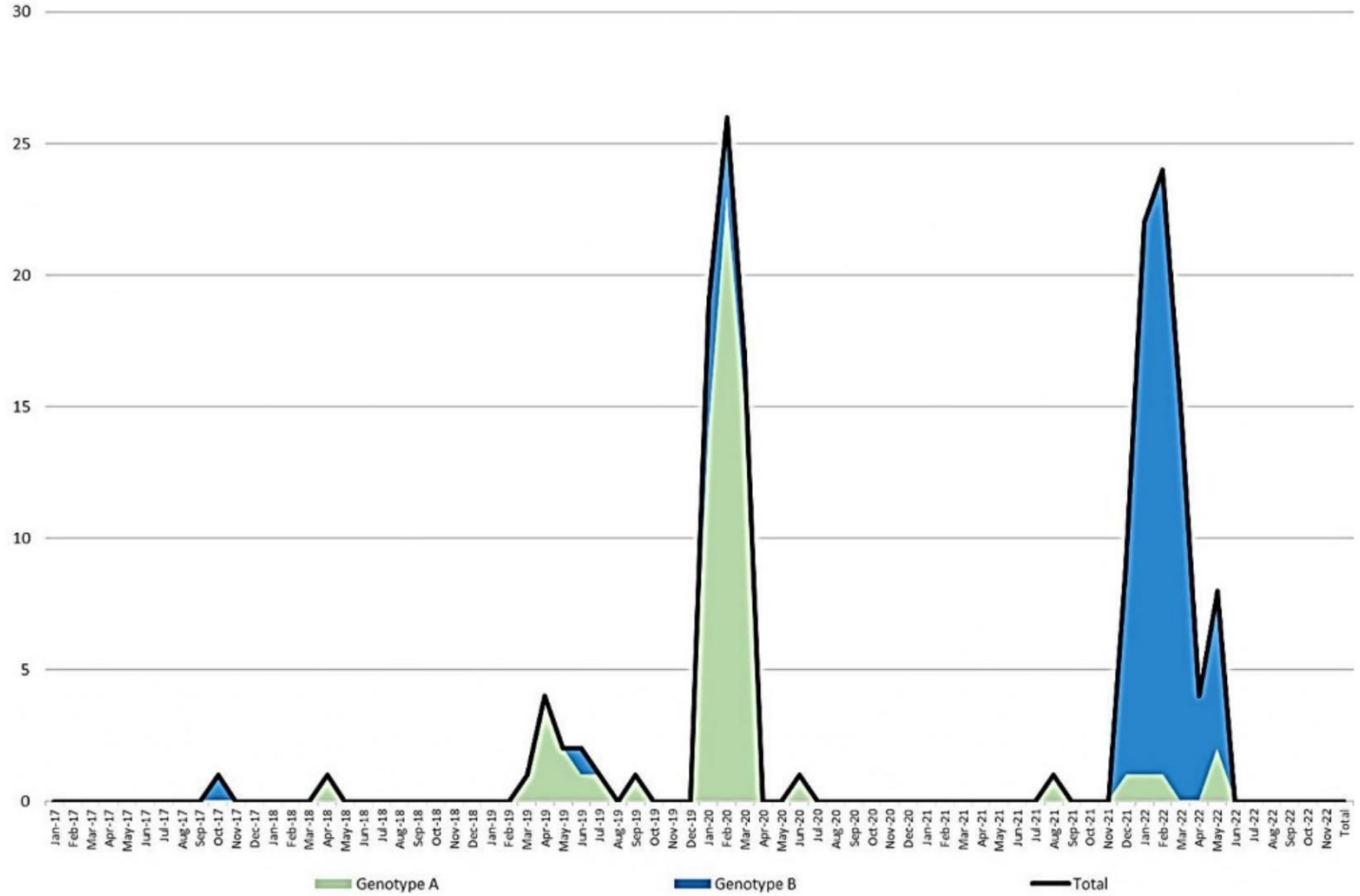


Figure 2: Phylogenetic tree based on HCoV-HKU1 genomes



Asterisks indicate genomes recovered from GenBank. All other genomes were obtained in the present study.

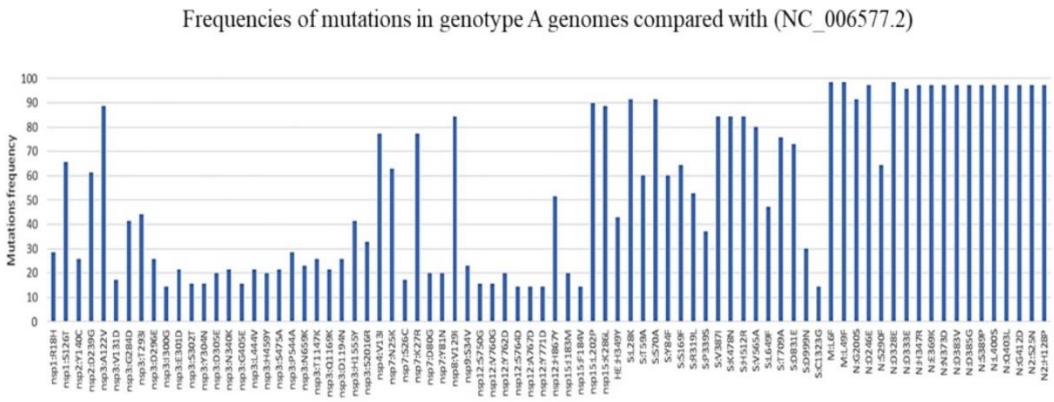
758 **Figure 3:** Distribution of genotypes over time for HCoV-HKU1 genomes obtained in the present study



759

Figure 4: Frequencies of mutations in genotype A and genotype B genomes compared with references

A.



B.

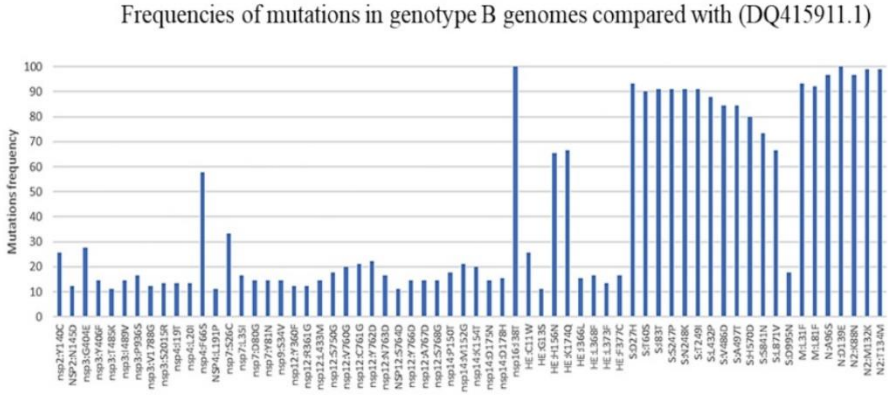
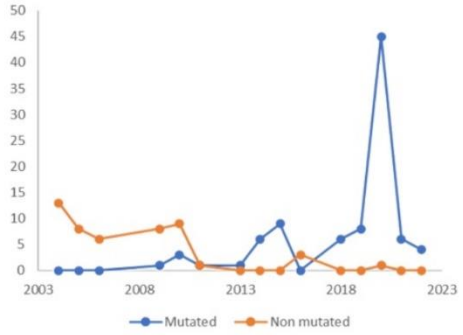
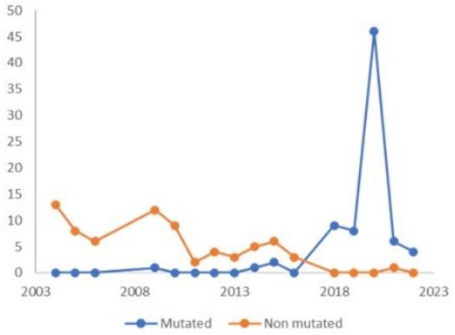


Figure 5: Prevalence of spike H512R, V387I and K487N mutations over time

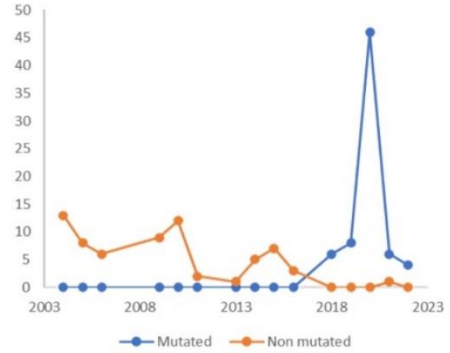
A. Prevalence of spike H512R mutation over time



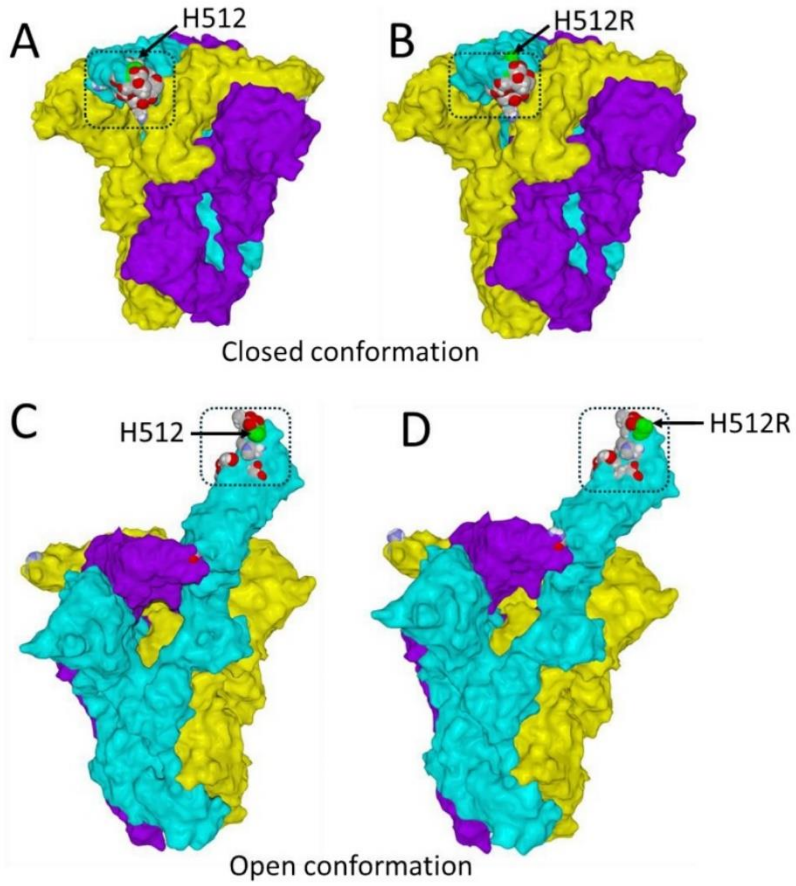
B. Prevalence of spike V387I mutation over time



C. Prevalence of spike K487N mutation over time



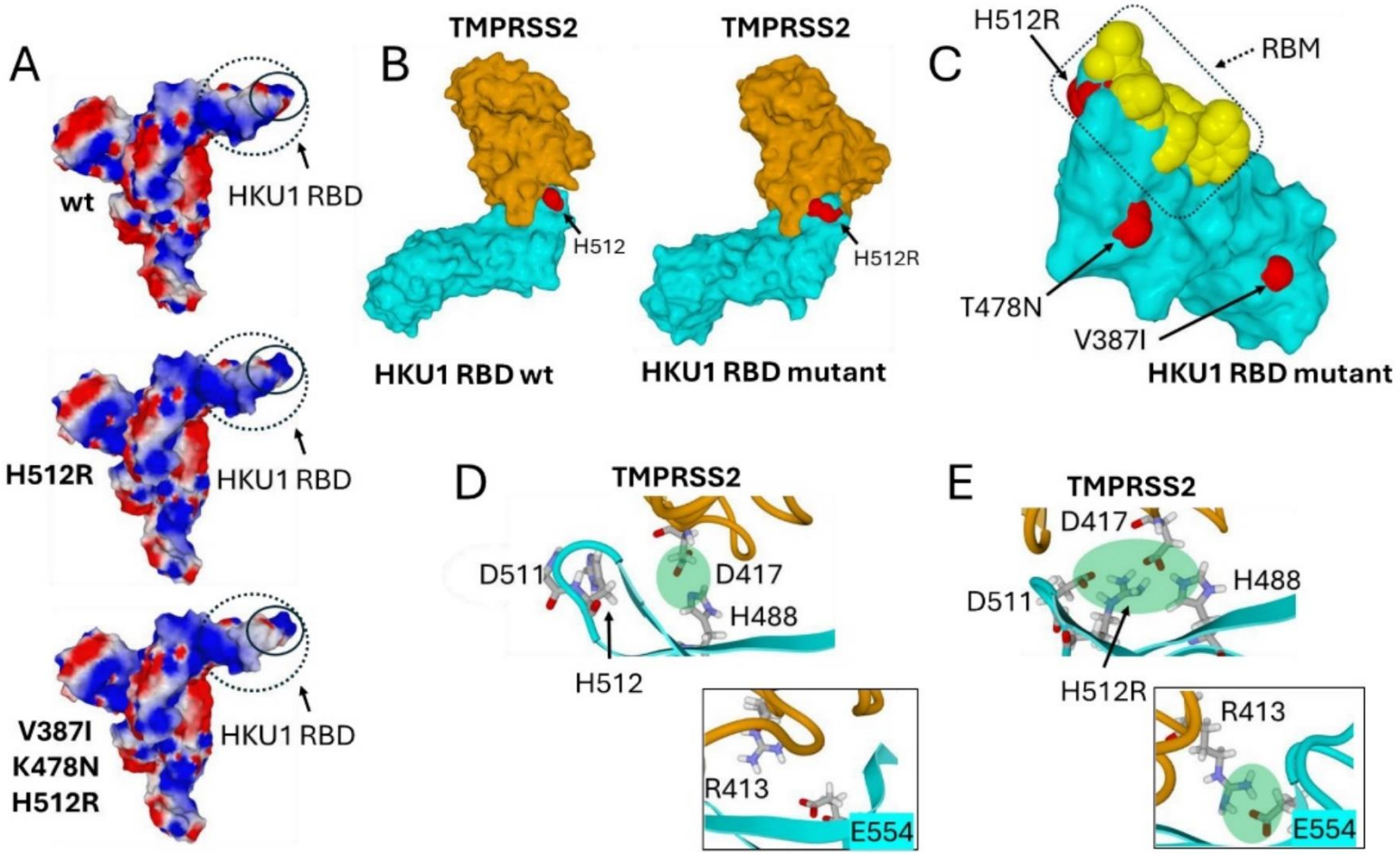
775 **Figure 6:** Location of H512R mutation in the trimeric spike of HKU1 in the closed and open
776 conformation.



777

778 **A.** Wild type trimer with H512 (green atomic spheres) in the closed conformation. The three
779 subunits are colored in cyan (chain A), yellow (chain B) and purple (chain C). Amino acid
780 residues belonging to the receptor binding motif (RBM, rectangle in dotted lines) are shown
781 in atomic spheres (carbon in green, oxygen in red, nitrogen in blue and hydrogen in grey). **B.**
782 H512R mutant in the closed conformation, same legend as in panel A. **C.** Wild type trimer in
783 the open conformation. **D.** H512R mutant in the open conformation. Note that in panels C and
784 D, the RBM is totally accessible to TMPRSS2. The structures of the wild-type HKU1 trimer
785 in the closed and open conformations were retrieved from pdb 8ohn and 8opn, respectively.
786 Mutation H512R was introduced in these files and submitted to molecular modeling.

Figure 7: Impact of mutation H512R, K478N, and V387I on HKU1 spike in the open conformation



790
791 **A.** Electrostatic surface potential of wild-type, single mutant H512R and triple mutant H512R, K478N, and V387I calculated in spike monomers.
792 These three mutations are all located within the receptor binding domain (RBD, dotted circle). Electronegative areas are in red, electropositive in
793 blue and neutral in white. Note the significant increase of electropositive surfaces in the RBM (solid circle) of the single and triple mutant spikes.
794 **B.** TMPRSS2 bound to wild-type HKU1 RBM (left panel) or triple mutant (right panel). TMPRSS2 is colored in orange, the HKU1 RBD in
795 cyan. Amino acids H512 (wild-type) and H512R (mutant) are represented in red atomic spheres. **C.** Focus on the amino acid residues that
796 constitute the RBM of HKU1 (dotted rectangle, amino acids in atomic spheres) in the RBD (colored in cyan). Mutant amino acids are
797 represented in red atomic spheres. Note that only H512R is in close contact with the RBM; T478N and V387I are not directly linked to the RBM.
798 **D.** TMPRSS2 (orange ribbons) bound to wild-type HKU1 RBM (cyan ribbons). Note that H488 (HKU1) interacts with D417 (TMPRSS2). The
799 binding zone is indicated by a green disk. Inset: lack of interaction between E554 (HKU1) and R413 (TMPRSS2). The atomic coordinates were
800 retrieved from pdb 8vgt. **E.** Molecular dynamics simulations (MDS) of TMPRSS2 binding to the triple mutant of HKU1 (open conformation).
801 Inset: during the binding process, an electrostatic bond is established between E554 (HKU1) and R413 (TMPRSS2). H512R is predicted to form
802 a high energy complex (green disk) involving D511 and H488 (HKU1) and D417 (TMPRSS2). This minimizes the initial electrostatic repulsion
803 between D511 and D417.
804

TABLES

Table1: Proportions of spike H512R, V387I and K487N mutations over time

	Year										
	<2009	2009	2010-2016	2017	2018	2019	2020	2021	2022	2023	**Any year**
Sequences harboring the mutation / covering the position (N (%)))											
H512R											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58/58**
GenBank	0/27	1/9	20/33	-	5/5	1/1	0/1	4/4	1/1	-	**32/81**
Total	**0**	**1**	**20 (60%)**	**-**	**6**	**8**	**45**	**6**	**4**	**-**	**90/139**
V387I											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58/58**
GenBank	0/27	1/13	3/35 (8%)	-	8/8	1/1	1/1	4/5	1/1	-	**19/91**
Total	**0**	**1**	**3 (8%)**	**-**	**9**	**8**	**46**	**6**	**4**	**-**	**77**
K478N											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58**
GenBank	0/27	0/9	0/30 (0%)	-	5/5	1/1	1/1	4/5	1/1	-	**12/79**
Total	**0**	**0**	**0 (0%)**	-	**6**	**8**	**46**	**6**	**4**	-	**70/137**
H512R + V387I											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58**
GenBank	0/27	1/9	3/29	-	5/5	1/1	0/1	4/4	1/1	-	**15**
Total	**0**	**1**	**3 (10%)**	**-**	**6**	**8**	**45**	**6**	**4**	**-**	**73/135**
H512R + K478N											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58**
GenBank	0/27	0/9	0/30 (0%)	-	5/5	1/1	0/1	4/4	1/1	-	**11/78**
Total	**0**	**0**	**0 (0%)**	**-**	**6**	**8**	**45**	**6**	**4**	**-**	**69/136**
H512R + V387I + K478N											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58**
GenBank	0/27	0/9	0/27 (0%)	-	5/5	1/1	0/1	4/4	1/1	-	**11/75**
Total	**0**	**0**	**0 (0%)**	-	**6**	**8**	**45**	**6**	**4**	-	**69/133**
V387I + K478N											
Present study	-	-	-	-	1/1	7/7	45/45	2/2	3/3	-	**58**
GenBank	**0/27**	**0/9**	**0/27 (0%)**	-	5/5	1/1	1/1	4/5	1/1	-	**12/76**
Total	**0**	**0**	**0 (0%)**	**-**	**6**	**8**	**46**	**6**	**4**	**-**	**70/134**

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

We used the same methodology than in a previous study [1] to classify genes into "structural", "informational", "other non-structural" and "accessory" genes. In the structural gene category, we included the genes encoding the hemagglutinin esterase and the N2 part of the nucleocapsid, which are absent in SARS-CoV-2 but are present in HCoV-HKU1, and in the accessory gene category we only included the ORF4 gene. In order to analyze three particular mutations in the spike, we retrieved the full-length spike amino acid sequences from the NCBI GenBank protein sequence database using "HCoV-HKU1" and "Spike" as keywords, and analyzed them concurrently with those obtained in the present study.

SUPPLEMENTARY RESULTS

HCoV-HKU1 mutational patterns

Genotype A sequences

Compared to reference genome LC315650.2 dating back to 2014, a total of 5,671 nucleotide substitutions, 417 deletions and 855 insertions were observed in the genomes obtained here. This corresponds to a substitution rate of $1.3\times10^{-3}\pm1.2\times10^{-3}$ per nucleotide per year. The gene that showed the greatest number of nucleotide substitutions relatively to its length was nsp9, which encodes a single-stranded DNA/RNA-binding protein [2], followed by nsp4, which was involved in deflecting host endomembranes into replication organelles [3], with 394 and 392 substitutions per 1,000 nucleotides, respectively; both these genes are part of the category of non-structural genes. In this category, the gene that showed the lowest number of nucleotide substitutions was nsp1, which encodes a host shutoff protein that controls anti-viral responses [4], with 112 substitutions per 1,000 nucleotides. Among

informational genes, as relatively to the 2004 reference, nsp12 was also the gene that exhibited the greatest number of nucleotide substitutions (339 substitutions per 1,000 nucleotides), and nsp15 exhibited the lowest number of nucleotide substitutions in this category (48 per 1,000 nucleotides). The HE encoding gene and the spike gene were the structural genes with the greatest number of mutations with 302 and 182 substitutions per 1,000 nucleotides, respectively, while the envelope gene carried the lowest number of nucleotides substitutions (12 per 1,000 nucleotides) (Supplementary Table S3). The accessory genes carried the same prevalence of susbstitutions (57 per 1,000 nucleotides) as relatively to the reference genome dating back to 2004.

Regarding amino acid mutations relatively to the reference genome LC315650.2 dating back to 2014, 1,629 amino acid substitutions were found. This corresponds to a substitution rate of $3.2\times10^{-4}\pm2.6\times10^{-4}$ per nucleotide per year. The number of amino acid substitutions ranged from 14 to 319 with a mean value of 55.0±46.1; the number of deletions ranged from 0 to 38 with a mean value of 9.0±8.7; and the number of insertions ranged from 1 to 42 with a mean value of 4.7±6.0. A total of 1,629 amino acid substitutions were detected. Among informational proteins, nsp12 was the protein with the greatest diversity with 299 amino acid substitutions per 1,000 amino acids; nsp15 exhibited the lowest diversity with 42 substitutions per 1,000 amino acids. Among structural proteins, HE followed by spike showed the greatest diversity with 290 and 144 substitutions per 1,000 amino acids, respectively, and envelope showed the lowest diversity (Supplementary Table S3). Of these 1,629 amino acid substitutions, only 41 were present in at least 10 genomes (Supplementary Figure S1A). The spike was the protein harboring the greatest number of substitutions. The proteins with no mutation present in at least 10 genomes were nsp4, nsp5, nsp6, nsp10, nsp11, nsp13, M, N2, E and ORF4. Interestingly, two stop codons (N193* in nsp4, and C1322* in the spike gene) were present in 11 and 10 of the genomes recovered here, respectively.

Genotype B sequences

A total of 6,149 nucleotide substitutions, 428 deletions and 789 insertions were detected in the genotype B sequences obtained in this study relatively to the reference genome LC315651.2 dating back to 2016. This corresponds to a substitution rate of $1.5\times10^{-3}\pm1.2\times10^{-3}$ per nucleotide per year. We observed results similar to those comparing these genomes to the older (dating back to 2005) genotype B reference genome (Supplementary Table S5). In comparison with this reference, genomes obtained here encoded a total of 1,789 amino acid susbtitutions. This corresponds to a substitution rate of $3.5\times10^{-4}\pm2.6\times10^{-4}$ per nucleotide per year. The number of amino acid substitutions ranged from 18 to 243 with a mean value of 56.7±44.6. The number of deletions was between 1 and 47 with an average of 9.2±9.5 and the number of insertions was between 0 and 35 with a mean value of 2.7±5.1. Of them, 61 were only present in at least 10 genomes. The spike was the protein with the greatest number of mutations. The proteins that harbored no amino acid substitution present in at least 10 genomes were nsp1, nsp5, nsp6, nsp8, nsp10, nsp11, nsp13, nsp15, nsp16, and ORF4 (Supplementary Figure S1B). Finally, in these genotype B genomes, three stop codons were detected, including L6* in the HE, F171* in nsp15, and N193* in nsp4, in 19, 14 and 10 of these genomes, respectively.

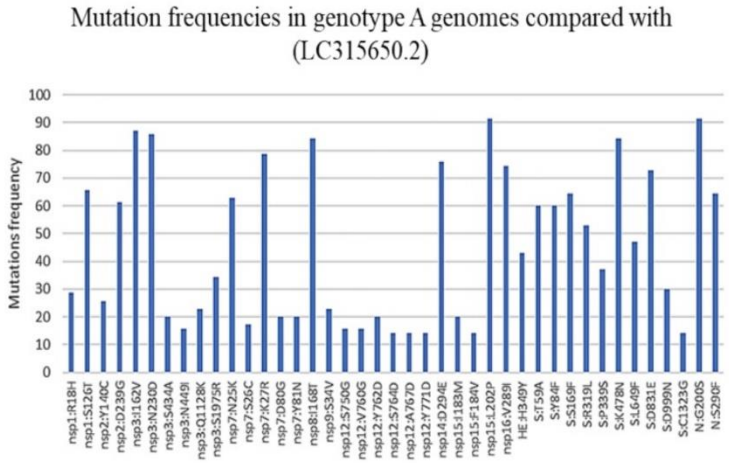
879 **SUPPLEMENTARY FIGURES**

880

881 **Supplementary Figure S1:** Frequencies of mutations in genotype A and genotype B

882 genomes compared with references

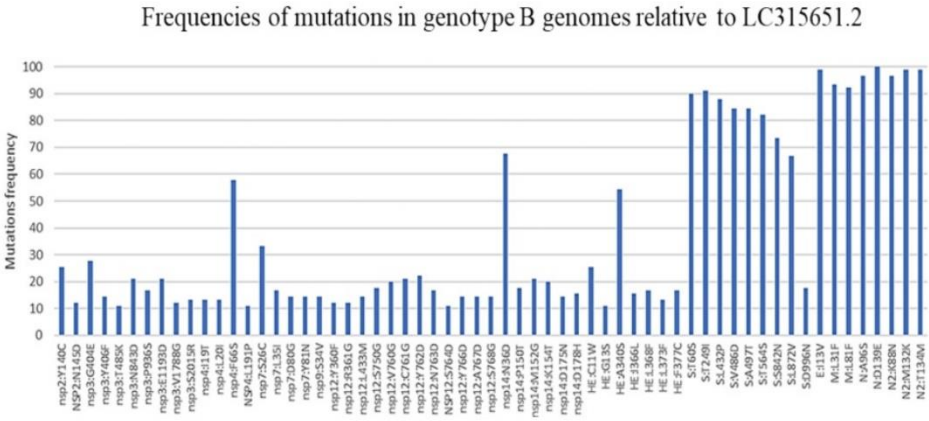
883 **A.**



884

885

886 **B.**



887

888

889 **SUPPLEMENTARY TABLES**

890

891 **Supplementary Table S1.** Sequences of primers used for PCR amplifcation prior to the next-

892 generation sequencing of the HCoV-HKU1 genomes

Primer name	Sequence (5'-3' orientation)	Pool	Concentration of use (pmol/µL)
HKU1_F1	ACGATCTCTTGTCAGATCYC	1	15.0
HKU1_R1	CAATCTACATAACCATACTGCA		15.0
HKU1_F2	GTGGTTTTGTTATGCAGTATG	2	15.0
HKU1_R2	GGTTGACAATACTCATAAGATG		15.0
HKU1_F3	TTTATTTTTTTGAAGGCGTTGCA	1	15.0
HKU1_R3	GTCACCAGTAACAACATCTTC		15.0
HKU1_F4_ALT	ACTGGTGACAATGACGATGA	2	12.6
HKU1_R4	CATACARTTGAGSTAACTCAAA		12.6
HKU1_F5	CTGTGGATGTAAACGTTTGTC	1	10.0
HKU1_R5	GCRAGCAATTGTTGCTGATC		10.0
HKU1_F6	GGTAATGTTTTCTGTGATGGC	2	10.0
HKU1_R6	CAACCCAAACAACATAATCACA		10.0
HKU1_F7	GTATCTGAGGAGTCACAAGG	1	10.0
HKU1_R7	AGGTCGTTTAAGCTCTTTAGA		10.0
HKU1_F8	TCTAAAGAGCTTAAACGACCT	2	10.0
HKU1_R8	GTGGAACYAAAGATGTATACAA		10.0
HKU1_F9	AAGAGGTGATGGTACACCAC	1	10.0
HKU1_R9	CATAACCAACAGATCCACATG		10.0
HKU1_F10	GCATTTCATGTTACTATGCGTA	2	15.0
HKU1_R10	CTACAATACTGCCACAACYTA		15.0
HKU1_F11_ALT	GAATAGGTGGCGTGCCAGT	1	12.6
HKU1_R11_ALT	ACAGACTTGACAAACATCATGT		12.6
HKU1_F12	CGTGTAGAGCATYCAGATGT	2	10.0
HKU1_R12	AAATCAACTGTACTACCCTCTT		10.0
HKU1_F13	GTAAAGGCTTGTTTAAAGAGG	1	12.6
HKU1_R13	ACTATCTAAGATCTGAKTACCA		12.6
HKU1_F14	GAGCGCTTTGTAAGTCTAGC	2	10.0
HKU1_R14_ALT	CCCAAAAAARATATCAGGACCT		10.0
HKU1_F15	GTGCTRTTGAGYAAGGGTTC	1	12.6
HKU1_R15	GCAACATGTGCACCTTTATGT		12.6
HKU1_F16	TTGTAGATATACAACAGTGGG	2	10.0
HKU1_R16	CATCGTCACCATCTTTTCTCA		10.0
HKU1_F17	GTCCRCAACGAGCTGATTTA	1	10.0
HKU1_R17	ACCRCCATTCCATGTTGTAC		12.6
HKU1_F18	AAATGCTTCTTCTAGTGAAGG	2	15.0
HKU1_R18	AATCAACAACCTGYACRGGTA		15.0
HKU1_F19	TACACTGGCGARGGTGATC	1	10.0
HKU1_R19	CACCAGMTTTAGGRAAATAAC		10.0
HKU1_F20	GCCTACAACAYTAGCTGTTAT	2	15.0
HKU1_R20	CAATTAAGAGGTGARGGTACA		15.0
HKU1_F21	CTGTWAAGCCTGTTGCAACT	1	15.0
HKU1_R21	ATTRGAACAAGTGGTRCCAC		15.0
HKU1_F22	AGTGGYACCACTTGTTCYAAT	2	15.0
HKU1_R22	CACTTTGWATTTTAGCAAGWGC		15.0
HKU1_F23	AATGGTTTGGGTGTTACTATG	1	15.0
HKU1_R23	GACATTTACTAAAACATGCWGA		15.0
HKU1_F24	TGGTTGYGGTTCWGCATGTT	2	15.0
HKU1_R24	GGCCTAACAWACATCTTRCC		15.0
HKU1_F25	GAACTGGMAGTTGGTGGAG	1	10.0
HKU1_R25	TCARATTGGTCARCCCAAGAA		10.0
HKU1_F26	GTCATYATGCTGGAAGTAGAA	2	10.0
HKU1_R26	TGAGAACGTGAACCTGGYC		10.0
HKU1_F27	TACGATTYTGCCTCAAGGCT	1	10.0
HKU1_R27	CTTTCATAATTGTCTCAAARCC		10.0
HKU1_F28	TACTTTACCWGGYTTTGAGAC	2	15.0
HKU1_R28	CAGCTAAACCTTATAACTACTC		15.0

893 Degenerated nucleotide positions are indicated by a red font.

Supplementary Table S5. Frequency of nucleotide and amino acid substitutions per 1000 nucleotides compared with the genotype B reference genome (LC315651.2)

Genotype	Gene	Nucleotide size	Number of mutations per gene	Number of mutations per 1,000 nucleotides
Genotype B (LC315651.2)	Nsp1	666	36	54
	Nsp2	1761	216	122
	Nsp3	6084	1009	166
	Nsp4	1488	443	297
	Nsp5	909	126	138
	Nsp6	861	66	76
	Nsp7	276	53	192
	Nsp8	582	95	163
	Nsp9	330	140	424
	Nsp10	411	79	192
	Nsp11	42	0	0
	Nsp12	2783	808	290
	Nsp13	1809	429	237
	Nsp14	1563	420	268
	Nsp15	1122	116	103
	Nsp16	897	168	187
	HE	1158	549	474
	S	4059	860	212
	ORF4	330	53	160
	E	249	42	168
	M	672	78	116
	N	1326	101	76
	N2	618	50	81

Genotype	Gene	Protein size	Number of mutations per gene	Number of mutations per 1,000 nucleotides
Genotype B (LC315651.2)	Nsp1	222	18	81
	Nsp2	587	95	162
	Nsp3	2029	342	168
	Nsp4	496	123	248
	Nsp5	303	39	128
	Nsp6	287	28	97
	Nsp7	92	30	326
	Nsp8	194	28	144
	Nsp9	110	57	518
	Nsp10	137	21	153
	Nsp11	14	0	0
	Nsp12	928	224	241
	Nsp13	603	103	171
	Nsp14	521	159	305
	Nsp15	374	44	117
	Nsp16	299	50	167
	HE	385	145	376
	S	1352	137	101
	ORF4	109	20	183
	E	82	24	292
	M	223	22	98
	N	441	42	95
	N2	205	38	185

REFERENCES

1 Colson Philippe, Chaudet Hervé, Delerce Jérémy, Pontarotti Pierre, Levasseur Anthony, Fantini Jacques, et al. Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic. *J Infect* 2024;**88**(5):106150. Doi: 10.1016/j.jinf.2024.106150.

2 V'kovski Philip, Kratzel Annika, Steiner Silvio, Stalder Hanspeter, Thiel Volker. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* 2021;**19**(3):155–70. Doi: 10.1038/s41579-020-00468-6.

3 Liu Ding X., Liang Jia Q., Fung To S. Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae). In: Bamford Dennis H., Zuckerman Mark, editors. *Encycl. Virol. Fourth Ed.* Oxford: Academic Press; 2021. pp. 428–40.

4 Bermudez Yahaira, Miles Jacob, Muller Mandy. Nonstructural protein 1 widespread RNA decay phenotype varies among coronaviruses. *iScience* 2023;**26**(1). Doi: 10.1016/j.isci.2022.105887.

Article n°9 :
Five-year (2017-2022) evolutionary dynamics of human coronavirus OC43 in southern France based on whole genome next-generation sequencing

Houmadi Hikmat, Lorlane Le Targa, Celine Boschi, Justine Py, Sarah Aherfi, Bernard La Scola, Philippe Colson (en cours d'écriture)

TITLE PAGE

Target journal and type of article: Research article

Full-length title: Five-year (2017-2022) evolutionary dynamics of human coronavirus OC43 in southern France based on whole genome next-generation sequencing

Short title (for the running head): Human coronavirus OC43 genomics

Author list: Hikmat HOUMADI[1], // Bernard LA SCOLA[1*], Philippe COLSON[1*]

Affiliations: [1] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Assistance Publique-Hôpitaux de Marseille (AP-HM), IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France

*** Corresponding authors:** Bernard La Scola, Philippe Colson IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: bernard.la-scola@univ-amu.fr, philippe.colson@univ-amu.fr

Key words: Human coronavirus OC43; respiratory infections; genomics; mutations; lineages

TEXT

INTRODUCTION

Coronaviruses are viruses that can affect both humans and animals (Millet et al, 2020). Human coronaviruses are represented by seasonal or endemic coronaviruses, including HCoV-OC43, HKU1, NL63 and 229E, as well as epidemic coronaviruses including SARS-CoV, MERS-CoV and SARS-CoV-2 (Otieno et al, 2022; Martínez et al, 2021). Seasonal coronaviruses are responsible for mild respiratory diseases in general. HCoV-OC43 has also been detected in patients with neurological disorders, as some other coronaviruses (Salinas and Simonin 2020). It was one the first human coronaviruses to be discovered, in England during the 1960s, being first isolated from an organic tissue culture (McIntosh et al, 1967).

HCoV-OC43 belongs to the *Betacoronavirus* genus and its HCoV-OC43 genome, as fror the case of other coronavirus genomes, is around 3 kilobases in size. Two-thirds of this genome encodes non-structural proteins (nsp1-nsp16) and the remaining third encodes structural proteins. The HCoV-OC43 genome encodes a haemagglutinin esterase, as for the case of other betacoronaviruses, and accessory proteins including NS2 and NS12.9 (Zhang et al, 2015). HCoV-OC43 was reported to use 9-O-acetylated sialic acid to bind host cells, but the cellular receptor that plays the role of receptor is not currently known (Saunders et al, 2023; Hulswit et al, 2019).

HCoV-OC43 is the most frequently diagnosed seasonal coronavirus (Gaunt et al, 2010; Lau et al, 2006; Liu et al, 2017; Yip et al, 2016; Shah et al, 2022; Kong et al, 2021; Monto et al, 2020). It is characterised by a winter and spring seasonality in temperate countries. In tropical countries, the seasonality of this virus is not well known (Monto et al, 2020; Liu et al, 2017; Hodinka 2016; Al-Khannaq et al, 2016; Otieno et al, 2022; Yip et al, 2016).

47 Since the advent of next-generation sequencing (NGS), genomic data for microorganisms and
48 viruses and study of their genetic evolution have expanded, but some viruses have been
49 neglected for such approach. Wherase there are currently more than 8 million SARS-CoV-2
50 genomes in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and more than 16 million in
51 Gisaid (https://gisaid.org/) (Burki, 2023), only 361 genomes (28861-31104) of HCoV-OC43
52 were available as of early October 2023 and few studies were carry out on the genomic
53 monitoring of this virus on large sample sets. The aim of the present study was therefore to
54 study retrospectively the genomes and evolutionary dynamics of HCoV-OC43 between 2017
55 and 2022 in south of France using an in-house PCR amplification system pre-NGS.
56
57 **MATERIALS AND METHODS**
58 **Clincal samples**
59 Detection of NGS of HCoV-OC43 genomes was performed on nasopharyngeal
60 samples diagnosted HCoV-OC43 RNA-positive by real-time reverse-transcription (RT)-PCR
61 (qPCR) using the FTD Respiratory pathogens 21 assay (FTD) (Fast Track Diagnosis,
62 Luxembourg) after extraction by the KingFisher Flex system (Thermo Fisher Scientific,
63 Waltham, MA, USA) according to the manufacter's protocol, or by the FilmArray
64 Respiratory panel 2 plus assay (BioFire) (Biomérieux, Marcy-l'Etoile, France). These clinical
65 samples had been sent to our clinical microbiology laboratory at University and Public
66 Hospitals of Marseille, southeastern France, for the purpose of routine diagnosis of respiratory
67 infections in the setting of clinical routine management, and they had been stored at -20°C or
68 -80°C after processing. A total of nasopharyngeal 1,007 samples were diagnosed HCoV-
69 OC43-positive between 02/2017 and 10/2022 (69 months), among which 434 were available
70 in sufficient volume to be applied the NGS procedure.
71 **Next-generation sequencing**

Primer design for PCR amplification of overlaping regions covering the whole viral genome

To design PCR primers, 220 complete HCoV-OC43 genomes were collected from the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and ViPR (https://www.bv-brc.org/view/Virus/10239) databases using as keywords "complete genomes". These genomes were then aligned using MAFFT (Multiple Alignment using Fast Fourier Transform) software (Katoh and Standley, 2013), and the primers were designed using GEMI software (Sobhy and Colson, 2012). This software is an automated tool that enables primers and probes to be designed quickly and easily on the basis of the bases conserved between the previously aligned genomes. GEMI was setted to select several parameters such as amplicon size (500-2,600 nucleotides), amplicon overlap size (at least 20 nucleotides), primer size (17-23 nucleotides), hybridization temperature (58-65°C), and also the possibility of insertions of degenerated nucleotides (maximum, n= 3). The specificity of the primers was then tested using the NCBI BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). After the primers were tested and optimised on HCoV-OC43 cell culture supernatants, two PCR primer pools were designed to ensure of the absence of overlaping amplicons within each primer pool.

PCR amplification of overlaping regions covering the whole genomes

Extarcted RNA was amplified by standard RT-PCR using the SuperScript III One-Step RT-PCR Kit with Platinum Taq High Fidelity (SS-III-HF kit; Invitrogen, Life Technologies) according to the following protocol: 12.5 µL of 2X mix were added to 7.8 µL of pure water, 0.7 µL of enzyme polymerase (SS-III), and 1 µL for each primer pool, and finally 3 µL of RNA were added to this mixture for a final volume of 25 µL. The PCR program used for the thermocycler consisted of an initial RT step at 50°C for 25 min followed by initial denaturation at 95°C for 2 min, and 40 PCR cycles including 15 sec at 95°C, 45 sec at 58°C, and 150 sec at 70°C. Final elongation consisted in 5 min at 70°C. Then, the

amplicons were purified using the NucleoFast 96-well plate (Macherey Nagel ref 743100.50, Hoerdt, France) for an elution volume of 40 µL. Eventually, the two PCR pools were mixted together as follows: 15 µL of pool no. 1 and 25 µL of pool no. 2.

Next-generation sequencing

To test the PCR primers separately and pooled, we performed NGS using the Oxford Nanopore Technology with the Ligation Sequencing Kit (SQK-LSK109) library that was deposited on a SpotON flow cell Mk I, R9.4.1 and sequenced a GridION instrument, following to the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK). Then, when the 'Artic-like' system was deemed functional, we used it on RNA extracts from the studied nasopharyngeal samples and HBoV genomes were sequenced using the Illumina technology on a NovaSeq 6000 instrument, using the COVIDSeq protocol (Illumina Inc., San Diego, CA, USA), but with replacement of the COVID-19 ARTIC PCR primers by the PCR primers designed here that were used according to set up conditions. NGS on the NovaSeq 6000 instrument, the classic Illumina loading procedure in a SP flow cell was used, according to the NovaSeq-XP workflow and a previously described procedure (Papa Mze et al, 2023), with a reading of 2 x 50 nucleotides.

Bioinformatic Analysis

A workflow was developed using the CLC genomics workbench software v.7, consisting of a quality control step by performing a trimming with a minimum of 15 and a maximum of 1,000 nucleotides in the reads. followed by a mapping step of NGS reads against a reference genome using as a parameters a coverage of at least 80% and a nucleotide identity of at least 90%. Then, the consensus sequences were recovered with 2, 3 and 3 as numbers of mismatches, deletions and insertions, respectively, and ambiguity symbols "N" at genome positions in case of an ambiguity. Once we had obtained our consensus sequences, we used the Mafft software (Katoh et al, 2013) with default parameters to perform the alignments.

Phylogenetic trees were constructed using Iqtree2 (Minh et al, 2020) using HCoV-OC43 genomes from the different genotypes with the maximum likelihood method with 1,000 boostrap replication, and iTOL software (Letunic and Bork, 2007) was used for tree visualisation and annotation. To search for mutations, the Nextclade and Nextrain tools (https://nextstrain.org) (Aksamentov et al, 2021) were modified to be used for this viral species. Amino acid and nucleotide diversities in studied viral genomes were determined according to the genotype to which they belong. For each of the three genotypes, one reference of genotype G (MN026164.1 dating back to 2018), one of genotype J (OK318939.1 also dating back to 2018) and one of genotype K (MW532118 dating back to 2017) were used. The same methodology used in a previous study by (Colson et al, 2023) to classify genes into "structural", "informational", "other non-structural" and "accessory" for SARS-CoV-2 was used in the present study. The hemagglutinin esterase (HE) encoding gene and the N2 part of the nucleocapsid encoding gene, which are part of the HCoV-OC43 genomes but absent in SARS-CoV-2, were added in the structural gene category, while accessory genes included NS2 and NS12.9. Finally, to search for recombinants of HCoV-OC43 genomes, we used the RDP4 tool (Martin et al, 2015), which contains several types of analyses to detect recombination events.

RESULTS

Primer design for PCR amplification of overlaping regions covering the whole genomes

From a total of 220 HCoV-OC43 genomes collected from GenBank according to the keywords used, 165 were selected for PCR primer design. A total of 200 PCR primer pairs were obtained by GEMI and 34 were eventually retained. These PCR primers had a maximum of 2 degenerated nucleotides, their size varied between 17 and 23 nucleotides and the expected sizes of the amplicon varied between 500 and 2,600 base pairs. The hybridization

147 temperature varied between 58°C and 65°C and the amplicons overlapped by at least 20 base
148 pairs. These 34 primers were tested individually and their functionality was verified on a
149 1.5% agarose gel by a band presence. PCR optimisation was carried out by increasing the
150 primer concentration in case the agarose gel band was weak. Eventually, the 34 PCR primers
151 were mixed into two pools of 17 primer pairs each in order to avoid overlaps between the
152 PCR amplicons, with primer concentrations ranging between 10 and 15.3 µM (Table 1).

153 **Next-generation sequencing**

154 Among the genomes obatined from the 434 HCoV-OC43-positive nasopharyngeal
155 samples sequenced using our in-house system on a NovaSeq 6000 instrument, 185 were
156 analysable according to our chosen parameters with a minimum of 80% coverage and of 3 X
157 as depth relatively to the reference (Supplementary Table S1). The 185 nasopharyngeal
158 samples from which they were obtained had been collected between 06/02/2017 and
159 16/07/2022 (over a period of 66 months). The qPCR cycle threshold (Ct) was available for 96
160 of these 185 samples and ranged between 11 and 25 with a mean (± standard deviation) of
161 17.8 ± 3.3. The total number of reads per sample mapped onto the reference genome for the
162 185 genomes ranged between 5,675 and 9,917,033, with a mean of 3,569,789±2,943,862
163 reads, and mean coverage was of 94.7 ± 5.2%.

164 **Phylogenic construction and distribution of genotypes as a function of time**

165 Phylogeny reconstruction included HCoV-OC43 genome sequences obtained here as
166 well as genomes of the different genotypes based on previous datasets Ye et al, 2023). Only
167 boostrap values greater than or equal to 70% were displayed on the tree. The viral genome
168 recovered here were clustered into three genotypes, namely G, J and K, that contained each
169 17, 79 and 89 sequences, respectively (Figure 1). The mean nucleotide diversity of our
170 sequences was 91.88 sequences ± 6.27 for genotype G, 93.17 ± 5.05 for genotype J
171 sequences, and 91.77 ± 5.32 for genotype K sequences. We then looked at the distribution of

these different genotypes as a function of time by plotting curves of their incidence (Figure 2). In some cases, a single HCoV-OC43 genotype circulated during given periods of time, such as for the case of genotype K in 2021 (Figure 2). For other periods of time, several genotypes co-circulated, as for the case of genotypes J and K in 2020, or of all three genotypes in 2018 and in 2022.

Mutation patterns

The patterns of mutations within genomes were analyzed according to the viral genotype. Regarding genotype G genomes, 991 nucleotide substitutions were detected compared with the reference genome GenBank accession no. MN026164.1 dating back to 2018. To gain a better picture of the distribution of the mutations, we classified the genes into "structural", "informative", "other non-structural" and "accessory" categories. In the informational gene category, NSP13 that encodes the helicase was the most affected gene, with 46 substitutions per 1,000 nucleotides and NSP15 that enccodes an endoribonuclease was the least affected, with 5 substitutions per 1,000 nucleotides (Table 2). In the structural gene category, the envelope gene showed the greatest diversity with 106 substitutions per 1,000 nucleotides and the matrix encoding gene was the least affected with 4 substitutions per 1,000 nucleotides. A total of 303 amino acid substitutions were detected relatively to the same reference genome. In the informational gene category, NSP13 was the most affected with 44 amino acid substitutions per 1,000 nucleotides and NSP8 exhibited no amino acid substitutions. In the structural genes, the most affected gene was the envelope gene and the least affected was the matrix gene (Table 2). We observed that 41 different amino acid substitutions were encoded by at least 10 of the genotype G genomes. The genes where these mutations stood were Nsp2, Nsp3, Nsp4, Nsp12, Nsp13, Nsp14, and Nsp16 regarding non-structural genes, and HE, S, N and N2 regarding structural genes. A majority of these mutations were located in the spike gene and were present in the majority of the genomes

197 analyzed here. The other genes including Nsp1, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10,
198 Nsp11, Nsp15, M, E and the accessory genes were considerably conserved (Figure 3A).
199 Regarding genotype J genomes, their mutation patterns were analyzed relatively to
200 reference genome OK318939.1 that dates back to 2018. In the informational gene category,
201 the gene with the greatest nucleotide diversity was Nsp13, and the one with the lowest
202 diversity was Nsp10, with 116 and 9 substitutions per 1,000 nucleotides, respectively. In the
203 structural gene category, the gene most affected by nucleotide diversity was the one encoding
204 the hemagglutinin esterase with 246 substitutions per 1,000 nucleotides, followed by the spike
205 gene with 174 substitutions per 1,000 nucleotides, and the gene the least affected was the
206 envelope gene with 45 substitutions per 1,000 nucleotides (Table 3). A total of 940 amino
207 acid substitutions were observed to be encoded by the genotype J genomes obtained here
208 relatively to genome OK318939.1. In the informational gene category, Nsp13 was the most
209 mutated with 89 amino acid substitutions per 1,000 nucleotides and Nsp10 was the least
210 affected with no amino acid substitution. In the informational gene category, Nsp13 was the
211 most affected with 89 amino acid substitutions per 1,000 nucleotides and Nsp10 was the least
212 affected with no amino acid substitution. In the structural gene category, the HE encoding
213 gene was the one exhibiting the greatest amino acid diversity, while the spike and
214 nucleocapsid genes were those the less mutated. In the accessory gene category, NS2 showed
215 the greatest amino acid diversity with 413 amino acid substitutions per 1,000 nucleotides
216 (Table 3). Overall, 64 amino acid substitutions were encoded by at least 10 o genotype J
217 genomes. A great number of mutations were observed in non-structural genes, the most
218 mutated of which being Nsp3. The structural genes with amino acid changes were HE, spike
219 and nucleocapsid. There were also mutations in the NS2 accessory gene. Finally, the most
220 conserved genes were Nsp7, Nsp10, Nsp11, Nsp14, M, E and the other accessory proteins
221 apart from NS2 (Figure 3B).

Regarding genotype K genomes, mutational diversity was analysed relatively to genotype K reference genome MW532118.1 that dates back from 2017. The gene with the greatest nucleotide diversity in the informational genes category was Nsp13 and the gene with with the lowest diversity was Nsp10, with 166 and 31 substitutions per 1,000 nucleotides, respectively. In the structural gene category, the HE gene was the one with the greatest diversity with 247 substitutions per 1,000 nucleotides, followed by the spike gene with 204 substitutions per 1,000 nucleotides, while the nucleocapsid gene was the one with the lowest diversity with 87 substitutions per 1,000 nucleotides (Table 4). NS2 was the accessory gene with the greatest diversity with 455 substitutions per 1,000 nucleotides. Besides, a total of 1,300 amino acid substitutions were detected in the genomes obtained in the present study relatively to the reference genome MW532118.1. In the informational gene category, the greatest amino acid diversity was observed for Nsp16 and Nsp13, with 130 and 129 amino acid substitutions per 1,000 nucleotides, respectively. Nsp10 was the gene the less affected by mutations in this category. In the structural gene category, the HE gene was the gene with the greatest diversity and the nucleocapsid gene was the gene with the lowest diversity, with 267 and 85 amino acid substitutions per 1,000 nucleotides, respectively. NS2 was the accessory gene with the greatest diversity (Table 4). Overall, 68 amino acid changes were found for at least 10 genomes. These amino acid changes were observed almost in every gene, in both non-structural and structural genes, including in Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp8, Nsp9, Nsp12, Nsp13, Nsp14, Nsp15, Nsp16, HE, S, N, N2 and the accessory gene NS2. In contrast, the majority of mutations were found in a minority of genomes, as only 20 of 68 substitutions were encoded by at least 50 genomes. The most conserved genes were Nsp1, Nsp7, Nsp10, Nsp11, M, E and the other accessory proteins apart from NS2 (Figure 3C)

Genome recombination events

Recombination events were searched with the RDP4 software for genome sequences

247 with a coverage of at least 90%. Those that were detected with at least 4 of the 7 tests
248 performed that were positive with a P value <0.01 were considered. Under these conditions,
249 we detected 61 putative recombination events, including 6 involving genotype G genomes, 24
250 involving genotype J genomes, and 31 involving genotype K genomes.
251
252 **DISCUSSION**
253 As the SARS-CoV-2 pandemic that emerged in 2019 reminded us, coronaviruses are
254 viruses that have a continuous potential of emergence both in non-human animals and in
255 humans and are zoonotic source of infections for humans. Therefore, it is critical to monitor
256 the evolution of these viruses to account for the different mutants and variants observed.
257 Among the four so-called 'endemic' coronaviruses that infect humans, HCoV-OC43 is the
258 one the most frequently detected. Here, we obtained and analyzed 185 HCoV-OC43 genomes,
259 whereas only 361genomes with a coverage >77% were available in GenBank at the beginning
260 of the present study (as of 04/10/2023). GenBank sequences were obtained primarily in China
261 and the USA. In addition, for the period between 2017 and 2022 (in all cases when the
262 respiratory sample collection date was available), there were a total of only 145 genomes
263 available worldwide, and for France, only 11 HCoV-OC43 genomes were available in
264 GenBank for any time period and none were available for the 2017-2022 period. Thus, here
265 we expanded by approximately 50% the set of HCoV-OC43 genomes available worldwide,
266 we more than doubled the set of genomes available worldwide for the 2017-2022 period, and
267 for France we expanded by approximately 17 times the number of genomes available and
268 provided the only genomes for the period from 2017 to 2022.
269 Here, the predominance of sequences for the period of time between 2020 and 2022 is
270 due to the Covid-19 pandemic which generated a massive increase in the sampling of patients
271 with signs of respiratory infection. We identified three HCoV-OC43 genotypes, G, J and K.

Genotypes J and K, which are among the latest detected HCoV-OC43 genotypes, were reported for the first time in China in 2022 (Zhang et al, 2022). We show here that these genotypes also circulated in France during the period of our study, and this findings questions on since how long it was circulating in our country. Otherwise, none of the genomes obtained here belonged to a new lineage of HCoV-OC43 that was reported to had emerged in China (Ye et al, 2023). Over the whole period of our study, either a unique genotype was found to circulate during some periods or two or three genotypes co-circulated during other periods. The HE gene was the structural gene showing the greatest diversity for genotypes J and K, while the envelope gene showed the greatest diversity for genotype G. The NSP13 gene was the informational gene with the greatest diversity The spike gene was one of those with the greatest diversity in the present study, which is in accordance with other studies showing that this gene evolves rapidly in coronavirus genomes (Ye et al, 2023; Lau et al, 2011). Oong et al reported that spike mutations D267K and I268D (which correspond to D266K and I267D, respectively, in the study by Ren et al (Ren et al, 2015)), are signature mutations of the genotype G (Oong et al, 2017). As a matter of fact, these mutations were not carried either by all the genomes obtained here or by all other genotype G genomes from GenBank, but only by some of these genomes. Oong et al also reported the presence of mutations P22T and T271S in two genomes from China (Oong et al, 2017). All the genomes we obtained carried mutation P22T, but some genomes available in GenBank did not. As for mutation T271S, some but not all genotype G genomes we obtained and from GenBank carried it while others did not. Thus, the classification of HCoV-OC43 genomes based on signature mutations needs to be checked and refined based on currently available genomes and over time, as it depends on the number and diversity of available genomes and on the geograophical areas and the periods of time they cover. Tracking HCoV-OC43 evolution based on genome analysis is therefore worthy to gain a better picture and understanding of the emergence and faith of viral

297 lineages within and between countries.

298 HCoV-OC43 is a coronavirus known to frequently recombine (Zhang et al, 2022; Zhu
299 et al, 2018; Wells et al, 2023). Recombination between two different 'parental' genomes
300 occurs when there is coinfection of two viral species or two genotypes or two viral variants in
301 the same host cell, resulting in the emergence of a virus that harbors parts of the genetic
302 material of both 'parents'. In coronaviruses, recombination is notably possible due to the
303 discontinuity of viral replication. Recombination phenomena have already been reported in
304 coronaviruses that infect both humans and animals (Wells et al, 2023). For HCoV-OC43, 6 of
305 the 11 genotypes are known to be derived from recombination events or from a recombinant
306 'parent'. Here, putative recombination events have been suspected based on bioinformatic
307 analyses in genotype K and G sequences, which is consistent with previous studies
308 demonstrating that these genotypes could recombine (Zhang et al, 2022; Zhu et al, 2018).

309 In summary, the present study demonstrates the circulation of new HCoV-OC43
310 genotypes in France between 2017 and 2022, initially detected in China and their dynamics
311 over time. It starts to fill a huge gap on HCoV-OC43 genome availability worldwide and in
312 France and points out the tremendous imbalance between the amounts of genomic data on
313 SARS-CoV-2 and on another human coronavirus much older known to cause infections in
314 humans. Further studies are needed to gain a better knowlege of whether the different HCoV-
315 OC43 genotypes are associated to particular transmissibility, antigenicity, replicative
316 capacity, or pathogenicity, and of the determinants of the rise and fall of their epidemics.
317
318

REFERENCES

Aksamentov, I., Roemer, C., Hodcroft, E. B., and Neher, R. A. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. doi: 10.5281/zenodo.5607694

Al-Khannaq, M. N., Ng, K. T., Oong, X. Y., Pang, Y. K., Takebe, Y., Chook, J. B., et al. (2016). Molecular epidemiology and evolutionary histories of human coronavirus OC43 and HKU1 among patients with upper respiratory tract infections in Kuala Lumpur, Malaysia. *Virol. J.* 13, 33. doi: 10.1186/s12985-016-0488-4

Burki, T. (2023). First shared SARS-CoV-2 genome: GISAID vs virological.org. *Lancet Microbe* 4, e395. doi: 10.1016/S2666-5247(23)00133-7

Colson, P., Chaudet, H., Delerce, J., Pontarotti, P., Levasseur, A., Fantini, J., et al. (2023). Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic. 2023.05.01.538506. doi: 10.1101/2023.05.01.538506

Gaunt, E., Hardie, A., Claas, E., Simmonds, P., and Templeton, K. (2010). Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years Using a Novel Multiplex Real-Time PCR Method. *J. Clin. Microbiol.* 48, 2940–2947. doi: 10.1128/JCM.00636-10

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*.34(23):4121-4123. doi: 10.1093/bioinformatics/bty407.

Hodinka, R. L. (2016). Respiratory RNA Viruses. *Microbiol. Spectr.* 4. doi: 10.1128/microbiolspec.DMIH2-0028-2016

Hulswit, R. J. G., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., et al. (2019). Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci.* 116, 2681–2690. doi: 10.1073/pnas.1809667116

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kong, D., Zheng, Y., Hu, L., Chen, J., Wu, H., Teng, Z., et al. (2021). Epidemiological and co-infection characteristics of common human coronaviruses in Shanghai, 2015-2020: a retrospective observational study. *Emerg. Microbes Infect.* 10, 1660–1668. doi: 10.1080/22221751.2021.1965498

Lau, S. K., Lee, P., Tsang, A. K., Yip, C. C., Tse, H., Lee, R. A., et al. (2011). Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. *J. Virol.* 85, 11325–11337. doi: 10.1128/JVI.05512-11

357 Lau, S. K. P., Patrick C Y Woo, Cyril C Y Yip, Herman Tse, Hoi-wah Tsoi, and Vincent C C
358 Chen (2006). Coronavirus HKU1 and other coronavirus infections in Hong Kong -
359 PubMed. Available at: https://pubmed.ncbi.nlm.nih.gov/16757599/ (Accessed
360 December 26, 2022).

361 Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for
362 phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi:
363 10.1093/bioinformatics/btl529

364 Liu, P., Shi, L., Zhang, W., He, J., Liu, C., Zhao, C., et al. (2017). Prevalence and genetic
365 diversity analysis of human coronaviruses among cross-border children. *Virol. J.* 14,
366 230. doi: 10.1186/s12985-017-0896-0

367 Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. (2015). RDP4: Detection and
368 analysis of recombination patterns in virus genomes. *Virus Evol.* 1(1):vev003. doi:
369 10.1093/ve/vev003.

370 Martínez, Y. A., Guo, X., Portales-Pérez, D. P., Rivera, G., Castañeda-Delgado, J. E., García-
371 Pérez, C. A., et al. (2021). The analysis on the human protein domain targets and host-
372 like interacting motifs for the MERS-CoV and SARS-CoV/CoV-2 infers the
373 molecular mimicry of coronavirus. *PloS One* 16, e0246901. doi:
374 10.1371/journal.pone.0246901

375 McIntosh, K., Dees, J. H., Becker, W. B., Kapikian, A. Z., and Chanock, R. M. (1967).
376 Recovery in tracheal organ cultures of novel viruses from patients with respiratory
377 disease. *Proc. Natl. Acad. Sci. U. S. A.* 57, 933–940.

378 Millet, J. K., Jaimes, J. A., and Whittaker, G. R. (2020). Molecular diversity of coronavirus
379 host cell entry receptors. *FEMS Microbiol. Rev.* 45, fuaa057. doi:
380 10.1093/femsre/fuaa057

381 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler,
382 A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
383 Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534. doi:
384 10.1093/molbev/msaa015

385 Monto, A. S., DeJonge, P. M., Callear, A. P., Bazzi, L. A., Capriola, S. B., Malosh, R. E., et
386 al. (2020). Coronavirus Occurrence and Transmission Over 8 Years in the HIVE
387 Cohort of Households in Michigan. *J. Infect. Dis.* 222, 9–16. doi:
388 10.1093/infdis/jiaa161

389 Otieno, J. R., Cherry, J. L., Spiro, D. J., Nelson, M. I., and Trovão, N. S. (2022). Origins and
390 Evolution of Seasonal Human Coronaviruses. *Viruses* 14, 1551. doi:
391 10.3390/v14071551

392 Papa Mze N, Kacel I, Beye M, Tola R, Sarr M, Basco L, Bogreau H, Colson P, Fournier PE.
393 (2023). High Throughput SARS-CoV-2 Genome Sequencing from 384 Respiratory
394 Samples Using the Illumina COVIDSeq Protocol. *Genes (Basel)*. 14(3):681. doi:
395 10.3390/genes14030681.

396 Salinas, S., and Simonin, Y. (2020). Les atteintes neurologiques liées au SARS-CoV-2 et
397 autres coronavirus humains. *médecine/sciences* 36, 775–782. doi:
398 10.1051/medsci/2020122

399 Saunders, N., Fernandez, I., Planchais, C., Michel, V., Rajah, M. M., Baquero Salazar, E., et
400 al. (2023). TMPRSS2 is a functional receptor for human coronavirus HKU1. *Nature*
401 624, 207–214. doi: 10.1038/s41586-023-06761-7

402 Shah, M. M., Winn, A., Dahl, R. M., Kniss, K. L., Silk, B. J., and Killerby, M. E. (2022).
403 Seasonality of Common Human Coronaviruses, United States, 2014-20211. *Emerg.*
404 *Infect. Dis.* 28, 1970–1976. doi: 10.3201/eid2810.220396

405 Sobhy, H., and Colson, P. (2012). Gemi: PCR Primers Prediction from Multiple Alignments.
406 *Comp. Funct. Genomics* 2012, 783138. doi: 10.1155/2012/783138

407 Wells HL, Bonavita CM, Navarrete-Macias I, Vilchez B, Rasmussen AL, Anthony SJ. The
408 coronavirus recombination pathway. (2023). *Cell Host Microbe.* 31(6):874-889. doi:
409 10.1016/j.chom.2023.05.003.

410 Ye, R.-Z., Gong, C., Cui, X.-M., Liu, J.-Y., Fan, H., Xie, H., et al. (2023). Continuous
411 evolution and emerging lineage of seasonal human coronaviruses: A multicenter
412 surveillance study. *J. Med. Virol.* 95, e28861. doi: 10.1002/jmv.28861

413 Yip, C. C. Y., Lam, C. S. F., Luk, H. K. H., Wong, E. Y. M., Lee, R. A., So, L.-Y., et al.
414 (2016). A six-year descriptive epidemiological study of human coronavirus infections
415 in hospitalized patients in Hong Kong. *Virol. Sin.* 31, 41–48. doi: 10.1007/s12250-
416 016-3714-8

417 Zhang, R., Wang, K., Ping, X., Yu, W., Qian, Z., Xiong, S., et al. (2015). The ns12.9
418 Accessory Protein of Human Coronavirus OC43 Is a Viroporin Involved in Virion
419 Morphogenesis and Pathogenesis. *J. Virol.* 89, 11383–11395. doi: 10.1128/jvi.01986-
420 15

421 Zhang, Z., Liu, W., Zhang, S., Wei, P., Zhang, L., Chen, D., et al. (2022). Two novel human
422 coronavirus OC43 genotypes circulating in hospitalized children with pneumonia in
423 China. *Emerg. Microbes Infect.* 11, 168–171. doi: 10.1080/22221751.2021.2019560

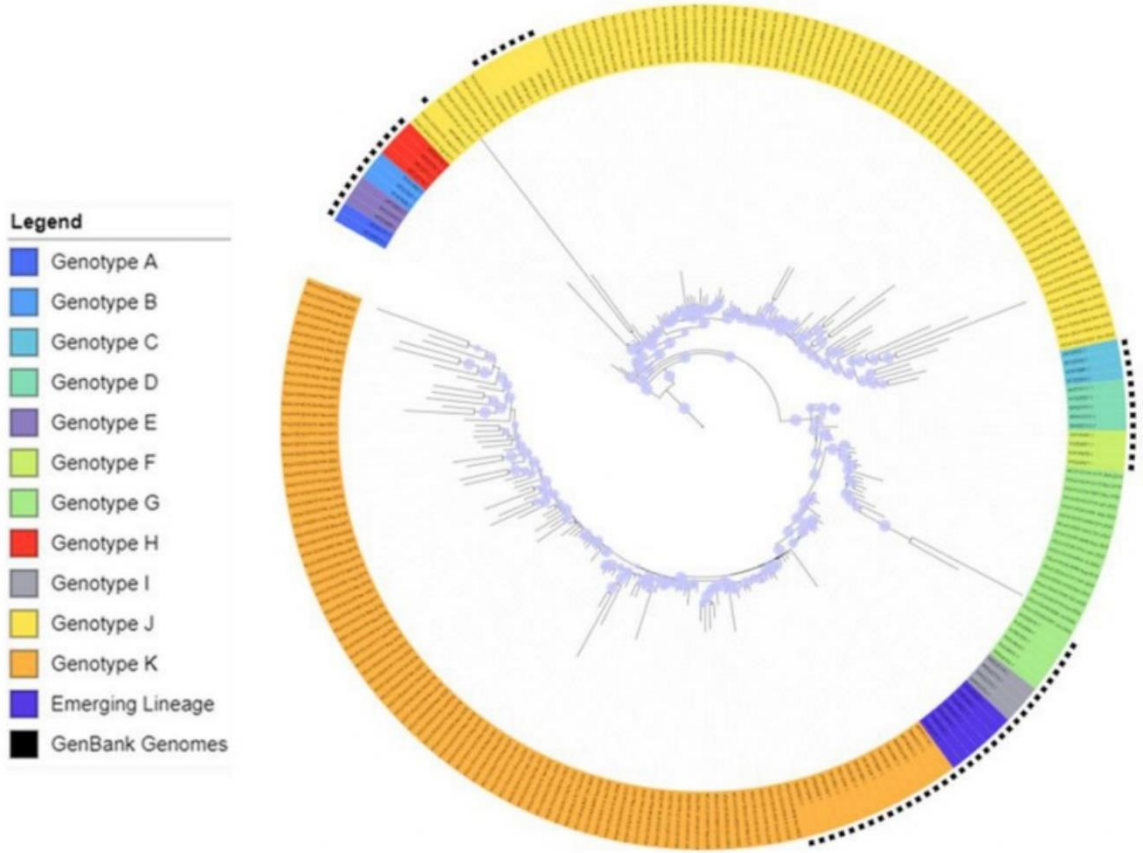
424 Zhu, Y., Li, C., Chen, L., Xu, B., Zhou, Y., Cao, L., et al. (2018). A novel human coronavirus
425 OC43 genotype detected in mainland China. *Emerg. Microbes Infect.* 7, 1–4. doi:
426 10.1038/s41426-018-0171-5

427

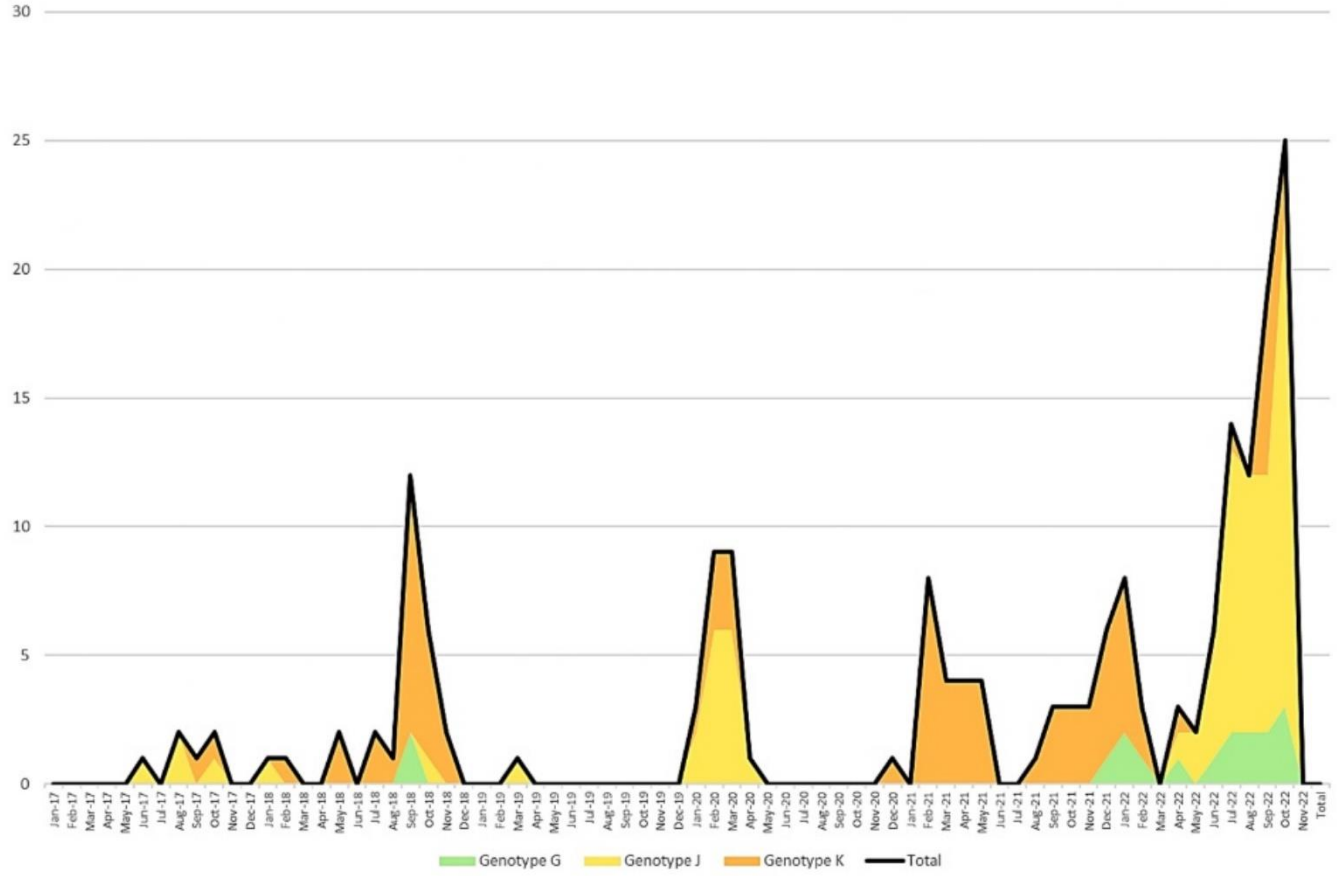
428

FIGURE LEGEND

Figure 1. HCoV-OC43 phylogenetic tree.



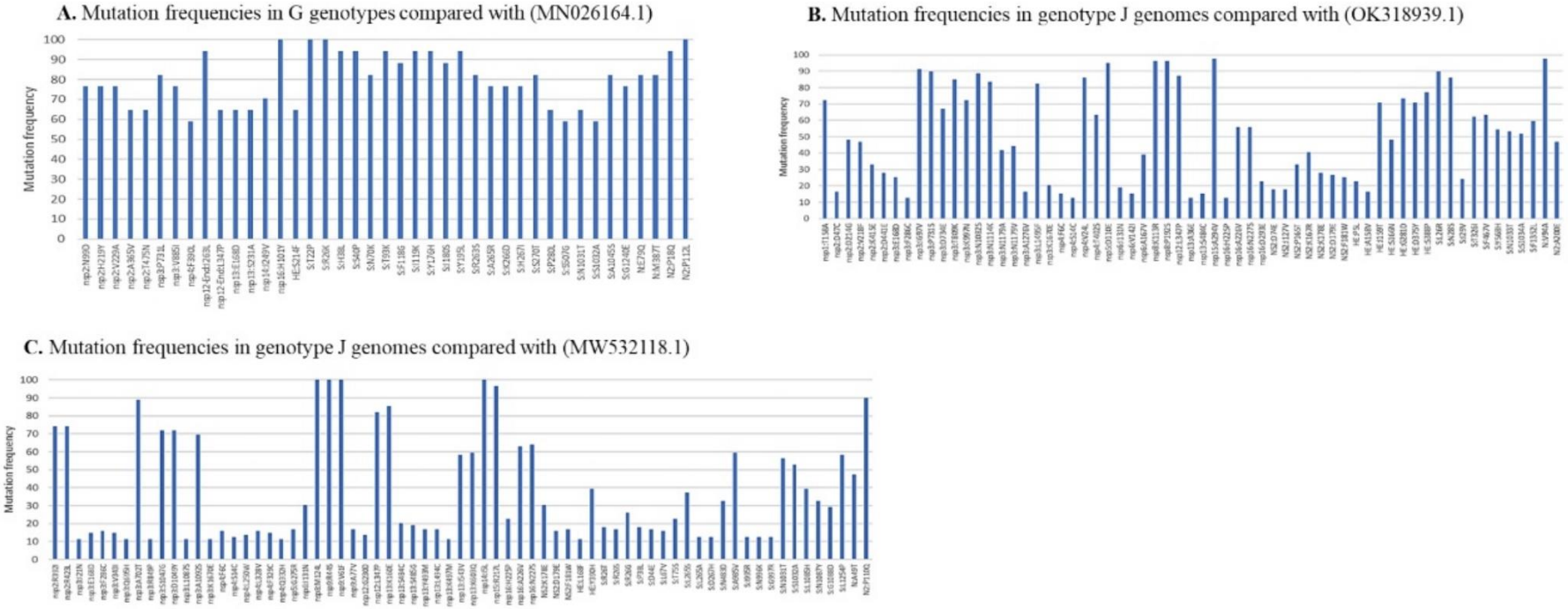
433 **Figure 2.** Distribution of HCoV-OC43 genotypes over time



434

435

Figure 3. Amino acid substitution frequencies compared to references



TABLE

Table 1. Sequences of primers used for PCR amplifcation

Primer name	Sequences (5'-3' orientation)	Pool	Concentration of use (pmol/µL)
OC43_F1	TAGGCAGTGRCCCDCCCA	1	15.3
OC43_R1	GTCTTCAGTAGGRTTAGCAACAT		
OC43_F2	GGCATGGCATGTTGTGCGT	2	12.6
OC43_R2	CACTSTCCACTACTATATAAGC		
OC43_F3	TGGAYTCTTTAGGTGCAGCT	1	15.3
OC43_R3	ACTTCAACYACACTATCYGCT		
OC43_F4	GCCTAGTCAAGTTCAGAAAGC	2	10
OC43_R4	CAGAKGTMTCCACACTTGGAC		
OC43_F5_alt2	TCCAAGTGTGGAKACMTCTGA	1	15.3
OC43_R5_alt2	TTAACATCCACAACACAAGCTG		
OC43_F6	GTTGCTGATTTTGCATATTGGT	2	10
OC43_R6	TCCAACAACATTAAGCACAGT		
OC43_F7_alt1	GCAGTAGCAGCTGGACAG	1	15.3
OC43_R7_alt1	TCTTRCCAAAACTWTCACCAAC		
OC43_F8_alt2	TGAGTGTACTGGAGGCATAG	2	10
OC43_R8_alt2	TACAACCCCTCTCATATCTCT		
OC43_F9_alt1	GTGGCCAACAGCTACAGG	1	10
OC43_R9_alt1	CCTGCCARGCAGAACTGAC		
OC43_F10	GGTTGCTAGAGGYGCTTGC	2	10
OC43_R10	AACATGCATATTAGGCACACTCT		
OC43_F11	CYACTTGGCTGCCWGAGC	1	15.3
OC43_R11	TCATCCGTCAATTCRABACCTG		
OC43_F12_alt1	TGCTTTAATAGCAACTGCGCA	2	10
OC43_R12_alt1	CCACAAAAGGTCCCAGGC		
OC43_F13	GTGAGGAAGCTGATGAGGG	1	15.3
OC43_R13	GAACATATYACRTGTCTGGGAC		
OC43_F14	ACTGTCCCAGACAYGTRATATG	2	15.3
OC43_R14	AGTATAGCAGCATARAGCCATG		
OC43_F15_alt1	TAGTACTGGTTGTCATACWGGT	1	15.3
OC43_R15_alt1	ACTGAGCAAGCTTTTCAAAAGC		
OC43_F16	GCCACTTCGGATCTGAGTG	2	10
OC43_R16	GTGGTRGCATCGGGTTTAACA		
OC43_F17	GCYGGTACYGGTATGGCC	1	10
OC43_R17	TACACGCTCATAGCATTTCATC		
OC43_F18_alt1	GCCCGTCTCGTACCCTG	2	10
OC43_R18_alt1	CTAAAACAGCAAGTGCGTAAAT		
OC43_F19	GCWGCTGATCCAGCTTTGC	1	10
OC43_R19	TAGCAAAAGCAGTAGTTGCATC		
OC43_F20	AAATTGTYATGTGTGGTGGCTGT	2	15.3
OC43_R20	CATCATTTACATCACATCCTGG		
OC43_F21	GGCGACTGATCATAAATATGTC	1	15.3
OC43_R21	ACATTGTTCTGAAACGTCTCRAG		
OC43_F22_alt1	GCRTCAGCAACAATACAAGAG	2	10
OC43_R22_alt1	ATCTGCTGCATATTTACAGCAC		
OC43_F23_alt2	ACAGTKTCYGCCTTGGTTTAT	1	15.3
OC43_R23_alt2	ATGGCGCCAACAACCATARTA		
OC43_F24	GGGCAGCCAACTTTGAGC	2	15.3
OC43_R24	ACCTGCTTAGCATCCATRCC		
OC43_F25	TTGAGYATTTGAAGCCTATGCCA	1	15.3
OC43_R25	CCCCAGATTACCTTGTGGG		
OC43_F26_alt2	CAGCTCTAACCAGAGCCCA	2	15.3
OC43_R26_alt2	TGGCATGCATAACATTTCCATC		
OC43_F27_alt2	AAGTTRGCTCTDGGTGGCAG	1	15.3
OC43_R27_alt2	ATCATCATTCTCRGGAAGATC		
OC43_F28_alt1	GCTCGGTGCTGAAAAGRAAG	2	15.3
OC43_R28_alt1	GCTAGTRAAACCAGCATCYC		
OC43_F29_alt1	CGGTGGAACAATGCYAGGC	1	15.3
)C43_R30/31_alt	AGTTAATRGGTTGCAGCTGTC		
OC43_F32_alt2	CGACAGCTGCAACCYATTAA	2	15.3
OC43_R32_alt2	AGCACATCCATRGWGACACC		
OC43_F33	GGCTGCCACYTCTGCTAG	1	15.3
OC43_R33	CATAGCTACACCARCAAGGS		
OC43_F34	ATGGCCTTGGTATGTATGGC	2	10
OC43_R34	GGGCCACATRAKCCACAAAATA		
OC43_F35	GCAATWTGGRTATACAAGTCGC	1	10
OC43_R35	AGCAGACCTTCCTGAGCC		
OC43_F36_alt1	ATCTGGGAACMGGACCGC	2	10
OC43_R36_alt1	GCTCRICCATCTTCTTAAGAA		

Degenerate bases are indicated by a red color

445 **Table 2.** Frequency of nucleotide and protein mutations in HCoV-OC43 genotype G

446

Nucleotide

Genotype	Gene	Nucleotide size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
G Genotype	Nsp1	738	1	1
	Nsp2	1815	71	39
	Nsp3	5697	130	23
	Nsp4	1488	107	72
	Nsp5	909	8	9
	Nsp6	861	4	4
	Nsp7	267	2	7
	Nsp8	591	4	6
	Nsp9	330	3	9
	Nsp10	411	5	12
	Nsp11	42	0	0
	Nsp12	2783	82	29
	Nsp13	1809	85	46
	Nsp14	1563	13	8
	Nsp15	1125	6	5
	Nsp16	897	22	24
	NS2	837	65	77
	HE	1284	58	45
	S	4077	176	43
	NS12.9	330	6	18
	E	255	27	106
	M	693	3	4
	N	1347	36	26
	N2	348	5	14

Protein

Genotype	Gene	Protein size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
G Genotype	Nsp1	222	1	4
	Nsp2	587	15	25
	Nsp3	2029	36	17
	Nsp4	496	35	70
	Nsp5	303	2	6
	Nsp6	287	4	14
	Nsp7	92	1	11
	Nsp8	194	0	0
	Nsp9	110	1	9
	Nsp10	137	1	7
	Nsp11	14	0	0
	Nsp12	928	35	37
	Nsp13	603	27	44
	Nsp14	521	2	4
	Nsp15	374	4	10
	Nsp16	299	12	40
	NS2	278	25	90
	HE	427	17	40
	S	1358	57	42
	NS12.9	109	2	18
	E	84	12	143
	M	230	1	4
	N	448	8	18
	N2	115	5	43

447

448

449

450 **Table 3.** Frequency of nucleotide and protein mutations in HCoV-OC43 genotype J

451

Nucleotide

Genotype	Gene	Nucleotide size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
J Genotype	Nsp1	738	101	137
	Nsp2	1815	157	86
	Nsp3	5697	329	57
	Nsp4	1488	168	113
	Nsp5	909	31	34
	Nsp6	861	70	81
	Nsp7	267	10	37
	Nsp8	591	11	18
	Nsp9	330	5	15
	Nsp10	411	4	9
	Nsp11	42	0	0
	Nsp12	2783	244	87
	Nsp13	1809	210	116
	Nsp14	1563	45	28
	Nsp15	1125	61	54
	Nsp16	897	57	63
	NS2	837	289	345
	HE	1272	313	246
	S	4083	713	174
	NS12.9	330	11	33
	E	267	12	45
	M	693	48	69
	N	1344	74	55
	N2	624	36	57

Protein

Genotype	Gene	Protein size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
J Genotype	Nsp1	222	28	126
	Nsp2	587	58	99
	Nsp3	2029	107	52
	Nsp4	496	58	117
	Nsp5	303	9	29
	Nsp6	287	25	87
	Nsp7	92	4	43
	Nsp8	194	4	20
	Nsp9	110	4	36
	Nsp10	137	0	0
	Nsp11	14	0	0
	Nsp12	928	47	50
	Nsp13	603	54	89
	Nsp14	521	8	15
	Nsp15	374	18	48
	Nsp16	299	20	67
	NS2	278	115	413
	HE	423	127	300
	S	1360	189	139
	NS12.9	109	3	27
	E	88	7	79
	M	230	18	78
	N	447	22	49
	N2	207	15	72

452

453

454

455 **Table 4.** Frequency of nucleotide and protein mutations in HCoV-OC43 genotype K

Nucleotide

Genotype	Gene	Nucleotide size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
G Genotype	Nsp1	738	64	86
	Nsp2	1815	246	135
	Nsp3	5697	672	118
	Nsp4	1488	313	210
	Nsp5	909	105	115
	Nsp6	861	85	98
	Nsp7	267	11	41
	Nsp8	591	46	78
	Nsp9	330	21	63
	Nsp10	411	13	31
	Nsp11	42	0	0
	Nsp12	2783	315	113
	Nsp13	1809	300	166
	Nsp14	1563	73	46
	Nsp15	1125	78	69
	Nsp16	897	103	115
	NS2	837	381	455
	HE	1284	318	247
	S	4077	833	204
	NS12.9	330	31	94
	E	255	49	192
	M	693	78	112
	N	1347	117	87
	N2	624	57	91

Protein

Genotype	Gene	Protein size	Number of mutation per gene	Number of mutations per 1,000 nucleotides
K Genotype	Nsp1	222	14	63
	Nsp2	587	82	139
	Nsp3	2029	182	89
	Nsp4	496	106	213
	Nsp5	303	34	112
	Nsp6	287	25	87
	Nsp7	92	6	65
	Nsp8	194	22	113
	Nsp9	110	10	91
	Nsp10	137	5	36
	Nsp11	14	0	0
	Nsp12	928	58	62
	Nsp13	603	78	129
	Nsp14	521	32	61
	Nsp15	374	25	67
	Nsp16	299	39	130
	NS2	278	138	496
	HE	427	114	267
	S	1358	228	168
	NS12.9	109	11	101
	E	84	15	178
	M	230	20	87
	N	448	38	85
	N2	207	18	87

456

457

458

Article n°10 :
Characterization of human parainfluenza virus type 2 (HPIV2) circulating in southern France between 2017 and 2022 using whole genome sequencing

Enzo Bisi, Céline Boschi, Houmadi Hikmat, Jérémy Delerce, **Lorlane Le Targa**, Lucile Lesage, Aurélie Morand, Bernard La Scola , Philippe Colson (Soumis dans Journal of Virologica Sinica)

TITLE PAGE

Target journal and type of article: Virologica Sinica, Research article

Characterization of human parainfluenza virus type 2 (HPIV2) circulating in southern France between 2017 and 2022 using whole genome sequencing

Short title (for the running head): HPIV2 whole genome sequencing

Author list: Enzo BISI[1,2], Céline BOSCHI[1,2,3], Houmadi HIKMAT[1,2], Jérémy DELERCE[2,3], Lorlane LE TARGA[1,2,4], Lucile LESAGE[5,6], Aurélie MORAND[5,6], Bernard LA SCOLA[1,2,3]* , Philippe COLSON[1,2,3]*

Affiliations : [1] Microbes Evolution Phylogeny and Infections (MEPHI), Aix-Marseille Univ. (AMU), Marseille, France; [2] IHU Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France; [3] Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France; [4] Biosellal, 27 Chemin des Peupliers, 69570 Lyon, France; [5] Service d'accueil des Urgences Pédiatriques, hôpital Nord, AP-HM, chemin des Bourrely, 13015 Marseille, France; [6] Service de Pédiatrie générale, hôpital Timone, AP-HM, 264 rue Saint-Pierre, 13005 Marseille, France

*** Corresponding authors:** Bernard La Scola and Philippe Colson IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: bernard.la-scola@univ-amu.fr, philippe.colson@univ-amu.fr

Keywords (MeSH): Next-generation sequencing; Parainfluenza virus 2; Genome; Human; Mutation; Genotype; Recombination; Evolution

Word counts: abstract, 249; text, 4,927

Figures: 5; **Tables:** 2; **References:** 39

Supplementary material: Figures: 1**; Tables:** 4

ABSTRACT

Genomic surveillance of respiratory viruses is an expanding field. As of 2023, only 71 (near-)complete genomes of human parainfluenza virus type 2 (HPIV2) were available in GenBank, none of which being from France. Here we aimed to obtain and analyse HPIV2 genomes from patients diagnosed HPIV2-positive in public and university hospitals of Marseille during the 2017-2022 period. Next-generation sequencing (NGS) used Illumina technology on a NovaSeq 6000 instrument. Prior to NGS, a PCR-based enrichment strategy was used with primers chosen with PrimalScheme. HPIV2 genomes were generated from NGS reads by mapping and *de novo* assembly using CLC Genomics. Mutations were identified by Nextclade, phylogeny was performed by MEGA, and recombinants were detected with RDP4. Seventy-seven genomes were recovered from 318 samples. The two major genotypes G1a (n= 129 genomes) and G3 (110) were identified, and 5 subgenotypes including G1a.5 (23), G1a.8 (68), G3.4 (4), G3.5 (7), and G3.7 (23) were newly proposed. Genotype-specific mutations were located on the L gene (RNA-dependent RNA polymerase) for G1a (T14612A/T11981A/T12374C) and on the HN (Hemagglutinin-neuraminidase) gene for G3 (G8171A/A14057G/10367G). G1a.5 and G3 subgenotypes were only detected in 2019. Genotype G1a was absent during winter 2021-2022 then became majoritary again during fall 2022. Nineteen putative recombinants were detected. Although preliminary, this work doubles the number of HPIV2 genomes available and produces the first genomes in France. It shows an evolution of genotypes and their distribution with significant genetic diversity, and suggests the presence of recombinants. These data justify genomic epidemiological surveillance of this virus.

TEXT

INTRODUCTION

Historically, viruses associated with respiratory diseases have caused global problems (Nicolle and Lebailly, 1918). The knowledge of their epidemiological, virological and clinical features has been largely improved by the development of molecular biology that includes PCR and sequencing technologies and led to today's genomics (Hilt and Ferrieri, 2022; Petrone, 2020). The development of whole genome sequencing of respiratory viruses has been a valuable step in the genetic analysis of these viruses. PCR, experimented for the first time in 1988 (Chamberlain et al., 1988), allowed three decades later in 2017 through the amplication of multiple overlaping genomic regions to obtain the complete sequence of the Zika virus genome, after a tedious optimization phase (Quick et al., 2017). This work served as a basis for the development in 2020 during the Covid-19 crisis of the 'ARTIC' protocol to optimize SARS-CoV-2 genome sequencing (Quick, 2020). This has been combined with bioinformatic tools to provide efficient mutational analyses such as the NextStrain/NextClade pipeline, not only for SARS-CoV-2 but also for other viruses responsible for respiratory infections.

Here, we performed genome sequencing of human orthorubulavirus 2 (commonly referred to as Parainfluenza virus type 2 or HPIV2). Discovered and named in the late 1950s (Andrewes et al., 1959; Chanock, 1956), HPIV2 was registered as an *Orthorubulavirus* subspecies in 2022 according to the ICTV (Walker et al., 2022) and is mainly responsible for lower respiratory tract infections in children under the age of 5 years. The majority of publications on its genetic features focused on the hemagglutine-neuraminidase (HN) or fusion protein (F) encoding genes, highlighting the HN gene as a hotspot of variation (Terrier et al., 2008; Tsurudome et al., 1989), while neglecting other parts of the HPIV2 genome. A

74 first description of HPIV2 genotypes was carried out by Santak et al in 2016 and refined in
75 2018 with six genotypes (G1a, G1b, G2, G3, G4a and G4b) solely based on a HN gene
76 analysis (Santak et al., 2016; Santak et al., 2018). This very important inaugural work
77 proposed the use of a genome-wide phylogenetic tree instead of a simple HN-based tree to
78 produce the phylogeny of this virus. A first PCR-based procedure for the whole-genome
79 amplification and sequencing of HPIV2 was proposed in 2018 using 17 sequences (Santak et
80 al., 2018).
81 As of 2023, 75 HPIV2 genomes were available worldwide in databases including
82 GenBank (Sayers et al., 2023) and VipR (Pickett et al., 2012). In the present work, 77
83 complete HPIV2 genomes and 239 partial genomes were obtained via a whole genome PCR
84 amplification protocol and sequencing with the Illumina technology on a NovaSeq 6000
85 instrument. Comprehensive mutational analysis of these sequences using a custom
86 NextStrain/NextClade pipeline was performed, which retrieved a high degree of HPIV2
87 genetic diversity, and led to the proposal of potential new subgenotypes and to the first
88 description of a potential HPIV2 recombinant genome.
89
90 **MATERIALS AND METHODS**
91 ***Respiratory samples***
92 HPIV2 RNA-positive nasopharyngeal respiratory samples that had been sent for the
93 purpose of routine diagnostic to the clinical virology laboratory of the University and Public
94 Hospitals of Marseille, southeastern France were used. HPIV2 had been diagnosed by qPCR
95 using FTD Respiratory pathogens 21 assay (FTD) (Fast Track Diagnosis, Luxembourg) after
96 MagMax viral/pathogenic nucleic acid using the isolation kit on the KingFisher Flex system
97 (Thermo Fisher Scientific, Waltham, MA, USA), or the BioFire FilmArray 2 plus assay
98 (BioFire) (Biomérieux, Marcy-l'Etoile, France), following the manufacturers' instructions.

Primer design for multiplex PCR

Primers for the multiplex PCR system were first designed using the Primal Scheme (Quick et al., 2017) and then manually modified for their optimization. An alignment was carried out with the MAFFT software ver:7.520 (Katoh et al., 2002) using all 50 complete genomes except those with fragments with stretches of more than five undetermined nucleotides available in the NCBI GenBank nucleotide sequence database (Sayers et al., 2023) that covered the 1990-2022 period. For the optimization step, samples were extracted from A culture supernatant filtered at 0.2 µm with the EZ1 Virus Mini Kit v2.0 assay (Qiagen, Antwerp, Belgium; Ref: 955134).

RNA extraction and cDNA amplification by PCR

We used the Invitrogen SuperScript III One-Step RT-PCR System with the Platinum Taq High-Fidelity DNA Polymerase (Thermo Fischer, ref: 12574030). Three different PCR primer pools were used to avoid cross-hybridization or primer overlap. The concentration of each primer stock solution was 100 µM for primers for which final concentration was 0.5 µM and 40 µM for primer for which final concentration was 0.2 µM. The constitution of the primer pools and the concentration of each primer are detailed in Supplementary Data Table 1. Equal volumes of each primer at the defined concentration were pooled. The PCR included 12.5 µL of 2X Reaction Mix, 8 µL of extracted RNA, 0.7 µL of SuperScript III RT/Platinum Taq High Fidelity Enzyme Mix, which was completed with a variable volume depending on the pool of primers used : 2.75 µL of Pool 1 or 2 µL of Pool 2 or 1 µL of Pool 3. DNAse/RNAse-free Ultra-Pure distilled water was added to reach 25 µL of final mixture. The final concentration of primers in the mixture was 0.5 µM. A reverse transcription step was carried out at 60°C for 30 seconds to synthesize a cDNA molecule. This was followed by denaturation at 94°C for 2 minutes. Then 40 PCR cycles were carried out with the following conditions: 94°C for 15 sec, 60.4°C for 30 sec, and 68°C for 2.5 min. A final elongation step

was carried out at 68°C for 5 min.

Next-generation sequencing

NGS was performed for the development step by the Oxford Nanopore Technology and the Ligation Sequencing Kit (SQK-LSK109) library with a SpotON flow cell Mk I, R9.4.1 and on a GridION instrument, following the manufacturer's instructions (Oxford Nanopore Technologies, Oxford, UK). Then, to obtain the HPIV2 genomes from the studied nasopharyngeal samples, NGS was performed by the Illumina technology with a NovaSeq 6000 instrument (Illumina Inc., San Diego, CA, USA) by the COVIDSeq protocol except with replaced the COVID-19 ARTIC PCR primers by those PCR primers we designed here. The NovaSeq 6000 instrument was loaded using a SP flow cell and the NovaSeq-XP workflow with a 2x50 reading, according to a previously described process (Papa Mze et al., 2023).

Bioinformatic analyses

Paired-end reads from Illumina NovaSeq were used to generate contigs via CLCGenomics (CLC Genomics Workbench Version 7.5 QIAGEN Aarhus A/S www.clcbio.com) with a minimum word size of 20 and bubble size of 50. The Dragen Bcl Convert pipeline (Illumina Inc.) was used to make the base call. Two reference sequences were used. The first one (NC_003443.1) was a sequence tagged as of reference by GenBank and for which sequences were first described in 1990 (Kawano et al., 1990) but the and released in 2018. The second one (KY674954.1) was a sequence from a sample collected in Seattle in 2010, which was chosen as the one the most similar to generate contigs. Contigs mapping was performed with the default setting of CLCGenomics.

Phylogenetic trees were generated using MEGA11 (Molecular Evolutionary Genetics Analysis) using the Maximum Likelihood method and 1,000 bootstrap replicates, from a alignment carried out by the Mafft web application

149 (https://mafft.cbrc.jp/alignment/server/index.html) with default parameters. The assignment to
150 HPIV2 genotypes was carried out in two steps. First, genotype assignment was done by
151 phylogeny reconstruction that included all HPIV2 sequences available from GenBank and
152 from Santak, et al. (M. Santak et al., 2018); each GenBank sequence that was clustered with a
153 bootstrap value ≥75% with those described by Santak, et al. were marked with the same
154 genotype than those latter. Second, a phylogenic tree was rebuilt that incoporated the
155 sequences obtained in the present study. HPIV2 genotype assignment method was the same
156 than above but was enriched with the addition of subgenotypes for each node with a bootstrap
157 value ≥ 75%. To study mutations of a specific genotype, a custom NextStrain/NextClade
158 pipeline was built. Data files were modified as outlined in the NextClade guidelines
159 (Aksamentov et al., 2021; "nextstrain/nextclade," 2024). KY674954.1 was the reference
160 HPIV2 genome sequence used in this pipeline.
161 For the study of sequence diversity, genome alignments were carried out with the
162 sequences obtained here combined with GenBank reference genomes. The sequences
163 diversity was calculated after removing the three reference sequences. To asses the diversity
164 intra-sample a distance matrix was done using BioEdit software version 7.7.1 (Dagona, 1999)
165 (https://thalljiscience.github.io). The nucleotide mutation rate for a given gene was calculated
166 as the number of nucleotide substitutions per 100 nucleotides. Nucleotide and amino acid
167 mutations frequencies in consensus genomes and proteins were determined using an in-house
168 Python script (https://www.python.org/) freely available at https://github.com/sanschiffre/ihu-
169 scripts/tree/main (Colson et al., 2024), which as input the output.csv file of the Nextclade tool
170 (https://clades.nextstrain.org/) (Aksamentov et al., 2021) and allows counting the number of
171 each different nucleotide or amino acid mutation.
172 Detection of recombinant viral genomes was performed with RDP4 Recombination
173 Detection Program version 4.101 (http://web.cbio.uct.ac.za/~darren/rdp.html) (Martin et al.,

2015). The HPIV2 genomes' alignment used consisted of all sequences obtained here with at least 70% coverage of genome KY674954.1, as well as of GenBank reference sequences for HPIV1, 2 and 3. A comprehensive exploratory recombinant analysis was performed with the default settings. Quality criteria were chosen based on the literature (Brnic et al., 2023; Dong et al., 2023; Gambaro et al., 2023; He et al., 2022; Lambrechts et al., 2023; Mencattelli et al., 2023). A recombinant was considered valid if both parental genomes were identified and at least five performed tests were positive with a p-value ≥ 0.01. Quality control was performed by removing recombination events located near a deletion breakpoint labeled "Unknown," those involving reference sequences, and those falling below the quality threshold.

Statistical analyses

Statistical analyses were performed using the OpenEpi web application (https://www.openepi.com/).

RESULTS

Obtaining and classification of whole HPIV2 genomes

The designed PCR-based enrichment pre-NGS consisted of three PCR primer pools (named P1, P2, and P3) to avoid potential overlaps of primers and PCR products. These pools contained 11, 8, and 4 primer pairs, respectively (Supplementary Table S1). They were used on residues from 318 nasopharyngeal swabs to perform PCR enrichment of HPIV2 genome fragments before NGS performed with the Illumina technology on a NovaSeq 6000 instrument. The contigs generated by CLC Genomics from NGS reads using the *de novo* assembly algorithm were mapped against the NCBI reference sequence NC_003443.1 for which sequences were first described during the 1990s. HPIV2 genome sequences with a coverage of NC_003443.1 greater than 90% were obtained for 77 (24%) of the 318 samples. This set will be used as a complete genome set. For these 77 genomes, mean (± standard

deviation) coverage was 92.7 ± 1.8 % (range 90-96%). Compared to another, more recent (2010) reference sequence (KY674954.1), 239 exhibited a coverage of ≥70%.

Among the set of 77 genomes with a coverage of NC_003443.1 ≥ 90%, 40 and 37 were classified as of genotype G1a and G3, respectively. Among the set of 239 genomes, 129 were classified as of genotype G1a and 110 were classified as of genotype G3.Beyond, both genotypes G1a and G3 split into a multitude of subgenotypes (Figure 1). Thus, a total of 17 subgenotypes were newly delineated here. They included subgenotypes 1 (n= 2 genomes), 5 (3), 6 (1), 7 (1) and 8 (24) within genotype G1a and subgenotypes 2 (2), 3 (2), 4 (2), 6 (2), 7 (9) and 8 (1) for genotype G3. In addition, two GenBank sequences (MK167027 and MF077311) were clustered together as a distinct lineage we named G1a.9. The delineation of and classification into these subgenotypes were supported by genome clustering in the phylogenetic analysis with bootstrap values ≥75%.

Mutational patterns of HPIV2 genomes

We thereafter considered the set of 77 most complete genomes. They included subgenotypes G1a.1 (n= 2 genomes), G1a.5 (3), G1a.6 (1), G1a.7 (1) and G1a.8 (24) and subgenotypes G3.2 (2), G3.3 (2), G3.4 (2), G3.6 (2), G3.7 (9) and G3.8 (1). For these 77 genomes, the number of different nucleotide mutations was 4784. Mean number of mutations per genome was 541.3 ± 368.5 (range, 121- 954). Firstly, regarding nucleotide mutations, 41,681 mutations were observed compared with reference genome KY674954.1. They consisted of 41,312 substitutions, 95 deletions and 274 insertions. Four mutations, namely G7844A, G12254A, C13012A, and C13436U were found in all 77 genomes compared with the reference used. These four mutations are located in the HN gene for the first one, and in the L (RNA-dependent RNA polymerase-encoding) gene for the threes. Four other mutations, namely G2243A, A7100G, U12830C, and G1298A, were present in between 73 and 76 genomes. Mutations at positions 2243, 12830 and 1298 were in fact present in all genomes for

224 which these positions were covered, and were therefore potentially present in all genomes.
225 These four mutations were present in the P (phospho-protein-encoding) , HN, L and NP
226 (nucleoprotein-encoding) genes, respectively. In addition, mutations U1777C, U4793C,
227 A9956G and U14511C were found in ≥62 genomes 80%) of all 77 genomes. They are located
228 between the NP and P (phospho-protein-encoding) genes for mutation U1777C, and in the L
229 gene for the three others.
230 A total of 29,046 nucleotide mutations, comprising 28,897 substitutions, 19 deletion
231 and 130 insertions, were G1a-specific, A total of 196 substitutions were detected in the 40
232 G1a genomes: 123 were located in the L gene, 31 in the HN gene, 14 in the F gene, 9 in the P
233 gene, 6 in the NP gene, and 3 in the M (matrix protein-encoding) gene. For the G3 genotype,
234 a total of 4,014 mutations were observed including 3,865 substitutions, 29 deletions and 120
235 insertions. Six substitutions were present in ≥25 of the 37 G3 genomes: G8171A, A14057G,
236 A10367G, A15029G, C5003A and A6525G, which are located in the HN, L, L, L, and F
237 genes and between F and HN genes, respectively. Overall, for genotypes G1a and G3, 196
238 and 6 nucleotide substitutions can be considered as signature mutations of these genotypes,
239 respectively. Considering all 77 genomes, the two most frequent types of nucleotide
240 substitutions were U > C (26.1%) and C > U (21.1%), which are potential signature mutations
241 of cellular enzymes Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
242 (APOBEC). The six other most frequent substitutions were A > G (16.4%), G > A (14.2%), C
243 > A (5.5%), U > A (4.6%), U > G (2.3%), and G > C (1.6%). In G1a genotype the three most
244 frequent substitutions were U > C (24.5%), C > U (22.9%), U > A (3.6%), and in G3
245 genotype they were U > C (18.8%), C > U (14.7%), U > A (11.8%).
246 As for amino acid mutations, a total of 5,950 mutations were detected relatively to
247 reference genome KY674954.1. They included 5,854 substitutions, 54 deletions and 42
248 insertions. The three most frequent mutations were F:T102P, NP:L474P, and NP:A184D,

249 which were present in 52, 50 and 46 of the 77 genomes, respectively; Nine other substitutions
250 were present in ≥36 genomes: F:G22D, F:I32V, P:A49V, P:R60K, P:T86I, L:A1438D,
251 M:L75I, M:M145T, and F:K163R. A total of 3,759 specific amino acid mutations were
252 observed in the G1a genotype, including 3,719 substitutions, 28 deletions, and 12 insertions.
253 Within this set, the following 12 substitutions were present in ≥35 of the 40 G1a genomes:
254 F:I32V, P:A49V, P:R60K, M:L75I, M:M145T, HN:V67I, HN:T114A, P:A137V, HN:K139E,
255 HN:S175I, HN:T195A, and M:E214D. For genotype G3, a total of 1,245 mutations were
256 detected, comprising 1,209 substitutions, 11 deletions, and 25 insertions. Of these, four were
257 present in at ≥17 of the 37 G3 genomes: F:N322S, P:N92T, L:I226V, and HN:T519A.
258 Overall, the three most prevalent amino acids substitution were L > P (3.6%), L > I (2.7%),
259 and I > T (1.8%). In G1a genotype, the three most frequent substitutions were I > V (6.2%), T
260 > A (5.5%), and V > I (3.9%). In G3 genotype, they were L > P (4.3%), S > T (2.8%), and A
261 >V (2.8%). Overall, for genotypes G1a and G3, 12 and 4 amino acid substitutions can be
262 considered as signature mutations of these genotypes, respectively.
263 We further considered the 77 genomes covering ≥90% of NC_03443.1. For genotype
264 G1a, the HN and Phosphoprotein (P) encoding genes exhibited the highest substitution rate,
265 with 30 and 23.8 substitutions per 100 nucleotides, respectively. The genes exhibiting the
266 lowest substitution rate were those encoding the large RNA-dependent RNA polymerase
267 protein (L) (11.1substitutions per 100 nucleotides) and the nucleocapsid protein (NP) (18.1).
268 The gene ranking regarding their diversity rate varied depending on the subgenotypes
269 (Supplementary Table S2A). The other highly mutated gene were the F gene for subgenotype
270 G1a.5 (7.6) and the M gene for G1a.8 (15.2). For G1a.5 the HN, P and F genes have a
271 substitution rate of 10.6, 8.4 and 7.6 respectively, while the lowest substitution rate are for the
272 M (6), NP (5.6) and L (4.4). Then for G1a.8 it's the HN, P and M encoding genes with the
273 highest substitution rate (20.2, 16.1 and 15.2 respectively) and the NP, F and L genes for the

274 lowest with 12.5, 12.4 and 8.1 respectively. In genotype G1a, 2,277 of 2,809 substitutions are
275 specific to G1a. In subgenotype G1a.5, 235 of 1,070 substitutions were subgenotype-specific,
276 while in subgenotype G1a.8 it was the case for 1,120 of 1,955 substitutions.
277 Regarding genotype-specific nucleotide substitutions, for G1a the genes exhibiting the
278 greatest substitution rate were the HN gene (24.1 substitutions per 100 nucleotides), the P
279 gene (18.9 substitutions per 100 nucleotides), and the Matrix protein (M) gene (18.1
280 substitutions per 100 nucleotides). For subgenotype G1a.5, the P, HN and Fusion protein (F)
281 genes were those carrying the greatest number of specific substitutions (3, 2.8 and 1.6
282 substitutions per 100 nucleotides, respectively), while for subgenotype G1a.8 this was the
283 case of the HN, P and M genes (12.5, 10.6 and 10.4 substitutions per 100 nucleotides,
284 respectively) (Table 1A; Figure 2). Otherwise, a nucleotide insertion (3376C) was present in
285 10 of 24 subgenotype G1a.8 genomes. These mutations could potentially serve as 'signature'
286 for the genotype G1a and its subgenotypes (Table 1B). For genotype G3, the second major
287 genotype, the HN and P genes carried the greatest substitutions rate with respectively 22.2
288 and 19.7 substitutions per 100 nucleotides (Supplementary Table S2A). The two genes with
289 the lowest substitutions rate were those corresponding to the L protein (10.9 substitutions per
290 100 nucleotides) and to the NP (13.5).The F gene is the other gene with a high substitution
291 rate (16.0). For subgenotype G3.4, the genes with the greatest substitution rate were the HN
292 (2.27 substitutions per 100 nucleotides), the P gene (2.03), and the M gene (1.32), while for
293 subgenotype G3.7 the genes with the greatest substitution rate were P (5.99 substitutions per
294 100 nucleotides), HN (4.84), and F (4.05) (Supplementary Table S2A). In genotype G3,
295 1,710 of 2,242 substitutions are specific to G3. In subgenotype G3.4, 166 of 198 substitutions
296 were subgenotype-specific, while in subgenotype G3.7 it was the case for 514 of 546
297 substitutions.
298 Regarding specific substitutions for G3 (Table 1A), the HN, P, and FF genes carried

299 the greatest substitutions rate (16.3, 14.8 and 12.3 substitutions per 100 nucleotides,
300 respectively) . Genotype G3 did not exhibit specific signature insertions, 2733:G being the
301 most detected in 3 genomes over 37. There's no specific frame shifts, but a specific deletion
302 of one nucleotide at position 6,636 is detected in 16 over 37 genomes (Table 1B). For G3.4,
303 the HN, P and M genes carried the greatest substitutions rate (2, 1.7 and 1.2 substitutions per
304 100 nucleotides). There's no insertion, frameshift or deletion present in the two G3.4's
305 genomes. Two specific insertion to G3.4 : 14532:T and 8219:AA each present in one of the
306 two G3.4 genomes. There were no frame shifts nor deletion specific to G3.4. Then for G3.7,
307 P, HN and F with 5.7, 4.6 and 3.8 substitutions per 100 nucleotides are the gens with the
308 higher specific substitutions rate. There's no insertion, frameshift or deletion detected in the 9
309 G.7's genotypes. There's 34 insertions present in only genome. Three genomes carry 20 over
310 34 insertions, one genome carry 8 insertions, another one carry 7 and the third carry 5. Seven
311 deletions are detected in only one genotype each, four of these deletions were detected in the
312 same genome.
313 As the substitution rate for amino acids, for G1a it's the HN, P and M genes with the
314 higher substitution rate with respectively 47.5 amino acids substitution per 100 amino acids,
315 40.5, and 20.5. For G1a.5 it's P (9.1), HN (8.8) and NP (4.8) when it's HN (31), P (24.1) and
316 M (21). There's 876 specific amino acids substitution over 1109 total substitutions for G1a,
317 73 over 186 for G1a.5 and 572 over 685 for G1a.8. Regarding genotype-specific amino acids
318 substitutions, for G1a it's the genes encoding for HN (39.8), P (33.4) and M (24.4). Then for
319 the subgenotype G1a.5 it's P (3.8), HN (3.7) and NP (1.8). For G1a.8 it's HN (25.9), M (19.6)
320 and P (18.7).
321 For the other genotype, G3, it's the genes encoding for HN (39.6), P (39.5) and F
322 (29.9). For the subgenotypes it's for G3.4 HN (3.5), P (3.3) and F (2.4), for G3.7 it's P (12.9),
323 HN (8.2) and F (6.7). There's 901 over 1130 specific amino acids substitutions for G3, 77

over 83 for G3.4 and 246 over 252 for G3.7. The genes with the higher specific amino acids substitutions rate for G3 are P (32.4), HN (31.9) and F (24.9). For G3.4 it's HN (3.5), P (3) and F (2) then for G3.7 it's P (12.7), HN (8.2) and F (6.4). There's no signature specific amino acids insertion or deletion observed in these genomes.

Temporal evolution of the incidence of HPIV2 genotypes and subgenotypes

Genotypes G1a and G3 were detected concurrently over years (Figure 3). Subgenotypes G1a.5 or G3.4/5 were detected before subgenotypes G1a.8 or G3.7, respectively. The peak of subgenotype G1a.5 incidence occurred during fall 2019, while that of subgenotype G1a.8 occurred during summer and early fall 2022. The peak of G3.4/5 incidence was not clearly delineated due to the scarcity of genomes studied but might be during late summer and fall 2019 or in August 2022. Regarding subgenotype G3.7, it was only detected between June to October 2022 and its peak of incidence occurred during summer 2022 (Supplementary Figure S1). Due to the limited number of HPIV2 genomes obtained here and their majority obtaining from nasopharyngeal samples collected in 2022, we could not draw reliable interpretations regarding the accumulation of mutations during the study period for a given lineage.

Comparison between genetic diversity of HPIV2 genomes obtained in the present study or available in GenBank

The set of 77 genomes that cover ≥90% of NC_003443.1 and were obtained here from samples collected between 2017 and 2022 tended to exhibit a greater diversity than the set of 71 GenBank genomes that originate from samples collected between 2002 and 2021 (656.3± 129.5 mutations on average compared to 639.4± 163.2; p= 0.4908). To limit the possible bias associated with samples' sizes, origins, and age, we considered the largest set of GenBank genomes available from a same country during a same year, which comprised ten GenBank genomes originating from Seattle, USA, in 2010. This dataset was compared to three sets of

10 genomes randomly selected from the set of 77 genomes obtained here. The intra-sample diversity for the 2010 Seattle set was significantly lower compared to the intra-sample diversity for the other sets (99.8 ± 0.1% on average percent of similarity compared to 95.7 ± 3.0%; $p < 0.0000001$) (Figure 5C-D).

Potential HPIV2 genome recombinations

A total of 204 recombination events (REs) were initially identified but only 19 REs remained after applying the quality criteria, as determined as five positive tests with a p.value threshold of 0.01 or less out of seven (Figure 4; Table 2). Recombinant sequences were 4.67 times more frequently of genotype G1a than G3 (14 vs. 5). Locations of breakpoints were found throughout the whole genome, without any hotspots. Eight potential recombination scenarios were detected in which parental genomes were detected before the recombinant genome (Table 3). Five involved genotype G1a and three involved genotype G3. Breakpoints were scattered in the six HPIV2 genes. The length of these recombination areas ranged from 316 to 1,497 nucleotides, with a mean ± standard deviation of 660 ± 372 nucleotides.

DISCUSSION

The literature on HPIV2 is scarce, particularly that describing genomes or sequences of this virus species. The present study approximately doubles the number of complete HPIV2 genomes recorded in GenBank. Indeed, according to GenBank (24), as of February 2024, 75 nucleotide sequences were labeled as complete genomes and covered at least 90% of genome NC_003443.1. Here, we provided 77 new genomes, and the first from France. In addition, our work also approximately tripled the number of sequences larger than 10 kilobases with 239 new sequences. Moreover, this work expands our knowledge of the HPIV2 genetic diversity, contributes to the characterization of HPIV2 genotypes, and describes for the first time putative recombinations between HPIV2 genomes.Beyond, our study provides an updated

374 strategy to obtain new genomes more efficiently with a one-step multiplex PCR system
375 specific to HPIV2. A previous work by Santak et al the reported a PCR-based procedure for
376 the whole-genome amplification that was based on only 17 genomes, which could not take
377 sufficiently broadly into account the diversity of HPIV2, and that did not include any
378 degenerated nucleotide (Maja Santak et al., 2018, p. 2).
379 Santak et al's 2016 and 2018 studies (Maja Santak et al., 2018, p. 2; M. Santak et al.,
380 2018; Santak et al., 2016) described 26 G1a genomes and 19 G3 genomes, while we add 129
381 G1a genomes and 110 G3 genomes. In their 2016 study, Santak et al delineated 6 HPIV2
382 genotypes or subgenotypes, namely G1a, G1b, G2, G3, G4a and G4b. This completed a
383 phylogenetic analysis reported in 2012 by Almajhdi et al that described four distinct clusters
384 (Almajhdi et al., 2012). Earlier, in 2008, Terrier et al (Terrier et al., 2008) had reported the
385 occurrence of natural HPIV2 variants. Santak et al allowed an in-depth analysis of the two
386 major genotypes, G1a and G3, based on an analysis of the entire genome and not only on the
387 HN gene. Based on the HN gene, Feng et al proposed in 2023 (Feng et al., 2023) the
388 delineation of five lineages and four sub-lineages within two major clusters. These two major
389 genotypes were consistently delineated phylogenetically based on the HN gene.
390 In a previous study, substitution of glutamine at position 202 of the NP, which encapsidates
391 the viral RNA, by other amino acids was reported to increase by 30-fold the activity of the
392 virus RNA polymerase (Matsumoto et al., 2017) Q202 was reported to be the single amino
393 acid in the NP RNA binding groove that directly contacts a nucleotide. Here, we detected
394 Q202T in one of 193 samples for which the codon was available. Saka et al further reported
395 that Q202A substitution in NP induced strong replication but prevented budding, and that its
396 effect was reported to be compensated by substitution I35L in a mini-replicon assay (Saka et
397 al., 2022). We detected the I35L substitution in none of the samples. Also, Matsumoto et al
398 reported that substitution of Y260 located in the RNA binding pocket of the NP dramatically

reduced polymerase activity in a mini-replicon assay (Matsumoto et al., 2019). Here, we detected the Y260D substitution in 1 of 77 samples which is a G3 genotype. Otherwise, Nishio et al (2006) (Nishio et al., 2006) reported that substitutions of some basic amino acids by alanines in the P protein specific region (amino acids 249-354) or V protein specific region (176-225), which are required for binding to viral RNA, were associated with RNA-binding abolition. Here, we detected no substitutions from basic amino acids to alanine in any sample.

The present study allowed characterizing HPIV2 diversity at the genome scale and over a 5 year-period. We described the occurrence of lineage-specific mutations, which are 'signatures' of these lineages, and the accumulation of other mutations within each lineages over time at substantial rates, with variations according to the viral gene, as for the case of other RNA viruses including SARS-CoV-2 (Carabelli et al., 2023). The detection of the major genotype G1a between 2017 and 2020 is in line with the observation by Santak et al. in 2018 of a change in the prevalence of HPIV2 genotypes. A change occurred again in 2021 and during early 2022 with G3 becoming the dominant genotype again, before reaching a quasi-balance in genotype prevalence in 2022. The more recent detection of subgenotypes G1a.8 and G3.7 in our study can be explained by genome evolution over time. For the first time, potential recombination scenarios were detected here for HPIV2. It is worthy to note that recombination in negative single-stranded RNA viruses has been described as less frequent than in other viruses (McVean et al., 2002; Patiño-Galindo et al., 2021).

CONCLUSION

Overall, the present study highlights the importance of expanding the activity of whole genome sequencing and analyses for HPIV2. There is currently a tremendous gap between the number of genomes available worldwide for HPIV2 and for other respiratory viruses such as SARS-CoV-2 or influenza virus. Further works are needed to obtain more precise pictures of

HPIV2 classification and of the diversity over time and the evolutionary dynamic of these respiratory viruses. Moreover, the availability of a broader and updated set of complete genomes would be helpful to investigate the potential impact of detected mutations on virus replication, transmission, neutralization and pathogenicity, and to check the adequacy of current universal PCR systems used for routine HPIV2 clinical diagnosis. Complementary studies aiming to investigate the effect of mutations could involve protein structural prediction and culture assays.

Acknowledgments

We are very grateful to the technical team of the next-generation sequencing platform of IHU Méditerranée Infection.

Data Availability

HPIV2 genomes analyzed here have been submitted to the GenBank sequence database (https://www.ncbi.nlm.nih.gov/genbank/; Sayers et al, 2023); (GenBank Accession no: XXXXX-XXXXX).

Author contributions

Conceived and designed the experiments: BLS, PC. Contributed materials, analysis tools: All authors. Analyzed the data: EB, BLS, and PC. Writing—original draft preparation: EB, BLS and PC. Writing—review and editing: All authors. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

LLT receives a grant from the BioSellal company. The other authors have no conflicts of interest to declare. Funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or

approval of the manuscript.

Funding

This work was supported by the French Government under the "Investments for the Future" program managed by the National Agency for Research (ANR), Méditerranée-Infection 10-IAHU-03; and by the French Ministry of Higher Education, Research and Innovation (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation) and the French Ministry of Solidarity and Health (Ministère des Solidarités et de la Santé).

Ethics

The present study regards data that have been registered on the Health Data Access Portal of AP-HM under No. PADS24-190 which has been approved by the Ethics and Scientific Committee of AP-HM.

FIGURE LEGENDS

Figure 1. Genome-based phylogenetic trees with sequences obtained here or taken from the NCBI GenBank sequence database (Sayers et al., 2023).
Trees were generated by Nextstrain (19; https://clades.nextstrain.org). A: Genomes of G1a genotype and subgenotypes. B: Genomes of G3 genotype and subgenotypes. C: Genomes of subgenotypes G1a.5, G1a.8, G3.4/5, and G3.7 obtained here. The X axis is the divergence in nucleotide and the Y axis is for representation only of lineage.

Figure 2. Number of genotype- and subgenotypes-specific substitutions per 100 nucleotides per gene.
A comparison of two substitutions rate between all genes was done with the "Compare 2 Rates" tool of OpenEpi (https://www.openepi.com/PersonTime2/PersonTime2.htm). The bold line represent the comparison of the HN gene to the others, the dotted line represent the comparison of the P gene compared to the others. P-value : * < 0.05 ; ** < 0.01 ; *** < 0.001 ; **** < 0.0001. NP: Nucleocapsid Protein ; P: Phospho-protein ; M: Matrix protein ; F: Fusion protein; HN: Hemagglutinin-Neuraminidase ; L: Large protein.

Figure 3. Chronological variation of detected HPIV2 among IHU Méditerranée Infection.
HPIV2 occurrence of genotypes and incidence of HPIV2 diagnosis (A). Number of genomes per genotype or sub-genotype per month among genomes obtain here with the two main genotypes G1a and G3 (B), and a quantification of the evolution of the two major HPIV2's genotype (C).

Figure 4. Recombinant analysis with RDP4.

Selection of the most reliable recombination event (R.E).

Figure 5. Genetic diversity among HPIV2 GenBank's genomes or IHU's genome.
A scatter plot representation of the intra-sample average percent of similarity obtain with a distance matrix *via* BioEdit. A-B : For the complete sequences set of GenBank (70 sequences) and the genomes obtained here labelled as "IHU" (77 sequences) . C-D : For three set of 10 sequences obtained here in 2022 and 10 sequences obtained at Seattle in 2010.

REFERENCES

Aksamentov, I., Roemer, C., Hodcroft, E.B., Neher, R.A., 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software 6, 3773. https://doi.org/10.21105/joss.03773

Almajhdi, F.N., Alshaman, M.S., Amer, H.M., 2012. Human parainfluenza virus type 2 hemagglutinin-neuramindase gene: sequence and phylogenetic analysis of the Saudi strain Riyadh 105/2009. Virol J 9, 316. https://doi.org/10.1186/1743-422X-9-316

Andrewes, C.H., Bang, F.B., Chanock, R.M., Zhdanov, V.M., 1959. Para-influenza viruses 1, 2, and 3: suggested names for recently described myxoviruses. Virology 8, 129–130. https://doi.org/10.1016/0042-6822(59)90027-3

Brnić, D., Vlahović, D., Gudan Kurilj, A., Maltar-Strmečki, N., Lojkić, I., Kunić, V., JemerSić, L., Bačani, I., Kompes, G., Beck, R., Mikuletič, T., Steyer, A., 2023. The impact and complete genome characterisation of viruses involved in outbreaks of gastroenteritis in a farrow-to-finish holding. Sci Rep 13, 18780. https://doi.org/10.1038/s41598-023-45994-4

Carabelli, A.M., Peacock, T.P., Thorne, L.G., Harvey, W.T., Hughes, J., COVID-19 Genomics UK Consortium, Peacock, S.J., Barclay, W.S., de Silva, T.I., Towers, G.J., Robertson, D.L., 2023. SARS-CoV-2 variant biology: immune escape, transmission and fitness. Nat Rev Microbiol 21, 162–177. https://doi.org/10.1038/s41579-022-00841-7

Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N., Caskey, C.T., 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 16, 11141–11156.

Chanock, R.M., 1956. Association of a new type of cytopathogenic myxovirus with infantile croup. Journal of Experimental Medicine 104, 555–576. https://doi.org/10.1084/jem.104.4.555

Colson, P., Chaudet, H., Delerce, J., Pontarotti, P., Levasseur, A., Fantini, J., La Scola, B., Devaux, C., Raoult, D., 2024. Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic. J Infect 88, 106150. https://doi.org/10.1016/j.jinf.2024.106150

Dagona, A.G., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series.

Dong, P.-P., Wang, R.-R., Abduriyim, S., 2023. Diversity and evolution of the MHC class II DRB gene in the Capra sibirica experienced a demographic fluctuation in China. Sci Rep 13, 19352. https://doi.org/10.1038/s41598-023-46717-5

Feng, Y., Zhu, Z., Xu, J., Sun, L., Zhang, H., Xu, H., Zhang, F., Wang, W., Han, G., Jiang, J., Liu, Y., Zhou, S., Zhang, Y., Ji, Y., Mao, N., Xu, W., 2023. Molecular Evolution of Human Parainfluenza Virus Type 2 Based on Hemagglutinin-Neuraminidase Gene. Microbiol Spectr 11, e04537-22. https://doi.org/10.1128/spectrum.04537-22

Gámbaro, F., Pérez, A.B., Prot, M., Agüera, E., Baidaliuk, A., Sánchez-Seco, M.P., Martínez-Martínez, L., Vázquez, A., Fernandez-Garcia, M.D., Simon-Loriere, E., 2023. Untargeted metagenomic sequencing identifies Toscana virus in patients with idiopathic meningitis, southern Spain, 2015 to 2019. Euro Surveill 28, 2200913. https://doi.org/10.2807/1560-7917.ES.2023.28.45.2200913

He, W.-T., Hou, X., Zhao, J., Sun, J., He, H., Si, W., Wang, J., Jiang, Z., Yan, Z., Xing, G., Lu, M., Suchard, M.A., Ji, X., Gong, W., He, B., Li, J., Lemey, P., Guo, D., Tu, C., Holmes, E.C., Shi, M., Su, S., 2022. Virome characterization of game animals in China reveals a spectrum of emerging pathogens. Cell 185, 1117-1129.e8. https://doi.org/10.1016/j.cell.2022.02.014

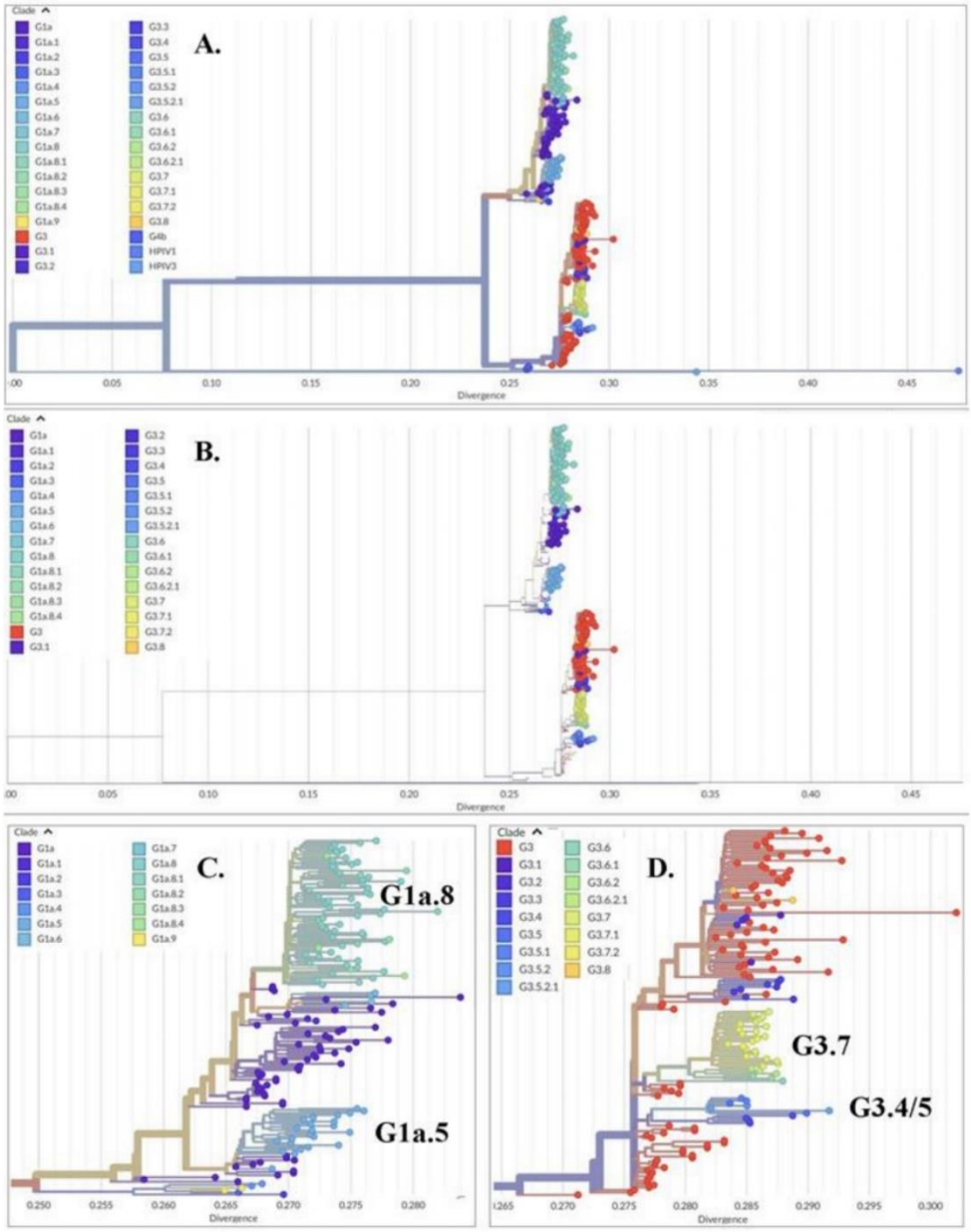
544 Hilt, E.E., Ferrieri, P., 2022. Next Generation and Other Sequencing Technologies in
545 Diagnostic Microbiology and Infectious Diseases. Genes (Basel) 13, 1566.
546 https://doi.org/10.3390/genes13091566
547 Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid
548 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research
549 30, 3059–3066. https://doi.org/10.1093/nar/gkf436
550 Kawano, M., Bando, H., Yuasa, T., Kondo, K., Tsurudome, M., Komada, H., Nishio, M., Ito,
551 Y., 1990. Sequence determination of the hemagglutinin-neuraminidase (HN) gene of
552 human parainfluenza type 2 virus and the construction of a phylogenetic tree for HN
553 proteins of all the paramyxoviruses that are infectious to humans. Virology 174, 308–
554 313. https://doi.org/10.1016/0042-6822(90)90081-2
555 Lambrechts, L., Bonine, N., Verstraeten, R., Pardons, M., Noppe, Y., Rutsaert, S.,
556 Van Nieuwerburgh, F., Van Criekinge, W., Cole, B., Vandekerckhove, L., 2023. HIV-
557 PULSE: a long-read sequencing assay for high-throughput near full-length HIV-1
558 proviral genome characterization. Nucleic Acids Res 51, e102.
559 https://doi.org/10.1093/nar/gkad790
560 Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and
561 analysis of recombination patterns in virus genomes. Virus Evolution 1, vev003.
562 https://doi.org/10.1093/ve/vev003
563 Matsumoto, Y., Ohta, K., Kolakofsky, D., Nishio, M., 2017. A Point Mutation in the RNA-
564 Binding Domain of Human Parainfluenza Virus Type 2 Nucleoprotein Elicits
565 Abnormally Enhanced Polymerase Activity. J Virol 91, e02203-16.
566 https://doi.org/10.1128/JVI.02203-16
567 Matsumoto, Y., Ohta, K., Nishio, M., 2019. Importance of tyrosine in the RNA-binding
568 domain of human parainfluenza virus type 2 nucleoprotein for polymerase activity.
569 Arch Virol 164, 1851–1855. https://doi.org/10.1007/s00705-019-04240-x
570 McVean, G., Awadalla, P., Fearnhead, P., 2002. A Coalescent-Based Method for Detecting and
571 Estimating Recombination From Gene Sequences. Genetics 160, 1231–1241.
572 https://doi.org/10.1093/genetics/160.3.1231
573 Mencattelli, G., Ndione, M.H.D., Silverj, A., Diagne, M.M., Curini, V., Teodori, L., Di
574 Domenico, M., Mbaye, R., Leone, A., Marcacci, M., Gaye, A., Ndiaye, E., Diallo, D.,
575 Ancora, M., Secondini, B., Di Lollo, V., Mangone, I., Bucciacchio, A., Polci, A.,
576 Marini, G., Rosà, R., Segata, N., Fall, G., Cammà, C., Monaco, F., Diallo, M., Rota-
577 Stabelli, O., Faye, O., Rizzoli, A., Savini, G., 2023. Spatial and temporal dynamics of
578 West Nile virus between Africa and Europe. Nat Commun 14, 6440.
579 https://doi.org/10.1038/s41467-023-42185-7
580 nextstrain/nextclade, 2024.
581 Nicolle, C., Lebailly, C., 1918. Quelques notions expérimentales sur le virus de la grippe. CR
582 Acad Sci Paris 167, 607–10.
583 Nishio, M., Tsurudome, M., Ito, M., Ito, Y., 2006. Identification of RNA-binding regions on
584 the P and V proteins of human parainfluenza virus type 2. Med Microbiol Immunol
585 195, 29–36. https://doi.org/10.1007/s00430-005-0244-7
586 Papa Mze, N., Kacel, I., Beye, M., Tola, R., Sarr, M., Basco, L., Bogreau, H., Colson, P.,
587 Fournier, P.-E., 2023. High Throughput SARS-CoV-2 Genome Sequencing from 384
588 Respiratory Samples Using the Illumina COVIDSeq Protocol. Genes (Basel) 14, 681.
589 https://doi.org/10.3390/genes14030681
590 Patiño-Galindo, J.Á., Filip, I., Rabadan, R., 2021. Global Patterns of Recombination across
591 Human Viruses. Molecular Biology and Evolution 38, 2520–2531.
592 https://doi.org/10.1093/molbev/msab046

593 Petrone, M., 2020. Virus genomics as a clinical and epidemiological tool. EBioMedicine 62,
594 103141. https://doi.org/10.1016/j.ebiom.2020.103141
595 Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar,
596 S., Zaremba, S., Gu, Z., Zhou, L., Larson, C.N., Dietrich, J., Klem, E.B.,
597 Scheuermann, R.H., 2012. ViPR: an open bioinformatics database and analysis
598 resource for virology research. Nucleic Acids Res 40, D593-598.
599 https://doi.org/10.1093/nar/gkr859
600 Quick, J., 2020. nCoV-2019 sequencing protocol.
601 Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira,
602 G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-Ximenez,
603 L.L., de Jesus, J.G., Giovanetti, M., Hill, S., Black, A., Bedford, T., Carroll, M.W.,
604 Nunes, M., Alcantara, L.C., Sabino, E.C., Baylis, S.A., Faria, N., Loose, M., Simpson,
605 J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2017. Multiplex PCR method for
606 MinION and Illumina sequencing of Zika and other virus genomes directly from
607 clinical samples. Nat Protoc 12, 1261–1276. https://doi.org/10.1038/nprot.2017.066
608 Saka, N., Matsumoto, Y., Ohta, K., Kolakofsky, D., Nishio, M., 2022. A Point Mutation in the
609 Human Parainfluenza Virus Type 2 Nucleoprotein Leads to Two Separate Effects on
610 Virus Replication. J Virol 96, e0206721. https://doi.org/10.1128/JVI.02067-21
611 Santak, M., Lang Balija, M., Mlinarić Galinović, G., Ljubin Sternak, S., Vilibić-Čavlek, T.,
612 Tabain, I., 2018. Genotype replacement of the human parainfluenza virus type 2 in
613 Croatia between 2011 and 2017 – the role of neutralising antibodies. Epidemiol Infect
614 146, 1372–1383. https://doi.org/10.1017/S0950268818001693
615 Santak, Maja, Mlinarić-Galinović, G., Vilibić-Čavlek, T., Tabain, I., 2018. Comparative
616 genomics of human rubulavirus 2. Arch Virol 163, 3141–3148.
617 https://doi.org/10.1007/s00705-018-3986-4
618 Santak, M., Slović, A., Ljubin-Sternak, S., Mlinarić Galinović, G., Forčić, D., 2016. Genetic
619 diversity among human parainfluenza virus type 2 isolated in Croatia between 2011
620 and 2014. Journal of Medical Virology 88, 1733–1741.
621 https://doi.org/10.1002/jmv.24532
622 Sayers, E.W., Cavanaugh, M., Clark, K., Pruitt, K.D., Sherry, S.T., Yankie, L., Karsch-
623 Mizrachi, I., 2023. GenBank 2023 update. Nucleic Acids Res 51, D141–D144.
624 https://doi.org/10.1093/nar/gkac1012
625 Terrier, O., Cartet, G., Ferraris, O., Morfin, F., Thouvenot, D., Hong, S.S., Lina, B., 2008.
626 Characterization of naturally occurring parainfluenza virus type 2 (hPIV-2) variants. J
627 Clin Virol 43, 86–92. https://doi.org/10.1016/j.jcv.2008.05.007
628 Tsurudome, M., Nishio, M., Komada, H., Bando, H., Ito, Y., 1989. Extensive antigenic
629 diversity among human parainfluenza type 2 virus isolates and immunological
630 relationships among paramyxoviruses revealed by monoclonal antibodies. Virology
631 171, 38–48. https://doi.org/10.1016/0042-6822(89)90508-4
632 Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-
633 Zerbini, P., Dempsey, D.M., Dutilh, B.E., García, M.L., Curtis Hendrickson, R.,
634 Junglen, S., Krupovic, M., Kuhn, J.H., Lambert, A.J., Łobocka, M., Oksanen, H.M.,
635 Orton, R.J., Robertson, D.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith,
636 D.B., Suzuki, N., Van Doorslaer, K., Vandamme, A.-M., Varsani, A., Zerbini, F.M.,
637 2022. Recent changes to virus taxonomy ratified by the International Committee on
638 Taxonomy of Viruses (2022). Arch Virol 167, 2429–2440.
639 https://doi.org/10.1007/s00705-022-05516-5
640

641

642 **FIGURES**

643 **Figure 1.**



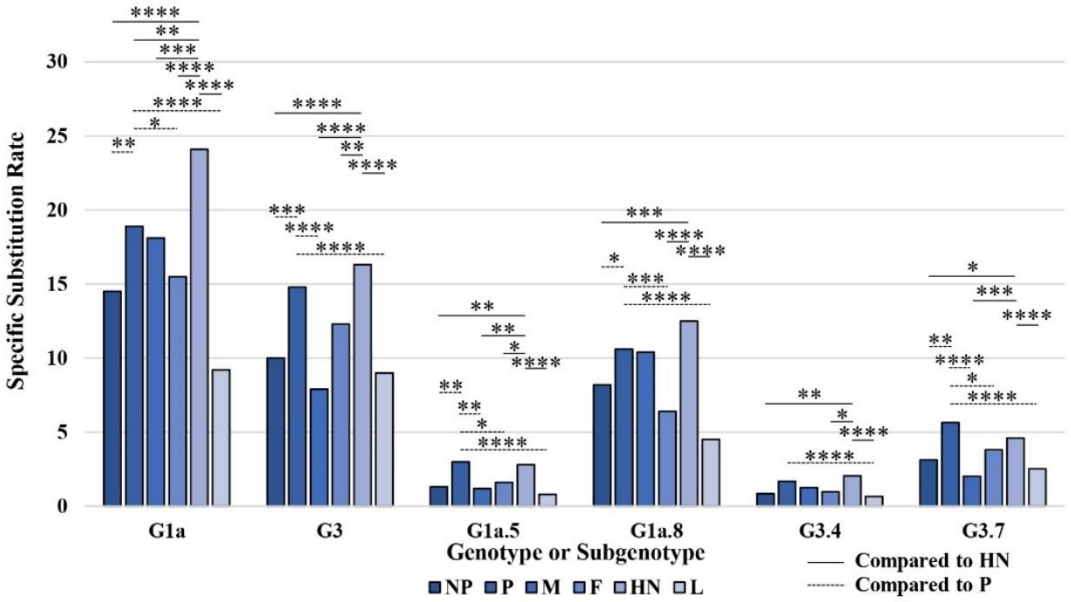
644

645

646

647 **Figure 2.**

648



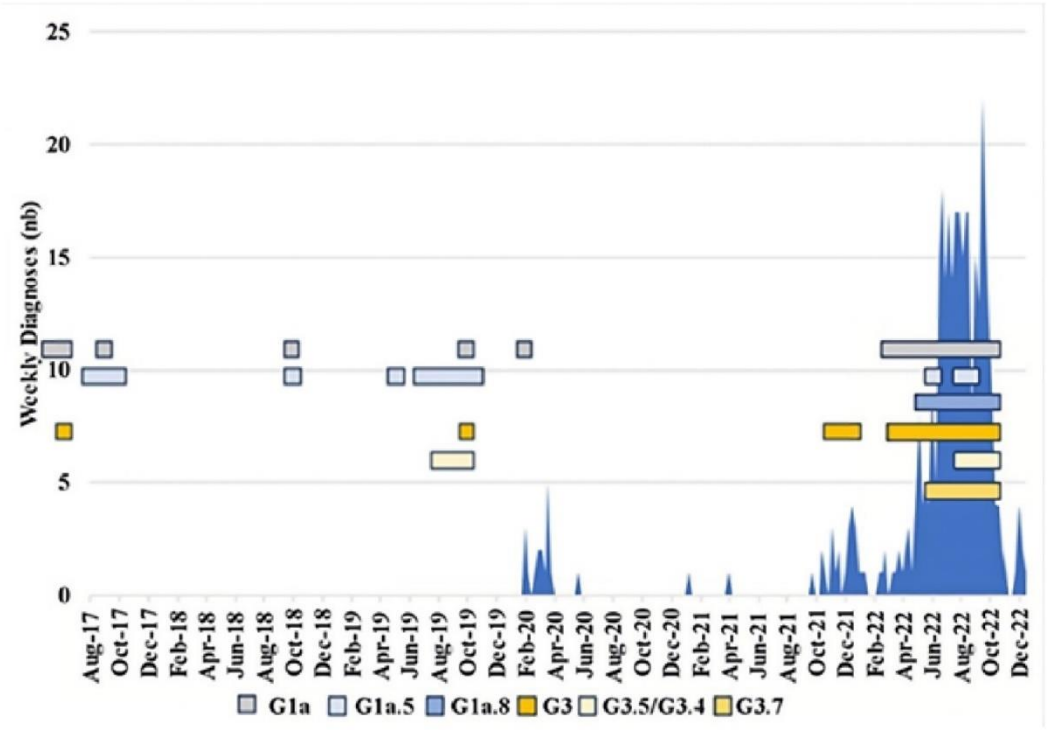
649

650

651 **Figure 3.**

652

653 **A.**

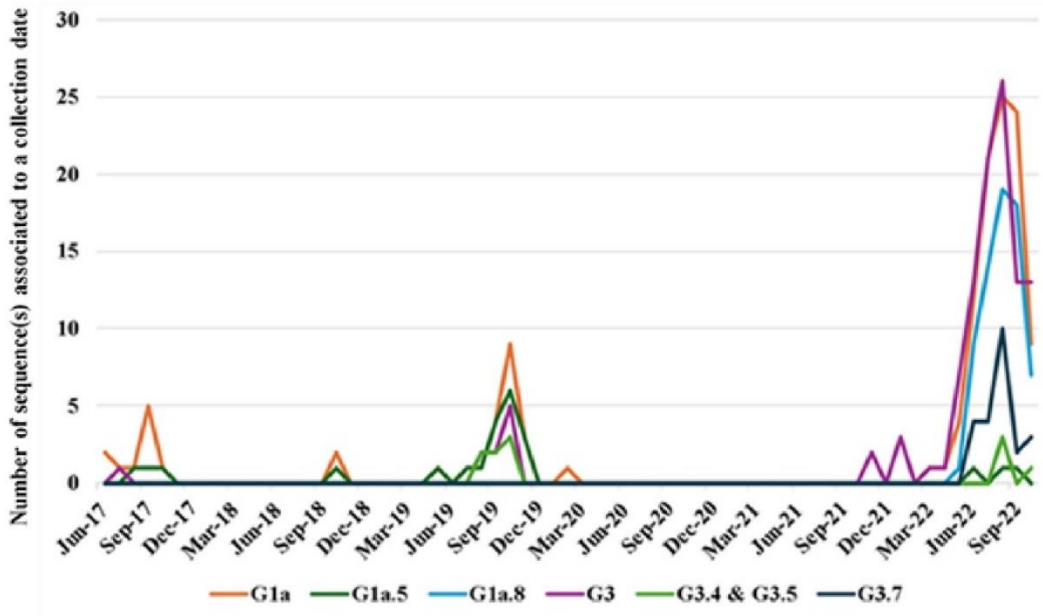


654

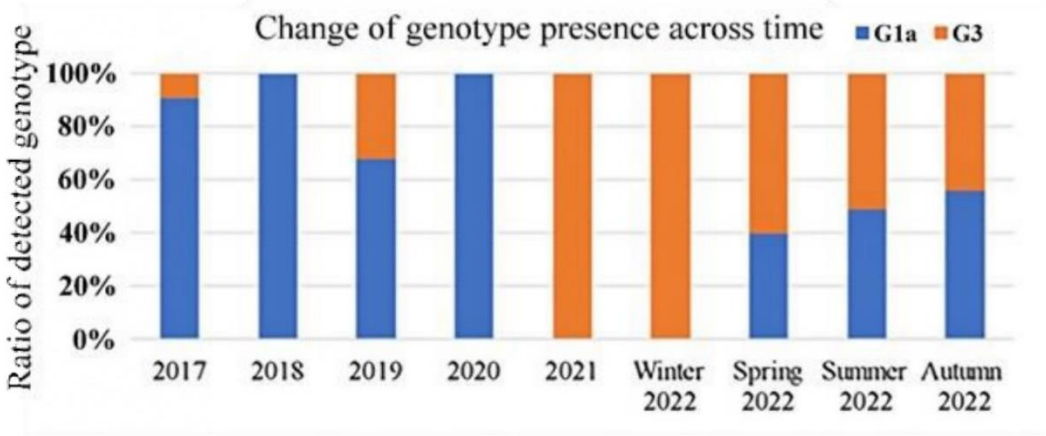
655

656

B.


Number of sequence(s) associated to a collection date
30
25
20
15
10
5
0
Jun-17
Sep-17
Dec-17
Mar-18
Jun-18
Sep-18
Dec-18
Mar-19
Jun-19
Sep-19
Dec-19
Mar-20
Jun-20
Sep-20
Dec-20
Mar-21
Jun-21
Sep-21
Dec-21
Mar-22
Jun-22
Sep-22
G1a
G1a.5
G1a.8
G3
G3.4 & G3.5
G3.7

660 **C.**

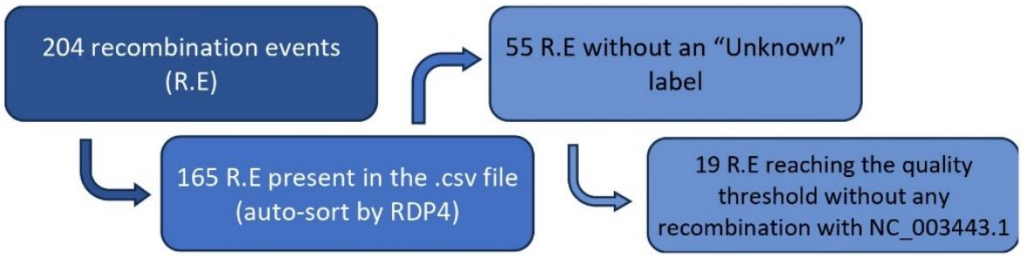

Change of genotype presence across time
G1a
G3
Ratio of detected genotype
100%
80%
60%
40%
20%
0%
2017
2018
2019
2020
2021
Winter 2022
Spring 2022
Summer 2022
Autumn 2022

661

662

663 **Figure 4.**

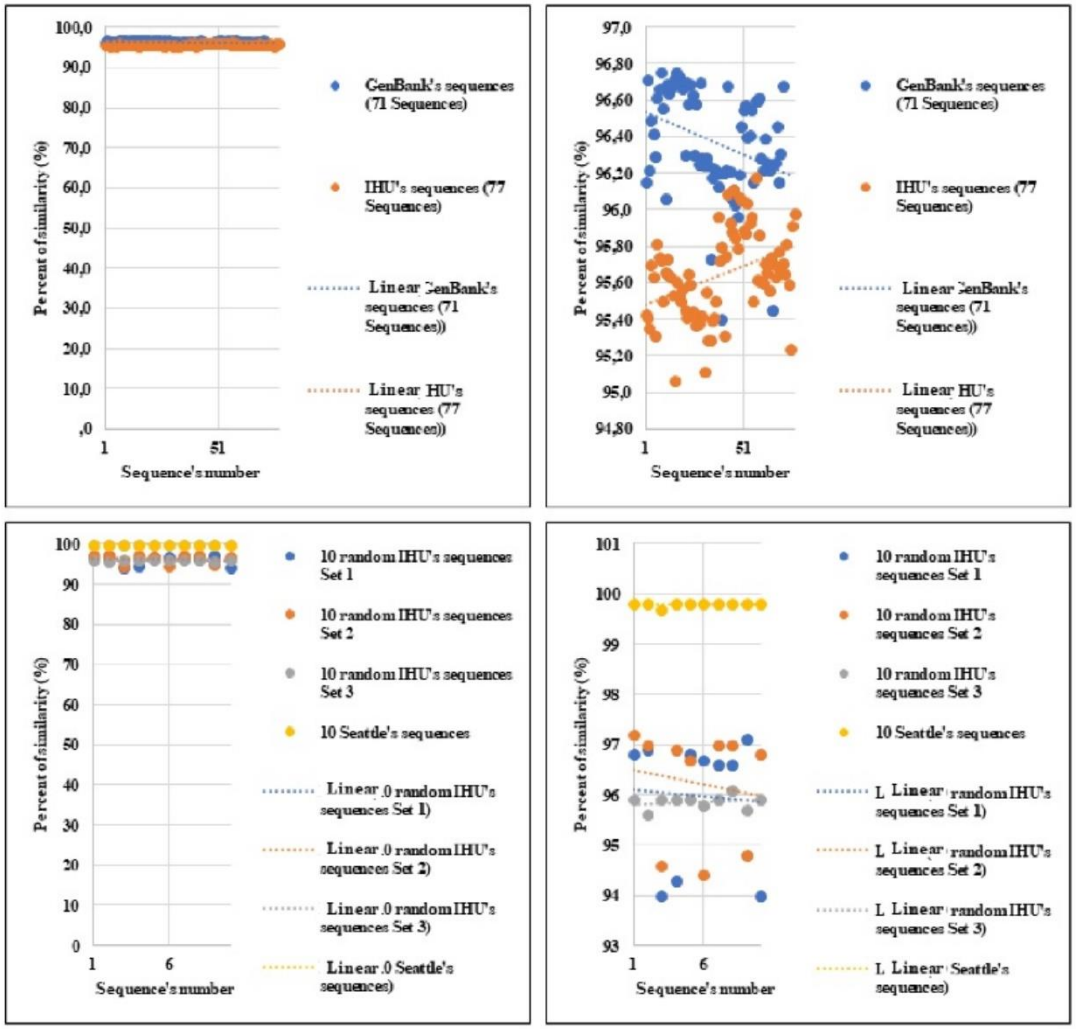
664



665

666

667 **Figure 5.**



668

669 # TABLES

670

671 **Table 1.** Identification of the most prevalent specific substitutions in each genotype.
672 Summary of the three, or more, most prevalent substitution specific to a given genotype or a
673 sub-genotype (A). Summary of the most prevalent mutation, other than substitution, with their
674 number of occurrence per genotype and sub-genotype (B). NP: Nucleocapsid Protein ; P:
675 Phospho-protein ; M: Matrix protein ; F: Fusion protein; HN: Hemagglutinin-Neuraminidase ;
676 L: Large protein ; Ins = Insertion ; Del = Deletion ; FS = Frame Shift.

677

678

679 **A.**

	G1a and G3 compared		G1a.5 and G1a.8 compared		G3.4/5 and G3.7 compared	
	G1a	**G3**	**G1a.5**	**G1a.8**	**G3.4/5**	**G3.7**
NP	G1422A (97.7%) A1638G (96.1%) A1364G (94.6%)	C1379T (28.2%) A1502G (26.4%) A1513T (26.4%) T1623C (26.4%)	A845G (97.1%) G1423A (97.1%) A1148G (94.1%)	T954C (100%) C1383T (91.3%) G704A (82.6%)	A1139G (100%) T1391C (100%) T1489C (100%)	C1501T (100%) A1502G (100%) A1513T (100%) T1553C (100%) T1623C (100%)
P	C2540T (90.7%) C2655T (90.7%) A2009G (89.2%)	A2669C (48.2%) T3009C (47.3%) G2072A (26.4%)	A2330G (94.1%) G2342T (94.1%) T2721C (92.7%) G2790A (92.7%)	C2081T (69.6%) T2925C (69.6%) C2754T (65.2%)	A2904G (90.9%) A2195G (45.5%) C2225A (45.5%) T2637C (45.5%)	C2137T (100%) A2127G (95.7%) C2477T (95.7%)
M	C3468T (82.2%) G4512A (80.6%) T3711C (78.3%)	A4443G (56.4%) C4437T (54.6%) C3915T (37.3%)	A3492G (82.4%) C3726T (77.9%) C4518A (77.9%)	C4453A (87%) A4482G (87%) T3531C (78.3%)	T3453C (81.8%) T4392C (54.6%) A4152G (36.4%) C4410T (36.4%) C4440A (36.4%)	A3549G (82.6%) G3567A (78.3%) T3439C (8.7%)
F	G5426A (96.9%) A5498C (96.9%) C5108T (96.1%) C5468T (96.1%) C5711T (96.1%) T5849C (96.1%)	A5192G (61.8%) A5668G (59.1%) C5081T (49.1%)	A5345G (97.1%) T5861A (97.1%) T6050C (95.6%)	T5249C (100%) C5399T (100%) T5528C (100%)	A5267G (100%) T5261C (63.6%) G5369A (63.6%) C5702A (63.6%) A5815G (63.6%)	C5120T (95.7%) G5693T (95.7%) T5786C (95.7%) C6068T (95.7%) T6095G (95.7%) T6130A (95.7%)
HN	G7255T (97.7%) A7553C (96.9%) A6857G (96.1%) C7022T (96.1%) T7052C (96.1%) T7190A (96.1%) C7363G (96.1%) C7463T (96.1%) C8443T (96.1%)	G8171A (99.1%) A7430G (61.8%) G8144A (52.7%) A8286G (52.7%)	T7283C (98.5%) G7460A (98.5%) C8166A (98.5%) T8185C (98.5%)	C7010T (100%) T7079C (100%) T7358C (100%) A7493T (100%)	G8181A (100%) C8411T (100%) T7784C (90.9%)	G7019A (100%) G7058A (100%) A7562G (100%) T8167C (100%) C8170T (100%) A8237G (100%) G8324A (100%)
L	T14612A (100%) T11981A (99.2%) T12374C (99.2%) T12494A (99.2%) A12495G (99.2%) A12797G (99.2%) T13211C (99.2%) A13679G (99.2%) A14800G (99.2%) T14885C (99.2%)	A14057G (93.6%) A10367G (76.4%) C13128T (62.7%)	G11948A (100%) A12080C (100%) A13193G (100%) T14582G (100%)	T14582A (100%) C9555T (95.7%) C10055T (95.7%) T10823C (95.7%) A11726G (95.7%) C13304T (95.7%) G15215A (95.7%)	A10445G (100%) T10601C (100%) G9794A (90.9%) A11111G (90.9%) C13502A (90.9%)	A9173G (100%) A9221G (100%) T9809C (100%) T11693G (100%) T12312C (100%) T13721C (100%) C13754A (100%) C14630T (100%) A14732G (100%) A15029G (100%) G15098A (100%)

680

681

B.

Variation	Total	G1a	G3	G1a.5	G1a.8	G3.4/5	G3.7
Ins : 3376:C	33	33	0	2	30	0	0
Ins : 2730:G	12	6	6	1	3	0	0
Ins : 2733:G	10	4	6	1	3	0	0
Del : 14446	37	37	0	11	13	0	0
Del : 6636	33	0	33	0	0	3	4
FS : L:1917-2229	14	14	0	4	6	0	0
FS : F:61-552	13	10	3	1	5	0	1
FS : F:474-552	13	6	7	2	2	1	1
FS : P:113-233	12	6	6	1	5	0	0

691 **Table 2.** Selection of recombination events temporally plausible.

692

Recombinant Sequence (Genotype)	Minor Parental Sequence (Genotype)	Major Parental Sequence (Genotype)	Recombination size (nt) / Region according to the NC_03443.1
512663666_KcBLu (G3)	511754512_kQQft (G1a.8)	577683690_UDllD (G3)	13978 - 14560 : 582 nt / Gene L
671122031_SgbRw (G1a)	662085176_UxsGw (G3.7)	301652630_wmIRF (G1a)	5508 - 6320 : 812 nt / Gene F
761032497_GAVVi (G1a.8.1)	851673922_TiEtE (G3.6)	881293965_Dhaaf (G1a)	708 - 1024 : 316 nt / Gene NP
845458065_oWOff (G3.3)	810158120_fwCeu (G1a)	261791997_JKnBe (G3.8)	2370 - 3867 : 1497 nt / Between Gene V and Gene M
530860623_QIeTf (G3.7)	574796828_nnrtw (G1a.8)	367977552_gnRjQ (G3.7)	6482 - 6896 : 414 nt / Between Gene F and Gene HN
516609874_jwZSj (G1a.8)	577683690_UDllD (G3)	545927507_CtQrR (G1a.8)	6425 - 6859 : 434 nt / Between Gene F and Gene HN
687194775_zrxbh (G1a.8)	809605673_OXFRv (G3)	423507134_HEFrl (G1a.8)	1 - 658 : 657 nt / Gene NP
158478296_TZkhZ (G1a.8)	512663666_KcBLu (G3)	245275920_dQIQq (G1a.8)	4156 - 4722 : 566 nt / Gene M

693

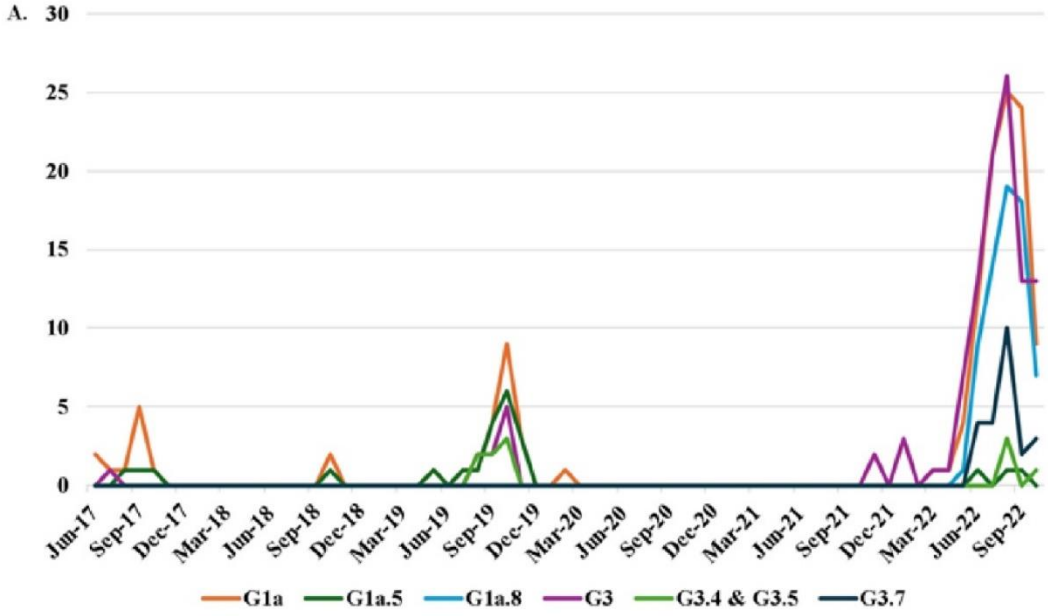
694 Recombination events temporarily plausible with a potential sequence originating from a
695 recombination event being detected after, or the same month and year as its two parents. NP:
696 Nucleocapsid Protein ; P: Phospho-protein ; M: Matrix protein ; F: Fusion protein; HN:
697 Hemagglutinin-Neuraminidase ; L: Large protein.

698

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY FIGURES

Supplementary Figure S1. Incidence of HPIV2 genotypes or sub-genotypes over time.



Number of genomes per genotype or sub-genotype per month among genomes obtain here with the two main genotypes G1a and G3.

SUPPLEMENTARY TABLES

Supplementary Table S1. Composition of the multiplex PCR's primer pool.

Primer's sequences (A). Composition for each pool and the number of primer duo per pools (B). Primer's final concentration in the complete reaction mix (C).

A.

Primer sequences			
1L	CATCCGCAAGTCAATGGTAGTA	2L	CAACAACAAGGGAGAATCAATGCT
1R	GTGGTGTATGTTGGTTCCTCGG	2R	TGTCCAGGCTCTCCTCTGTTCT
3L	ATYCAMTCAATAATCGGAGGTATGAC	4L	GYAAAGGCAGTATCCCACCAGG
3R	GATTGGTAGCTTCTCAGGTGGR	4R	GTGAAGCCTCCACTAGTYGGAC
5L	AACAACAATTAAAATCATGGATCCCAGYAC	6L	CAGATCCMACTTCACCCAGTCA
5R	AGGYTCATTTAGATACCTAATTCKRATCT	6R	TTGTTGGTCCCCAAAGATCTGC
7L	RGGATGCCTYGCATCAGTCTGG	8L	AGGGGTCATTCAATCAAAGATAAGGT
7R	CATGTTCCATTGYTTGGGGGAA	8R	AGATCAATTACCTTCGCACCGG
9L	AGYTCCGGACTGTTAACTGGTC	10L	AGTGCTGAGGAKTGGATTGCAGAT
9R	TGTAGGCAATCARCAATCCCACRA	10R	TKAATAATGCCTTGCTGTGAAT
11L	GAACTGCCACAATTCTTGGMAT	12L	TGYTATGTAGCTACAAGRTCTGARA
11R	CAYTGCCCTGTTGTATTTGGAAG	12R	ATGCAGTGTARAATGTGGGATTR
13L	CAGATGTGTGGCCGCTTAAY	14L	AGGGCACTTCCCGCATGATCTT
13R	RGGCACACGTTTAATTAATTCAGAYTTAG	14R	ACACCCATATACTAAAGATTCAAGAGATG
15L	CWCCACTAGGGCCAAGGATMGA	16L	AAGTGCTTTGACTTTGATCCTGG
15R	TGGRTCAAAACTGTCATCTTCTAAGAAATT		
17L	GTTCCACATTTATTTGAATGGATTCA	18L	GGGTCATCARCTAAAGGCTCAAGA
17R	ATCTCCGATATTGCGGTTAAATAATC	18R	GACAACTTCAATTGGATCTGGAGTTTCTA
19L	GACCAAACTTGCAGTATTGATATATCTAGA	20L	GCACACTAGCTCATCTTGTTGTG
19R	GCTAGYGGRTCAGGATCATATA	20R	CTGGTGGCATCCCTTTAATTTTYG
21L	TCTTGATTCTTTCAGARGATTCAATGG	22L	CAGGAGATACTGTCTATTATAATAGTCTTTTT
21R	TTGCCACAATTTATATGGAACACTC	22R	RGTGATAAGTGTCACYAGAGCATCT
23L	ACTACTGTTCTTAAGAGGGGTG	L =Left (Forward)	
23R	AGGACTGARTGAGARTTAAAGAAGG	R =Right (Reverse)	

B.

	Pools composition	Number of primer duo per pool
Pool 1	3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23	11
Pool 2	2, 4, 6, 8, 10, 14, 20, 22	8
Pool 3	1, 11L/13R, 16L/17R, 18	4
All primer in pool are paired with the corresponding L/R, except for the primer mix for pool 3		

719 **C.**

Primer pair	Final primer concentration in PCR Reaction Mix (μM)
1	0.2
2	0.2
3	0.5
4	0.2
5	0.5
6	0.2
7	0.5
8	0.2
9	0.5
10	0.2
11	0.5
11L/13R	0.5
13	0.5
14	0.2
15	0.5
16L/17R	0.5
17	0.5
18	0.2
19	0.5
20	0.2
21	0.5
22	0.2
23	0.5

720

721

722

723 **Supplementary Table S2.**

724 **A.**

Gene \ Genotype	G1a	G3	G1a.5	G1a.8	G3.4/5	G3.7
NP	3,98	0,81	3,87	3,92	0,68	1,06
P	4,85	**0,97**	4,52	5,03	**0,91**	**1,15**
M	4,36	0,64	4,24	4,37	0,7	0,34
F	**5,67**	0,93	**5,96**	**5,55**	0,77	0,93
HN	**7,74**	**1,14**	**7,89**	**7,83**	**1,27**	**1,16**
L	3,96	0,75	3,76	4,05	0,69	0,7

725

726 **B.**

Gene \ Genotype	G1a and G3 compared		G1a.5 and G1a.8 compared		G3.4/5 and G3.7 compared	
	G1a	**G3**	**G1a.5**	**G1a.8**	**G3.4/5**	**G3.7**
NP	1,68	0,36	0,42	0,45	0,51	**0,87**
P	1,44	0,46	0,71	**1,20**	**0,73**	**0,88**
M	2,39	0,35	**0,87**	1,02	0,62	0,28
F	**2,88**	**0,52**	0,71	0,55	0,64	0,86
HN	**4,04**	**0,57**	**1,04**	**1,07**	**0,92**	0,83
L	2,41	0,45	0,39	0,64	0,57	0,58

727

728

729 **Supplementary Table S3.**

A.

Gene \ Genotype	G1a	G3	G1a.5	G1a.8	G3.4/5	G3.7
NP	3.98	0.81	3.87	3.92	0.68	1.06
P	4.85	**0.97**	4.52	5.03	**0.91**	**1.15**
M	4.36	0.64	4.24	4.37	0.7	0.34
F	**5.67**	0.93	**5.96**	**5.55**	0.77	0.93
HN	**7.74**	**1.14**	**7.89**	**7.83**	**1.27**	**1.16**
L	3.96	0.75	3.76	4.05	0.69	0.7

730

B.

Gene \ Genotype	G1a and G3 compared		G1a.5 and G1a.8 compared		G3.4/5 and G3.7 compared	
	G1a	**G3**	**G1a.5**	**G1a.8**	**G3.4/5**	**G3.7**
NP	1.68	0.36	0.42	0.45	0.51	**0.87**
P	1.44	0.46	0.71	**1.20**	**0.73**	**0.88**
M	2.39	0.35	**0.87**	1.02	0.62	0.28
F	**2.88**	**0.52**	0.71	0.55	0.64	0.86
HN	**4.04**	**0.57**	**1.04**	**1.07**	**0.92**	0.83
L	2.41	0.45	0.39	0.64	0.57	0.58

731 **Supplementary Table S4.**

Score	Recombination Event no.	Recombinant Sequence (Genotype)	Minor Parental Sequence (Genotype)	Major Parental Sequence (Genotype)	Breakpoint Positions In Alignment		Recombination Size (nt) on alignment	Breakpoint Positions In NC_03443.1		Recombination Size (nt) on NC_03443.1	Location on NC_03443.1
					Start	End		Start	End		
7	1	687194775_zrxbh (G1a)	809605673_OXFRv (G3)	423507134_HEFrl (G1a)	1	658	657	1	658	657	Gene NP
7	2	574796828_nnrtw (G1a)	672827805_jqzcA (G1a)	369522769_rNVzQ (G1a)	438	546	108	438	546	108	
6	3	574796828_nnrtw (G1a)	672827805_jqzcA (G1a)	404745068_SnfDS (G1a)	574	734	160	574	699	125	
5	4	761032497_GAVVi (G1a)	851673922_TiFtE (G3)	881293965_Dhaaf (G1a)	745	1680	935	708	1024	316	
6	5	856750032_HvIITc (G1a)	367977552_gnRjQ (G3)	672827805_jqzcA (G1a)	3861	5676	1815	2231	3517	1286	Between gene V and M
6	6	845458065_oWOff (G3)	810158120_fwCcu (G1a)	261791997_JKnBc (G3)	4047	6065	2018	2370	3867	1497	
7	7	564720042_fYnzm (G3)	810158120_fwCcu (G1a)	439560848_Wwxlh (G3)	5996	6828	832	3803	4426	623	Gene M
7	8	158478296_TZkhZ (G1a)	512663666_KcBLu (G3)	245275920_dQIQq (G1a)	6372	6746	374	4156	4722	566	
6	9	881293965_Dhaaf (G1a)	523702868_zvCpK (G3)	687194775_zrxbh (G1a)	6470	7146	676	4209	4691	482	
6	10	672827805_jqzcA (G1a)	872307260_AJAFa (G3)	424652147_Fihhr (G1a)	6536	7702	1166	4264	5125	861	Between gene M and F
5	11	245275920_dQIQq (G1a)	903462368_PsrBC (G3)	160310781_jlmAu (G1a)	6964	7140	176	4514	4685	171	Gene M
7	12	639357067_xCpZf (G1a)	662085176_UxsGw (G3)	427459559_klKBy (G1a)	6966	7474	508	4516	4960	444	Between gene M and F
7	13	671122031_SgbRw (G1a)	662085176_UxsGw (G3)	301652630_wmIRF (G1a)	8193	9350	1157	5508	6320	812	Gene F
5	14	516609874_jwZSj (G1a)	577683690_UDllD (G3)	545927507_CtQrR (G1a)	9456	9921	465	6425	6859	434	Between gene F and HN
6	15	264789848_ZhcMg (G1a)	577683690_UDllD (G3)	642692991_QVOiN (G1a)	9526	10084	558	6476	6915	439	
6	16	530860623_QIcTf (G3)	574796828_nnrtw (G1a)	367977552_gnRjQ (G3)	9532	10001	469	6482	6896	414	
5	17	357495727_zfqLy (G3)	623332855_FIbXI (G3)	707275196_ttJWc (G3)	10660	15392	4732	7337	11256	3919	Between gene HN and L
6	18	378365121_xvNQg (G1a)	222356540_JhRnF (G3)	450041525_NtumP (G1a)	14066	14649	583	10156	10623	467	Gene L
7	19	512663666_KcBLu (G3)	511754512_kQQft (G1a)	577683690_UDllD (G3)	18952	19706	754	13978	14560	582	

732