

Une équipe internationale de chercheurs vient de découvrir un gène mystérieux dans le code génétique du coronavirus SARS-CoV-2 – un segment pratiquement caché dans le génome du virus, et largement ignoré jusqu'à présent. Ce gène inconnu pourrait expliquer pourquoi le virus est si difficile à combattre. Cette découverte a fait l'objet, ce 10 novembre, d'une publication approuvée par les pairs dans la revue scientifique *eLife*.

Le gène nouvellement identifié – appelé ORF3d – est un exemple de ce que les généticiens appellent un gène de chevauchement en raison de la façon dont il chevauche les séquences codées d'autres gènes. Il s'agit d'une sorte de « gène dans un gène » qui est effectivement caché dans une chaîne de nucléotides,



Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic



Chase W Nelson , Zachary Ardern , Tony L Goldberg, Chen Meng, Chen-Hao Kuo, Christina Ludwig, Sergios-Orestis Kolokotronis, Xinzhu Wei 

Biodiversity Research Center, Academia Sinica, Taiwan; Institute for Comparative Genomics, American Museum of Natural History, United States; Chair for Microbial Ecology, Technical University of Munich, Germany; Department of Pathobiological Sciences, University of Wisconsin-Madison, United States; Global Health Institute, University of Wisconsin-Madison, United States; Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University of Munich, Germany; Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Health Sciences University, United States; Institute for Genomic Health, SUNY Downstate Health Sciences University, United States; Division of Infectious Diseases, Department of Medicine, SUNY Downstate Health Sciences University, United States; Departments of Integrative Biology and Statistics, University of California, Berkeley, United States; Departments of Computer Science, Human Genetics, and Computational Medicine, University of California, Los Angeles, United States

« *En termes de taille du génome, le SRAS-CoV-2 et ses parents sont parmi les virus à ARN les plus longs qui existent* », [explique](#) le bioinformaticien Chase Nelson du Musée américain d'histoire naturelle, co-auteur de la recherche. Ils sont donc peut-être plus enclins à la « tromperie génomique » que les autres virus à ARN ». Les virus sont en fait assez enclins à héberger des gènes qui se chevauchent, ce n'est donc pas vraiment une découverte choquante. Il reste à voir si l'ORF3d représente vraiment une supercherie génomique destinée à rendre le virus plus difficile à combattre.

Les gènes qui se chevauchent sont difficiles à identifier dans les séquences génétiques, car les systèmes de balayage génomique peuvent souvent les manquer lorsqu'ils parcourent les chaînes du code génétique : programmés pour détecter des gènes individuels, ils ne voient pas nécessairement les instructions globales partagées entre les nucléotides des gènes adjacents dans une séquence.

Dans le contexte d'un virus comme le SRAS-CoV-2, cela pourrait constituer un sérieux point aveugle. Depuis le début de l'année, les scientifiques s'efforcent d'en savoir le plus possible sur ce virus dévastateur et si certains aspects de sa composition génétique ont été élucidés (notamment le consensus ferme selon lequel il n'a pas été « fabriqué en laboratoire »), il reste beaucoup de choses que nous ne savons pas encore.

« Le chevauchement des gènes peut faire partie de l'arsenal des moyens par lesquels les coronavirus ont évolué pour se répliquer efficacement, contrecarrer l'immunité de l'hôte ou se transmettre » explique Chase Nelson. Quant à l'ORF3d, il reste encore beaucoup à apprendre sur sa raison d'être, pourquoi et comment il se cache dans le génome et chevauche d'autres gènes.

En parcourant les bases de données génomiques, les chercheurs ont découvert que le gène avait déjà été identifié, mais seulement dans une variante du coronavirus qui affecte les pangolins (trouvé dans le Guangxi, en Chine). Il a également été précédemment classé à tort comme un gène non apparenté, ORF3b – qui est présent dans d'autres coronavirus, dont le SRAS-CoV – mais les scientifiques auteurs de l'article de *eLife* pensent qu'il ne s'agit pas vraiment de la même chose. « *Les deux gènes ne sont pas apparentés et codent pour des protéines entièrement différentes* », explique ainsi le professeur Nelson. « *Cela signifie que les connaissances sur l'ORF3b du SRAS-CoV ne devraient pas être appliquées à l'ORF3d du SRAS-CoV-2.* » En clair, ce gène qui était absent du SRAS-CoV original – qui a provoqué une épidémie mortelle entre 2002 et 2004 – pourrait expliquer pourquoi le nouveau coronavirus est si efficace pour se propager parmi les humains par rapport à son prédécesseur.



UniProtKB

Advanced

Search

BLAST Align Retrieve/ID mapping Peptide search SPARQL

Help Contact

UniProtKB - A0A166ZLS1 (A0A166ZLS1_9NIDO)

Basket

Display [Help video](#)

BLAST Align Format Add to basket History

Add a publication Feedback

Entry

Protein Submitted name: **3d protein**

Publications

Gene **ORF3d**

Feature viewer

Organism *Bat coronavirus*

Feature table

Status Unreviewed - Annotation score: ●○○○○○ - Protein predictedⁱ

None

Names & Taxonomyⁱ

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Processing

Expression

Interaction

Protein names ⁱ	Submitted name: 3d protein
Gene names ⁱ	Name: ORF3d
Organism ⁱ	Bat coronavirus
Taxonomic identifier ⁱ	1508220 [NCBI]
Taxonomic lineage ⁱ	Viruses > Riboviria > Orthornavirae > Pisuviricota > Pisoniviricetes > Nidovirales > Cornidovirineae > Coronaviridae > Coronavirinae > unclassified Coronavirinae
Proteomes ⁱ	UP000107989 Component ⁱ : Genome

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcell. location

Pathol./Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar proteins

Cross-references

Entry information

Subcellular locationⁱ

Topology

Feature key	Position(s)	Description	Actions	Graphical view	Length
Transmembrane ⁱ	39 – 59	Helical Sequence analysis	Add BLAST		21
Transmembrane ⁱ	71 – 90	Helical Sequence analysis	Add BLAST		20
Transmembrane ⁱ	96 – 113	Helical Sequence analysis	Add BLAST		18

GO - Cellular componentⁱ

- [integral component of membrane](#) [Source: UniProtKB-KW](#)

[Complete GO annotation on QuickGO ...](#)

Keywords - Cellular componentⁱ

Membrane

Family & Domainsⁱ

Keywords - Domainⁱ

Transmembrane, [Transmembrane helix](#) [Sequence analysis](#)

Sequenceⁱ

Sequence statusⁱ: Complete.

A0A166ZLS1-1 [UniParc] [FASTA](#) [Add to basket](#)

[« Hide](#)

Length: 227
Mass (Da): 25,562
Last modified: July 6, 2016 - v1
Checksum:ⁱ AD1061C874E095E7