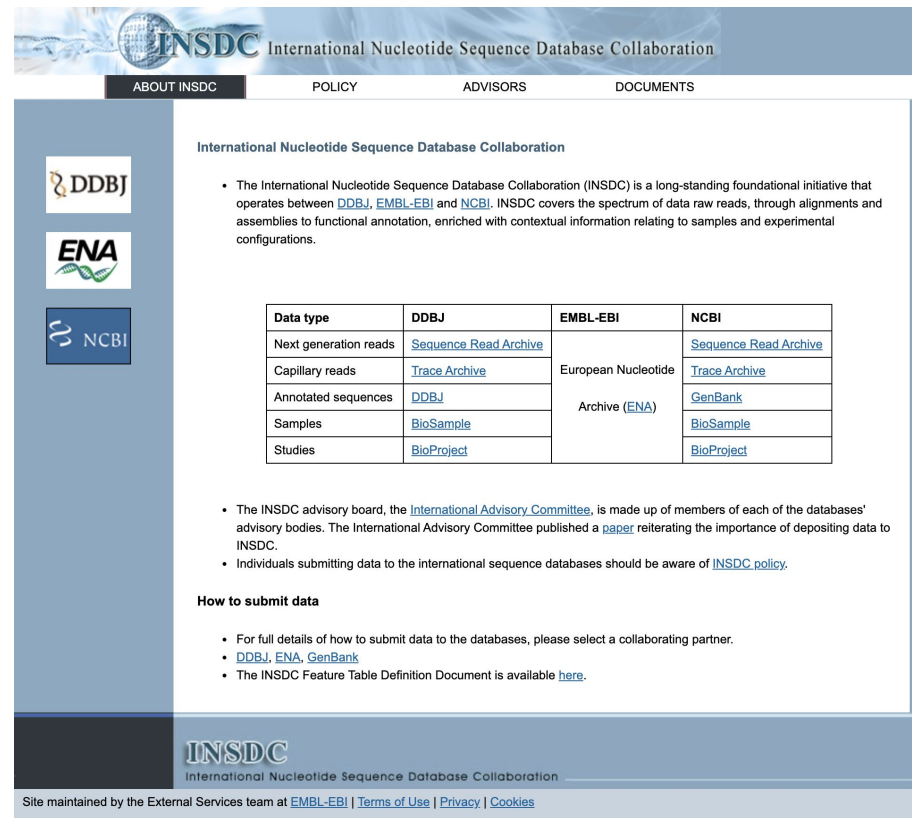


Bases de données de séquences biologiques

International Nucleotide Sequence Database Consortium (INSDC)

- Avant de publier un article scientifique qui repose sur des séquences, les biologistes sont tenus de déposer ces séquences dans l'une des trois bases de données internationales de référence:
 - ❑ NCBI (Etats-Unis)
 - ❑ EMBL-EBI-ENA (Europe)
 - ❑ DDBJ (Japon)
- Ces bases de données se sont organisées en un consortium : International Nucleotide Sequence Database Consortium (INSDC).
- Les séquences soumises à chaque base de donnée sont automatiquement copiées dans les deux autres.



The screenshot shows the official website of the International Nucleotide Sequence Database Collaboration (INSDC). The header features the INSDC logo and navigation links: ABOUT INSDC, POLICY, ADVISORS, and DOCUMENTS. The main content area is titled 'International Nucleotide Sequence Database Collaboration' and includes a descriptive paragraph about the consortium. A table lists the data types and their corresponding archives for DDBJ, EMBL-EBI, and NCBI. Below the table, there are sections for 'How to submit data' and a list of links for more information. The footer contains the INSDC logo and a note about the site being maintained by the External Services team.

International Nucleotide Sequence Database Collaboration

ABOUT INSDC POLICY ADVISORS DOCUMENTS

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Database Collaboration (INSDC) is a long-standing foundational initiative that operates between [DDBJ](#), [EMBL-EBI](#) and [NCBI](#). INSDC covers the spectrum of data raw reads, through alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations.

Data type	DDBJ	EMBL-EBI	NCBI
Next generation reads	Sequence Read Archive	European Nucleotide Archive (ENA)	Sequence Read Archive
Capillary reads	Trace Archive		Trace Archive
Annotated sequences	DDBJ		GenBank
Samples	BioSample		BioSample
Studies	BioProject		BioProject

- The INSDC advisory board, the [International Advisory Committee](#), is made up of members of each of the databases' advisory bodies. The International Advisory Committee published a [paper](#) reiterating the importance of depositing data to INSDC.
- Individuals submitting data to the international sequence databases should be aware of [INSDC policy](#).

How to submit data

- For full details of how to submit data to the databases, please select a collaborating partner.
- [DDBJ](#), [ENA](#), [GenBank](#)
- The INSDC Feature Table Definition Document is available [here](#).

INSDC
International Nucleotide Sequence Database Collaboration

Site maintained by the External Services team at [EMBL-EBI](#) | [Terms of Use](#) | [Privacy](#) | [Cookies](#)

- Le National Center for Biotechnology Information (NCBI) est le plus grand centre international de référence pour les données biologiques.
- Via son site Web « Entrez », le NCBI donne accès à une série de bases de données pour différents types d'informations
 - ❑ Séquences nucléiques (ADN, ARN)
 - ❑ Génomes
 - ❑ Séquences protéiques
 - ❑ Taxonomie des espèces vivantes
 - ❑ Littérature biomédicale
 - ❑ ...

The screenshot shows the NCBI Entrez homepage. At the top, there's a navigation bar with 'NCBI', 'Resources', and 'How To' links. A search bar is prominently displayed. Below the navigation bar, there is a banner for the 'Ending Structural Racism' NIH initiative. The main content area is divided into sections: 'Welcome to NCBI' with a brief description of the center's mission, 'Popular Resources' listing various databases like PubMed, Bookshelf, and BLAST, and 'NCBI News & Blog' with recent updates. A sidebar on the left provides a 'Resource List (A-Z)' with links to various biological data categories.

- La section Genomes du NCBI permet d'accéder rapidement à l'information disponible pour les génomes complètement séquencés.

The screenshot shows the NCBI GenBank page for the Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. The page is displayed in a web browser with the URL www.ncbi.nlm.nih.gov/nuccore/NC_045512.2. The top navigation bar includes links for NCBI, Resources, and How To. A search bar is present with the text "Nucleotide" and a "Search" button. The main content area displays the following information:

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[FASTA](#) [Graphics](#)

[Go to:](#) [GenBank](#) [Send to:](#) [Change region shown](#) [Customize view](#) [Analyze this sequence](#) [Run BLAST](#) [Pick Primers](#) [Highlight Sequence Features](#) [Find in this Sequence](#)

LOCUS NC_045512 29903 bp ss-RNA linear VRL 18-JUL-2020

DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome.

ACCESSION NC_045512

VERSION NC_045512.2

DBLINK BioProject: [PRJNA485481](#)

KEYWORDS RefSeq.

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Coronavirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1 (bases 1 to 29903)

AUTHORS Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y.Z.

TITLE A new coronavirus associated with human respiratory disease in China

JOURNAL Nature 579 (7798), 265-269 (2020)

PUBMED [32015508](#)

REMARK Erratum:[Nature. 2020 Apr;580(7803):E7. PMID: 32296181]

REFERENCE 2 (bases 13476 to 13503)

AUTHORS Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. and Howard, M.T.

TITLE Programmed ribosomal frameshifting in decoding the SARS-CoV genome

JOURNAL Virology 332 (2), 498-510 (2005)

PUBMED [15680415](#)

REFERENCE 3 (bases 29728 to 29768)

AUTHORS Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M. Jr. and Scott, W.G.

TITLE The structure of a rigorously conserved RNA element within the SARS virus genome

JOURNAL PLoS Biol. 3 (1), e5 (2005)

PUBMED [15630477](#)

REFERENCE 4 (bases 29609 to 29657)

NCBI Virus Retrieve, view, and download SARS-CoV-2 coronavirus genomic and sequences.

Related information

- [Assembly](#)
- [BioProject](#)
- [Protein](#)
- [PubMed](#)
- [Taxonomy](#)
- [Full text in PMC](#)
- [Gene](#)
- [Genome](#)
- [Identical GenBank Sequence](#)
- [Mature Peptides](#)
- [Other INSDC Genome Sequence](#)
- [PubMed \(Weighted\)](#)

- <https://www.uniprot.org/>
- KB: knowledge base
- UniProt KB vise à rassembler l'information sur toutes les séquences protéiques caractérisées par les biologistes.
- Swiss-Prot contient des séquences protéiques "annotées" par des biologistes. L'annotation consiste à associer à une séquence les connaissances résultant d'expérimentation.
 - ☐ Fonction de la protéine
 - ☐ Domaines structurels
 - ☐ Sites actifs (enzymes)
 - ☐ ...

UniProtKB - PODTC2 (SPIKE_SARS2)

Display video

Entry

Publications

Feature viewer

Feature table

Protein | **Spike glycoprotein**

Gene | **S**

Organism | *Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)*

Status | **Reviewed** - Annotation score: ●●●●●●
- Experimental evidence at protein levelⁱ

Functionⁱ

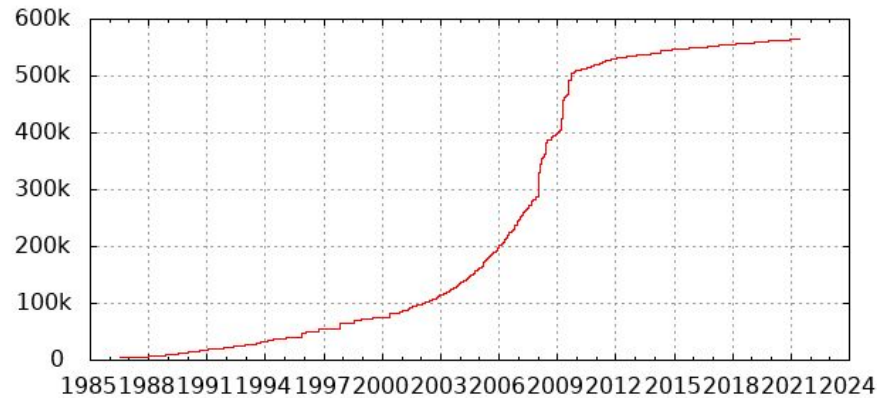
Spike protein S1:
attaches the virion to the cell membrane by interacting with host receptor, initiating the infection. Binding to human ACE2 receptor and internalization of the virus into the endosomes of the host cell induces conformational changes in the Spike glycoprotein (PubMed:32142651, PubMed:32221306, PubMed:32075877, PubMed:32155444).

Binding to host NRP1 and NRP2 via C-terminal polybasic sequence enhances virion entry into host cell (PubMed:33082294, PubMed:33082293).

This interaction may explain virus tropism of human olfactory epithelium cells, which express high level of NRP1 and NRP2 but low level of ACE2

- Deux limitations à l'annotation
 - Le nombre de publications augmente tellement qu'il n'est pas possible à l'équipe de Swiss-Prot de tout annoter
 - Le nombre de séquences augmente de façon tellement rapide qu'il est impossible de toutes les caractériser expérimentalement
- TREMBL
 - annotation automatique des séquences traduites de la base de données EMBL.
- Uniprot = Swiss-Prot + TREMBL
 - Swiss-Prot 580.000 séquences
 - TREMBL >200 millions !
- Conséquence: la vaste majorité des séquences sont annotées automatiquement, sans possibilité de les vérifier individuellement

Number of entries in UniProtKB/Swiss-Prot



Number of entries in UniProtKB/TREMBL

