

Comment étudier l'origine de SARS-CoV-2 ?

Extrait d'une activité de l'Atelier 5 : <http://education.expasy.org/bioinformatique/>
(Atelier 26 dans la liste <http://unige.ch/-/bioinformatique>)

Expasy Ateliers de bioinformatique

You are here: Home > Education > Ateliers de bioinformatique

00118926

Ateliers de bioinformatique

Une activité proposée par le **SIB Institut Suisse de Bioinformatique** (contact: [outreach\(at\)sib.swiss](mailto:outreach(at)sib.swiss)).

Définition: La **bioinformatique** est une discipline des sciences de la vie qui s'appuie sur des outils informatiques pour **stocker**, **analyser** et **visualiser** des données biologiques. C'est un domaine de recherche et d'expertise en pleine expansion et les applications sont très variées.

Exemples:

- Stocker la **séquence en acides nucléiques** du chromosome 1 humain
- Visualiser la **structure 3D** de l'insuline
- Prédire la **structure d'un gène** à partir d'une séquence ADN
- Choisir le meilleur traitement en fonction d'un **profil génétique**
- Découvrir **différentes cellules**

Vous allez découvrir des sites internet et des outils bioinformatiques utilisés par les biologistes du monde entier.
Le but de ces ateliers est de vous familiariser avec certains d'entre eux et surtout vous faire une idée de ce qu'on peut faire avec...

Une exposition virtuelle sur le **génomme humain**: www.ChromosomeWalk.ch (dossier pédagogique)
Comment les chercheurs conçoivent-ils **les médicaments de demain?**: www.atelier-drug-design.ch
NEW Un tour d'horizon pour comprendre **la médecine de précision**: www.PrecisionMed.ch

Atelier 1	Atelier 2
Des chromosomes et des gènes	Protéines et drug design
Atelier 3	Atelier 4
Phylogénie, biodiversité et pizza...	Médecine de précision et profil génétique
Atelier 5	Atelier 6
Coronavirus et protéines...	Découverte de BLAST...
Atelier 7	Quelques liens
L'insuline de A à Z	Pour les plus curieux.....



Les coronavirus infectent de nombreuses espèces de mammifères



- SARS coronavirus (*homme*)
- Civet coronavirus (*civette*)
- Hedgehog coronavirus (*hérisson*)
- Bat coronavirus (*chauve-souris*)
- Rabbit coronavirus (*lapin*)
- Camel coronavirus (*chameau*)
- Dog coronavirus (*chien*)
- Rat coronavirus (*rat*)
- Bovine coronavirus (*vache*)
- Equine coronavirus (*cheval*)
- Yak coronavirus (*yak*)
- Pangolin coronavirus (*pangolin*)
- Porcine coronavirus (*cochon*)



Certains coronavirus provoquent également des bronchites chez les poulets et une maladie intestinale mortelle chez les porcs

*La classification des virus est particulièrement complexe...
Il n'est pas toujours facile de s'y retrouver...*

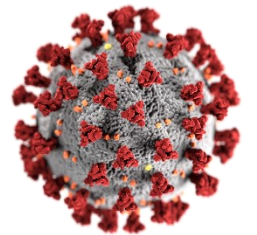
Question à se poser pour étudier l'origine de SARS-CoV-2:

**Quel coronavirus est le plus similaire
à SARS-CoV-2 ?**

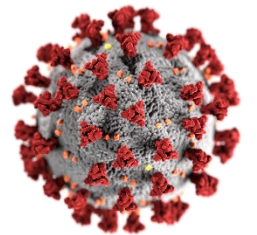
Bioinformatique

Outil de comparaison de sequence: align@UniProt

Banque de données de protéines: [UniProtKB/Swiss-Prot](https://www.uniprot.org/)



Comparer la protéine Spike de différents coronavirus (chauve-souris, civette, pangolin, ...)



Démarche 'manuelle'

La protéine Spike chez différents coronavirus (1)

Voici des séquences partielles de la protéine Spike de différents coronavirus infectant différentes espèces, différentes années.

Quelle séquence est la plus similaire à human SARS-CoV-2 ?

Quelle(s) hypothèse(s) pourriez-vous faire ?

Human SARS-CoV-2	(2020)	IRGDEVQRQIAPGQTGKIAD
Pangolin coronavirus	(2020)	VRGDEVQRQIAPGQTGRIAD
Human SARS-CoV	(2003)	VKGDDVQRQIAPGQTGVIAD
Bat coronavirus	(2020)	ITGDEVQRQIAPGQTGKIAD



wikipedia

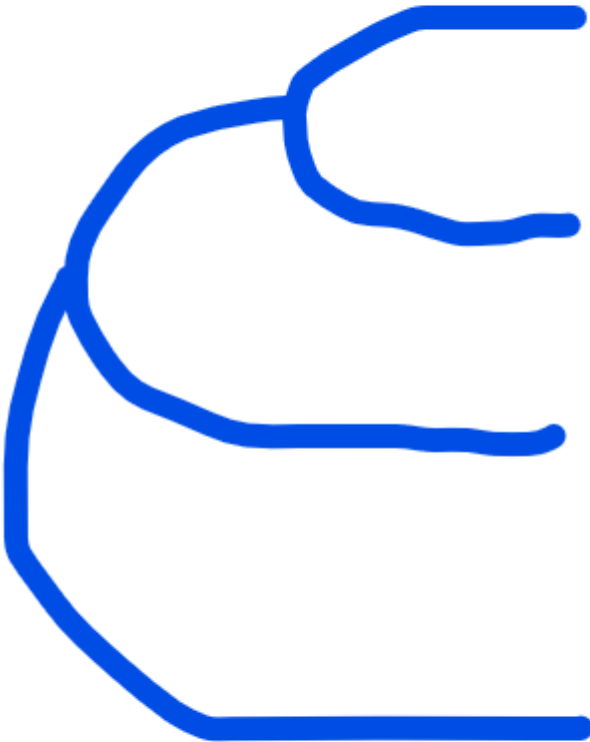


wikipedia

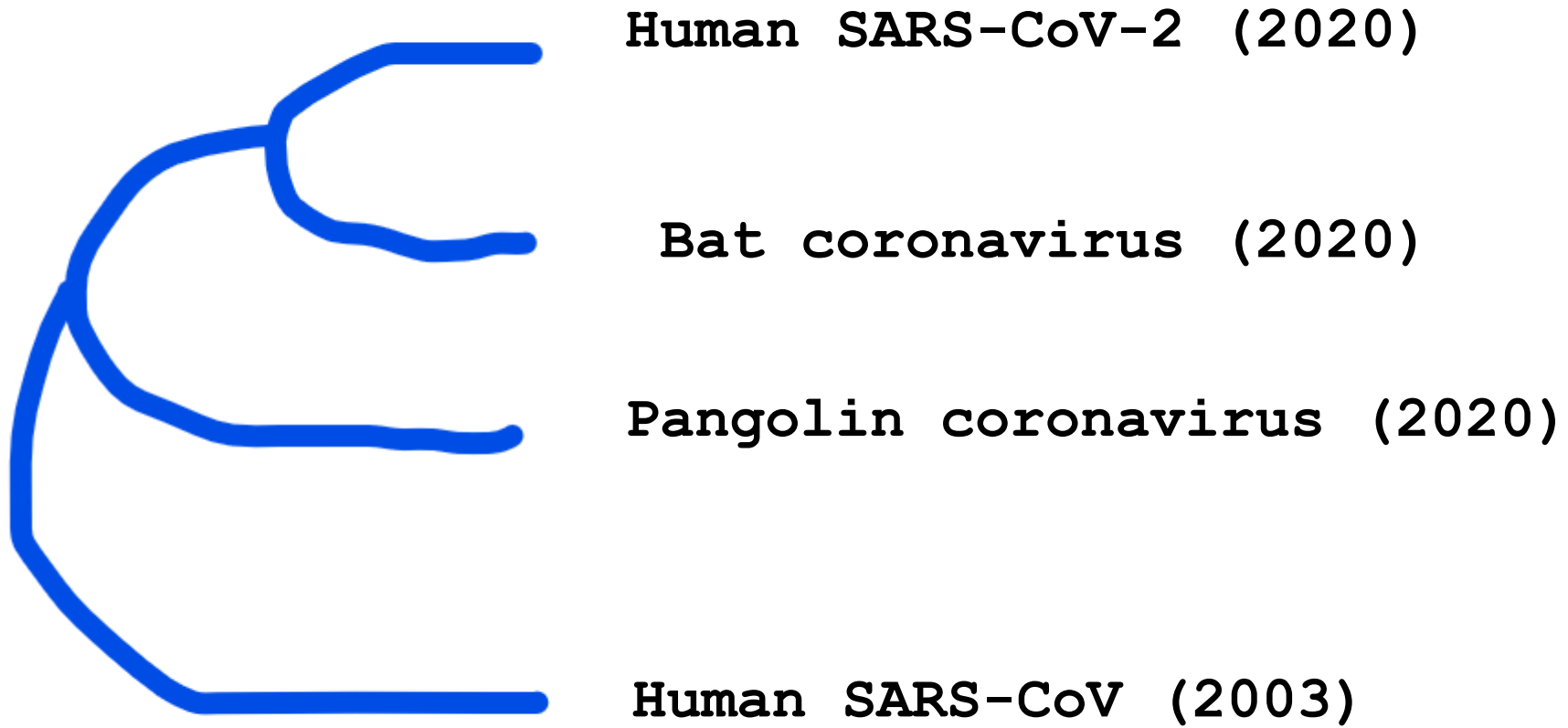
identité	Human SARS-CoV-2 2020	Pangolin coronavirus 2020	Bat coronavirus 2020	Human SARS-CoV 2003
Human SARS-CoV-2	100 %			
Pangolin coronavirus		100		
Bat coronavirus			100	
Human SARS-CoV				100

identité	Human SARS-CoV-2 2020	Pangolin coronavirus 2020	Bat coronavirus 2020	Human SARS-CoV 2003
Human SARS-CoV-2	100 %			
Pangolin coronavirus	2 mutations/19 89,5 %	100		
Bat coronavirus	1 mutation/19 94.7 %	3 mutations/19 84 %	100	
Human SARS-CoV	3 mutations/19 84 %	3 mutations/19 84 %	4 mutations/19 79 %	100

Human SARS-CoV-2 (2020)



?



Démarche 'bioinformatique' (1)

La protéine Spike chez différents coronavirus (2)

Voici des séquences partielles de la protéine Spike de différents coronavirus infectant différentes espèces, différentes années.

```
>Human_SARS2020
FSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD
>Civet_2003_coronavirus
FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVQRQIAPGQTGVIAD
>Pangolin_2020_coronavirus
FSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVQRQIAPGQTGRIAD
>Human_SARS2003
FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVQRQIAPGQTGVIAD
>Bat_2020_coronavirus
FSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVQRQIAPGQTGKIAD
>Human_MERS2012
VNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQ
>Bat_2007_coronavirus
VDEFSCNGISPDSIARGCYSTLTVDYFAYPLSMKSYIRPGSAGNIPL
```

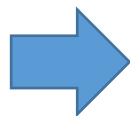
Result information

Query sequences



```
>Human_SARS2020
FSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD
>Civet_2003_coronavirus
FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVQRQIAPGQTGVIAD
```

Date of job execution	2020-10-30
Job identifier	A20201030E5A08BB0B2D1C45B0C7BC3B55FD26556001CE75 (jobs are stored for 7 days)
Running time	13.6 seconds
Identical positions	41
Identity	87.234%
Similar positions	4
Program	CLUSTALO



La protéine Spike chez différents coronavirus (2)

Comparer [les séquences partielles de la protéine Spike de différents coronavirus](#), en faisant des alignements 2 à 2 avec la protéine Spike de SARS-CoV-2 (2020) à l'aide de [Align@UniProt](#).

Remplir le tableau suivant:

% identité	Human_SARS2020
Human_SARS2020	100
Human_SARS2003	
Civet_2003_coronavirus	
Pangolin_2020_coronavirus	
Bat_2020_coronavirus	
Human_MERS2012	
Bat_2007_coronavirus	

% identité	Human_SARS2020
Human_SARS2020	100
Human_SARS2003	87,2
Civet_2003_coronavirus	87,2
Pangolin_2020_coronavirus	96,7
Bat_2020_coronavirus	97,9
Human_MERS2012	18,1
Bat_2007_coronavirus	22,7

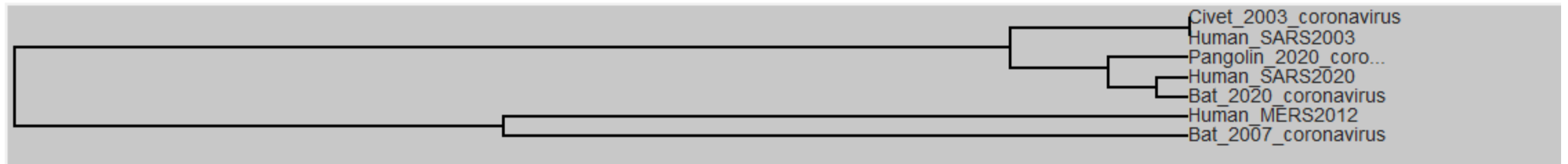
Réponse:

```

Human_SARS2020      1  FSTFKCYGVSPTKLNLDLFCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD
Civet_2003_coro..  1  FSTFKCYGVSATKLNLDLFCFTNVYADSFVVKGDDVRQIAPGQTGVIAD
Pangolin_2020 c..  1  FSTFKCYGVSPTKLNLDLFCFTNVYADSFVVRGDEVQRQIAPGQTGRIAD
Human_SARS2003     1  FSTFKCYGVSATKLNLDLFCFTNVYADSFVVKGDDVRQIAPGQTGVIAD
Bat_2020 corona..  1  FSTFKCYGVSPTKLNLDLFCFTNVYADSFVITGDEVQRQIAPGQTGKIAD
Human_MERS2012    1  VNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLVSSAGPISQ
Bat_2007_corona..  1  VDEFSCNGISPDSIARGCYSTLTVDYFAYPLSMKSYIRPGSAGNIPL
.. *.* : * : *::: * * . : ..* *

```

Tree



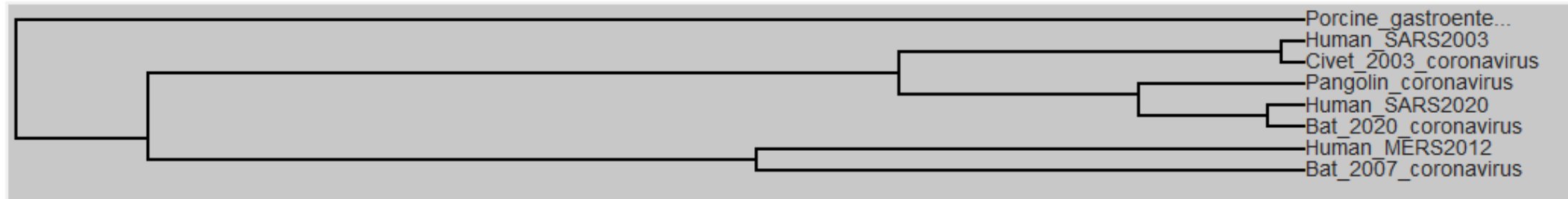
Démarche 'bioinformatique' (3)

La protéine Spike chez différents coronavirus (3)

Copier/coller les [séquences complètes des protéines Spike de différents coronavirus](#) dans [Align@UniProt](#)

Voici une représentation très simplifiée des relations évolutives existant entre différents coronavirus infectant différentes espèces de mammifères

Tree



Quelle(s) chaîne(s) de transmission de SARS-CoV-2 serait envisageable selon ces résultats ?

homme -> homme	1
cochon -> homme	2
civette -> homme	3
chauve-souris -> homme	4
pangolin -> homme	5
chauve-souris -> pangolin -> chauve-souris -> homme	6

Attention: il s'agit de découvrir la démarche scientifique et de formuler des hypothèses et en aucun cas des conclusions !
Des conclusions sur les chaînes de transmission sont impossibles à faire avec si peu de données!

L'échantillonnage (présence ou absence d'une séquence (coronavirus du chameau 2012)) et/ou des erreurs de séquençage peuvent influencer de manière significative l'interprétation