# COVID-19, SARS and Bats Coronaviruses Genomes Unexpected Exogeneous RNA Sequences

Jean Claude Perez, Luc Montagnier

**Jean-Claude Perez**, PhD Maths § Computer Science Bordeaux University, RETIRED interdisciplinary researcher (IBM Emeritus, IBM European Research Center on Artificial Intelligence Montpellier), Bordeaux metropole, France

**Luc Montagnier** , Paris, France

## ABSTRACT :

We human are facing the worldwide invasion of a new coronavirus. This follows several limited outbreaks of related viruses in various locations in a recent past (SARS, MERS). Although the main objective of researchers is to bring efficient therapeutic and preventive solutions to the global population, we need also to better understand the origin of the newly coronavirus-induced epidemic in order to avoid future new outbreaks. The present molecular appraisal is to study by a bio-infomatic approach the facts relating to the virus and its precursors.

This article demonstrates how 16 « Exogeneous Informative Elements » fragments (Env Pol and Integrase genes) from different strains, both diversified and very recent, of the HIV1, HIV2 and SIV retroviruses most likely are present into the genome of COVID-19. Among these fingerprints, 12 of them would be concentrated in a very small region of the genome COVID-19 of length less than 900bases, i.e. less than 3% of the total length of this genome. In addition, these footprints are positioned in 2 functional genes of COVID-19: the orf1ab and S spike genes.

To sum up, here are the two main facts which contribute to our hypothesis of a partially synthetic genome: A contiguous region representing 2.49% of the whole COVID-19 genome is 40.99% made up of 12 diverse Exogeneous Informative Elements (EIE) fragments originating from various strains of HIV SIV retroviruses. On the other hand, these 12 Exogeneous Informative Elements, some of them appear concatenated, that is to say placed side by side in the genome of COVID-19, and this despite natures, strains, and years of emergence all different.

Notably, the retroviral part of these regions, which consists of 8 motifs from various strains HIV1, HIV2 and SIV, covers a length of 275 contiguous bases of COVID-19. The cumulative length of these 8 HIV SIV motifs represents 200 bases. Consequently, the HIV SIV density rate of this region of COVID-19 is 200/275 = 72.73%, which is considerable.

A major part of these 16 EIE Elements already existed in the first SARS genomes as early as 2003. However, we demonstrate how and why a new region including 4 HIV1 HIV2 Exogeneous Informative Elements radically distinguishes all COVID-19 strains from all SARS and Bat strains.

Particularly, we will be interested in the Bat RaTG13 strain whose genomic proximity to COVID-19 will be thoroughly analyzed. Then, we gather facts about the possible origins of COVID_19, we have particularly analyze this small region of 225 bases common to COVID_19 and batRaTG13 but totally absent in all SARS strains.

Then, we discuss the case of bat genomes presumed to be at the origin of COVID_19. In the strain of bat RaTG13 bat coronavirus isolated in 2013, then sequenced in 2020, the homology profile for HIV1 Kenya 2008 fingerprint is identical to that of COVID_19. (collected end december 2019, then sequenced in 2020).

Finally, we have studied the most recent genetic evolution of the COVID_19 strains involved in the world epidemic. We found an astoneshing occurrence of mutations and deletions in the 225b region.

On sampling genomes, we finally show that this 225b key region of each genome, rich in "EIE", evolves much faster than the corresponding whole genome.

# INTRODUCTION :

We are facing the worldwide invasion of a new coronavirus. This follows several limited outbreaks of related viruses in various locations in a recent past [1, 2]. The human civilization has been very successful in the last centuries with regard demographic and economic growths. However, in our times, the economic power has been concentrated in the hands of a few individuals and consequently economic interests are prevailing over the well beeing of humanity.
Although the main objective of researchers is to bring efficient therapeutic and preventive solutions to the global population, we need also to better understand the origin of the newly coronavirus-induced epidemic in order to avoid future new outbreaks. The present molecular appraisal is to study by a bio-infomatic approach the facts relating to the virus and its precursors .

We had analyzed the evolution of coronaviruses from the first SARS (2003), to the first genomes of COVID-19, when it was still called 2019-nCoV [3]. It is then that we have knowledge of this online article [4] according to which a region of around 1kb is totally new in the genome of COVID-19.

According to our methods having made it possible to evaluate the level of cohesion and organization of a genome, we then discover that the deletion of this region of 1kb suspected new would INCREASE the level of organization of the genome.
We can then suggest that this region was "added" to the genome. This is the line of research we have followed. We then have the intuition to search in this genome for possible traces of HIV or even SIV. A first publication [5] reports the discovery if 6 HIV SIV RNA pieces relates to crucial retroviruses genes like Enveloppe and RT Pol. The present article confirms then extends these initial results.

# MATERIALS and METHODS :

**Access to data banks :**
Preliminary Note :
The COVID-19 genome sequence initially studied in this article is NC_045512.2 .More generally, we are interested in the first genomes published under the reference "Wuhan market". However, these sequences published in January 2020 evolved somewhat during the first quarter of 2020. Thus, NC_045512.2
has evolved from 29866bp to 29903bp without its GENBANK NCBI reference was changed.

All these sequences of genomes referenced "Wuhan market" relate to individual patients, were deposited on January 30, 2020 and then re-published on March 6, 2020. For these reasons we will have to specify and adjust here the addresses of the key regions "A" and "B " which we analyze in this article.

The Wuhan market referenced genomes are presently:
https://www.ncbi.nlm.nih.gov/nuccore/LR757995.1
https://www.ncbi.nlm.nih.gov/nuccore/LR757996.1
https://www.ncbi.nlm.nih.gov/nuccore/LR757997.1
**https://www.ncbi.nlm.nih.gov/nuccore/LR757998.1**
and
https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2

Thus, the start address of the region of 330bp named in this article "region B" is positioned at 21673bp in our article while it is now shifted at 21698bp in NC_045512.2 , at 21683bp in LR757995.1, at 21678bp in LR757996.1, , and at 21673bp in **LR757998.1**. The sequence LR757997.1, is unavailable because it contains more than 10,000 indeterminate « N » bases.
Then, finally, this region « B » has the same starting adress in our NC_045512.2 reference sequence and in LR757998.1 .

Then, finally the reference sequence used in this article is : **https://www.ncbi.nlm.nih.gov/nuccore/LR757998.1**

### Then we use as reference the formal referenced genome : Wuhan market ID: LR757998.1

**Validation of nucleotides fragments as « Exogeneous Informative Elements » (« EIE ») :**

We have chosen this minimal length of 18 nucleotides ( 6 amino acids ) for the support of information ( thus as an antigenic motif ). This is also the size of the primers used for PCR which allows high specificity of sequence selection on DNA recognition.

**Main COVID_19 genes involved :**

The two main genes involved in COVID-19 genome are Orf1ab and « S » Spike.
Their relative adresses in our referenced genome are :
266. 21555. Orf1ab
21563..25384. S spike

**The main analysed regions :**

Region « A », Location of the 600 first bp from COVID_19 reference genome Wuhan market **ID: LR757998.1**.

Their length was between 21072 and 21672 nucleotides.

AGGGTTTTTTCACTTACATTTGTGGGTTTATACAACAAAAGCTAGCTCTTGGAGGTTCCGTGGCTATAAAGA
TAACAGAACATTCTTGGAATGCTGATCTTTATAAGCTCATGGGACACTTCGCATGGTGGACAGCCTTTGTTA
CTAATGTGAATGCGTCATCATCTGAAGCATTTTTAATTGGATGTAATTATCTTGGCAAACCACGCGAACAAAT
AGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTCCTATTCTT
TATTTGACATGAGTAAATTTCCCCTTAAATTAAGGGGTACTGCTGTTATGTCTTTAAAAGAAGGTCAAATCAA
TGATATGATTTTATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGT
GATGTTCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGT
GTTAATCTTACAACCAGAACTCAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTG
ACAAAGTTTTCAGATCC

see details alignment in supplementary materials « a ».

Region « B », Location of the 330 first bp from COVID_19 reference genome Wuhan market **ID:** LR757998.1**.**

Their length was between 21672 and 22002 nucleotides (then immediately following region « A » :

TCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTGGTTCCATGCTATACATG
TCTCTGGGACCAATGGTACTAAGAGGTTTGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTT
CCACTGAGAAGTCTAACATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGACCCAGTCCCTA
CTTATTGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCCATTTTTGGG
TGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGT

see details alignment in supplementary materials « b ».

Then, we analysed also this larger region which starts at the same address as our region "B" :
entitled « Region Lyons-Weiler » [4].

Their length was between 21672 and 23050 (1378 nucleotides) within reference genome Wuhan market **ID: LR757998.1**

In the Discussion §, we will more particularly analyze a small region of 225 nucleotides located between the bases. and. of the reference genome:

TGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACT
CAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTT
TCAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTAC
TTGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTAA

**Alignments : Analysing COVID-19 DNA sequences ,** We use BLAST NCBI public tool.

BLASTn - NIH

**NCBI** National Center for Biotechnology Information.

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch

**Relating The "DNA Master Code" of DNA method used in Conclusions§ :**

Full details on this numerical method in [6-8], and recall Methods in supplementary Materials2 « m ».

# RESULTS :

We are now looking for possible traces of HIV1, HIV2 or SIV « EIE » into our Wuhan market reference genome **LR757998.1** .

We will only use as significant EIE those which have at least 18 nucleotides of homology, i.e. 6 codons.

Note: We will present below 12 + 4 HIV/SIV "EIE" elements in the sequential order of their addresses within COVID_19 genome. However, it is worth recalling here the history of successive discoveries from these regions. Initially, by focusing on the genome region mentioned in [4], we discovered and published [5] 6 first "EIE" located at the very beginning of this region.
Then, by, on the one hand, a more in-depth exploration of this region (region "B" 330bp), then, on the other hand, by exploring region "A" (of 600bp) immediately located upstream of this region "B ", we discover, concentrated on less than 930bp, 12 HIV SIV "EIE". We then complete them with the last 4 "EIE" located upstream in the genome. It is this set of 16 "EIE" which will be detailed below.
In the "Discussion" chapter, we will show that 12 of this 16 "Exogeneous Informative Elements" were yet presents in SARS genomes. On the other hand, a small region of 225bp contains the 4 "EIE" absent from SARS but present in both COVID_19 and in the bat genome bat RaTG13. We will devote an in-depth analysis of these important facts in the "Discussion" chapter.

see details alignment in supplementary materials « c ».

## 1/ Evidence for 12 HIV/SIV "EIE" sequences in regions "A" and "B" of COVID-19 genome:

Following, the 14 HIV/SIV "Exogeneous Informative Elements":

**Region A : 600bp (21072 to 21672)**

**==> HIV2 env 66-81 France 2012**

**HIV-2 isolate 56 from France envelope glycoprotein (env) gene, partial cds**
**Sequence ID: JN230738.1Length: 234Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 30.1 bits(32) | 2.8 | 16/16(100%) | 0/16(0%) | Plus/Plus |

```
Query  66  ATAAAGATAACAGAAC  81
           ||||||||||||||||
Sbjct  77  ATAAAGATAACAGAAC  92
```

we consider this length of 16 nucleotides as insufficiently significant.

**==> HIV1 154-174 Sweden 2017**

**HIV-1 isolate 060SE from Sweden, partial genome**
**Sequence ID: MF373163.1Length: 8732Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 39.2 bits(42) | 0.87 | 21/21(100%) | 0/21(0%) | Plus/Minus |

```
Query  154  ATGCGTCATCATCTGAAGCAT  174
            |||||||||||||||||||||
```

Sbjct 5634 ATGCGTCATCATCTGAAGCAT 5614

**==> HIV2 236-253 Guinea Bisseau 2012**

**HIV-2 isolate CA65410.13 from Guinea-Bissau envelope gene, partial cds**
**Sequence ID: JN863831.1Length: 2028Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 29.2 bits(31) | 9.7 | 17/18(94%) | 0/18(0%) | Plus/Minus |

```
Query 236 TGCAAATTACATATTTTG 253
          ||||||||||||| |||||
Sbjct 436 TGCAAATTACATCTTTTG 419
```

**==> SIV 366-386 2016 Africa**

**Simian immunodeficiency virus isolate VSAA2001, complete genome**
**Sequence ID: KR862351.1Length: 9053Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 34.6 bits(37) | 2.5 | 20/21(95%) | 0/21(0%) | Plus/Minus |

```
Query 366  GATATGATTTTATCTCTTCTT 386
           |||| ||||||||||||||||
Sbjct 8635 GATAGGATTTTATCTCTTCTT 8615
```

These first 4 HIV SIV motifs are located in this COVID-19 gene **« orf1ab »**:

```
                     /collection_date="Dec-2019"
     5'UTR           1..265
     gene            266..21555
                     /gene="orf1ab"
                     /locus_tag="GU280_gp01"
                     /db_xref="GeneID:43740578"
```

**==> HIV1 471-501 Kenia 2008 [9]**

This HIV1 signature region is located SIMULTANEOUSLY between the end of the "orf1ab" gene and the start of the "S spike" gene.

**HIV-1 clone ML1592n from Kenya nonfunctional vpu protein (vpu) gene, complete sequence; and nonfunctional envelope glycoprotein (env) gene, partial sequence**
**Sequence ID: EU875177.1Length: 601Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 37.4 bits(40) | 3.0 | 28/32(88%) | 1/32(3%) | Plus/Minus |

```
Query 471 TGTTTTTCTTG-TTTTATTGCCACTAGTCTCT 501
          ||||||| ||  |||||||||||||| |||||
Sbjct 270 TGTTTTTATTACTTTTATTGCCACTATTCTCT 239
```

**==> HIV2 512-529 Cap vert 2012 au delà de cette région de gene...**

**HIV-2 isolate 05HANCV37 from Cape Verde envelope glycoprotein (env) gene, partial cds**
**Sequence ID: JF267434.1Length: 342Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 33.7 bits(36) | 0.23 | 18/18(100%) | 0/18(0%) | Plus/Plus |

```
Query  512  TTAATCTTACAACCAGAA  529
            ||||||||||||||||||
Sbjct  284  TTAATCTTACAACCAGAA  301
```

**Region B : 330bp (21672 to 22002)**

**Details:**
Hiv2. Côté ivoire 23. 42 *
Siv Tanzania 29 50 partial overlap
Siv p18. 77 96 *
Hiv1. Netherlands. 85. 112.  Usa 85 108 *
Hiv2 UC1. 132 157 *
Hiv2 Sénégal. 179 194 *
Hiv1 Malawi. 212 243 *
Hiv1. Russia. 242 280 *
SivagmTan 279 298 *

We consider only the 8 (*) HIV SIV motifs, the 9th is partially in overlap.

**==> Region HIV2a 24-43:**

**HIV-2 isolate 106CP_RT from Cote d'Ivoire reverse transcriptase gene, partial cds**
**Sequence ID: KJ131112.1Length: 924Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 32.8 bits(35) | 0.38 | 19/20(95%) | 0/20(0%) | Plus/Minus |

```
Query  24  ACTTGTTCTTACCTTTCTTT  43
           ||||||||||| ||||||||
Sbjct  85  ACTTGTTCTTATCTTTCTTT  66
```

**==> Region SIV 29-50** Siv Tanzania 29 50 is partially in overlap with the above fingerprint, therefore, we will not retain it for our analyzes. .

**Simian immunodeficiency virus isolate TAN5 from Tanzania, complete genome**
**Sequence ID: JN091691.1Length: 9893Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 31.9 bits(34) | 4.7 | 20/22(91%) | 0/22(0%) | Plus/Minus |

```
Query  29    TCTTACCTTTCTTTTCCAATGT  50
             |||| |||| ||||||||||||
Sbjct  2801  TCTTTCCTTCCTTTTCCAATGT  2780
```

**==> Region SIV P18  77-96**

**Simian immunodeficiency virus isolate P18 patient P1, gp120 (env) gene, partial cds**
**Sequence ID: AF003044.1Length: 1089Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 32.8 bits(35) | 4.7 | 19/20(95%) | 0/20(0%) | Plus/Plus |

```
Query  77   CTGGGACCAATGGTACTAAG  96
            ||||||| ||||||||||||
Sbjct  845  CTGGGACTAATGGTACTAAG  864
```

**==> Region Hiv1. Netherlands. 85. 112.  Usa 85 108**

**HIV-1 isolate 19828.PPH11 from Netherlands envelope glycoprotein (env) gene, partial cds**

**Sequence ID: HQ644953.1Length: 1143Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 38.3 bits(41) | 1.6 | 25/28(89%) | 0/28(0%) | Plus/Plus |

```
Query  85   AATGGTACTAAGAGGTTTGATAACCCTG  112
            ||||||||||| ||||| ||||||  |||
Sbjct  967  AATGGTACTAAAAGGTTAGATAACACTG  994
```

**==> Region Hiv2 UC1. 132 157**

**Human immunodeficiency virus type 2 complete genome from strain HIV-2UC1**
**Sequence ID: L07625 .1Length: 10271Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 30.1 bits(32) | 1.5 | 22/26(85%) | 0/26(0%) | Plus/Plus |

```
Query  132   TGTTTATTTTGCTTCCACTGAGAAGT  157
              |||||||||||||| | ||| | ||||
Sbjct  6701  TGTTTATTTTGCTCCTACTTATAAGT  6726
```

**==> Region Hiv2 Sénégal. 179 194**

**HIV-2 isolate H2A62_111808_CINT_WBC_25 from Senegal pol gene, partial sequence**
**Sequence ID: JF811228.1Length: 981Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 30.1 bits(32) | 1.5 | 16/16(100%) | 0/16(0%) | Plus/Minus |

```
Query  179  TTTTTGGTACTACTTT  194
             ||||||||||||||||
Sbjct  803  TTTTTGGTACTACTTT  788
```

we consider this length of 16 nucleotides as insufficiently significant.

**==> Region Hiv1 Malawi. 212 243**

**HIV-1 isolate 4045_Plasma_Visit1_amplicon9 from Malawi envelope glycoprotein (env) gene, complete cds**

**Sequence ID: KC187066.1Length: 2571Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 37.4 bits(40) | 1.6 | 28/32(88%) | 1/32(3%) | Plus/Minus |

```
Query  212  CCCTACTTATTGTTAATAACGCTACTAATGTT  243
             ||||||| || |||| |||| ||||||||||
Sbjct  424  CCCTACTAAT-GTTACTAACCCTACTAATGTT  394
```

**==> Region Hiv1. Russia. 242 280**

**HIV-1 isolate 07.RU.SP-R497.VI.F5 from Russia envelope glycoprotein (env) gene, complete cds**

**Sequence ID: GU481454.1Length: 2580Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 38.3 bits(41) | 1.6 | 32/39(82%) | 3/39(7%) | Plus/Minus |

```
Query  242   TTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATC  280
              |||||||||||||| ||  |||||||||||||| | |||
Sbjct  1064  TTGTTATTAAAGTATTT---TTTCAATTTTGTACTTATC  1029
```

**==> RegionSivagmTan 279 298**

**Simian immunodeficiency virus partial pol gene for Pol, isolate SIVagmTAN-CM545-pol**
**Sequence ID: LM999945.1Length: 3111Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 32.8 bits(35) | 4.7 | 25/30(83%) | 0/30(0%) | Plus/Minus |

```
Query 269  TTTGTAATGATCCATTTTTGGGTGTTTATT 298
           || ||||| |||| | || ||||||||||||
Sbjct 1098 TTGGTAAAGATCTACTTCTGGGTGTTTATT 1069
```

So, to sum up these 14 HIV SIV signatures, here is Table1:

**Table1** - Synoptic table of 12 significant gene «EIE » motifs from HIV SIV strains in the "A" and "B" regions of the COVID-19 genome.

| Origines | HIV SIV type | Relative Location | « Exogeneous Informative Element » Label | Access | Homology | Bp identities | ORF1ab | Sspike | Real location |
|---|---|---|---|---|---|---|---|---|---|
| Region A 600bp : 21072 to 21672 | | | | | | | | | |
| | | | 266. 21555. Orf1ab. Relative locations 484/600 (end Orf1ab gene), then 492/600 start Spike gene | | | | | | |
| 2012 France | HIV2 | 66-81 | **HIV-2 isolate 56 from France envelope glycoprotein (env) gene, partial cds** | JN230738.1 | 100,00% Unsignificant | 16/16 Unsignificant | X | | 21137 21152 |
| 2017 Sweden | HIV1 | 154-174 | **HIV-1 isolate 060SE from Sweden, partial genome** | MF373163.1 | 100,00% | 21/21 | X | | 21225 21245 |
| 2012 Guinea | HIV2 | 236-253 | **HIV-2 isolate CA65410.13 from Guinea-Bissau envelope gene, partial cds** | JN863831.1 | 94,00% | 17/18 | X | | 21307 21324 |
| 2016 Africa | SIV | 366-386 | **Simian immunodeficiency virus isolate VSAA2001, complete genome** | KR862351.1 | 95,00% | 20/21 | X | | 21437 21457 |
| | | | 21563..25384. S spike | | | | | | |
| 2008 Kenia [9] | HIV1 | 471-501 | **HIV-1 clone ML1592n from Kenya nonfunctional vpu protein (vpu) gene, complete sequence; and nonfunctional envelope glycoprotein (env) gene, partial sequence** | EU875177.1 | 88,00% | 28/32 | X | X | 21542 21572 |
| 2012 Cap verde | HIV2 | 512-529 | **HIV-2 isolate 05HANCV37 from Cape Verde envelope glycoprotein (env) gene, partial cds** | JF26 | 100,00% | 18/18 | | X | 21583 21600 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 7434.1 | | | | |

**Region B : 330bp (21672 to 22002)**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2014 Cote d'ivoire | HIV2 | 23-42 | **HIV-2 isolate 106CP_RT from Cote d'Ivoire reverse transcriptase gene, partial cds** | KJ131112.1 | 95,00% | 19/20 | X | 21694 21713 |
| 2016 Tanzania Partially overlap | SIV | 29-50 | **Simian immunodeficiency virus isolate TAN5 from Tanzania, complete genome** | AF003044.1 | 91,00% | 20/22 | X | 21700 21721 |
| 2016 Africa | SIV | 77-96 | **Simian immunodeficiency virus isolate P18 patient P1, gp120 (env) gene, partial cds** | AF003044.1 | 95,00% | 19/20 | X | 21748 21767 |
| 2016 Netherlands | HIV1 | 85-112 | **HIV-1 isolate 19828.PPH11 from Netherlands envelope glycoprotein (env) gene, partial cds Sequence ID: HQ644953.1** | HQ644953.1 | 89,00% | 25/28 | X | 21756 21783 |
| 1993 côté ivoire | HIV2 | 132-157 | **Human immunodeficiency virus type 2 complete genome from strain HIV-2UC1** | L07625.1 | 85,00% | 22/26 | X | 21803 21828 |
| 2011 Sénégal | HIV2 | 179-194 | **HIV-2 isolate H2A62_111808_CINT_WBC_25 from Senegal pol gene, partial sequence** | JF811228.1 | 100,00% Unsignificant | 16/16 Unsignificant | X | 21850 21865 |
| 2013 Malawi | HIV1 | 212-243 | **HIV-1 isolate 4045_Plasma_Visit1_amplicon9 from Malawi envelope glycoprotein (env) gene, complete cds** | KC187066.1 | 88,00% | 28/32 | X | 21883 21914 |
| 2010 russia | HIV1 | 242-280 | **HIV-1 isolate 07.RU.SP-R497.VI.F5 from Russia envelope glycoprotein (env) gene, complete cds** | GU481454.1 | 82,00% | 32/39 | X | 21913 21951 |
| 2015 Cameroun. | SIV | 279-298 | **Simian immunodeficiency virus partial pol gene for Pol, isolate SIVagmTAN-CM545-pol** | LM999945.1 | 83,00% | 25/30 | X | 21950 21969 |

## 2/ Evidence for 4 others HIV/SIV “EIE” sequences in others areas of COVID-19 genome:

We also found 4 other non-contiguous HIV SIV regions summarized in Table 2 below. The reader will find the details of these searches in the supplementary materials "d".

==> **Region 8751 to 8770, SIV Germany 2015**

> Simian immunodeficiency virus isolate D4 from Germany gag protein (gag) gene, complete cds; pol protein (pol) gene, partial cds; vif protein (vif), vpx protein (vpx), vpr protein (vpr), tat protein (tat), rev

protein (rev), and envelope glycoprotein (env) genes, complete cds; and nef protein (nef) gene, partial cds
Sequence ID: KM378564.1Length: 9051Number of Matches: 1

**Sequence ID: KM378564.1Length: 9051Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 37.4 bits(40) | 3.3 | 20/20(100%) | 0/20(0%) | Plus/Minus |

```
Query  8751  TGATTTTGACACATGGTTTA  8770
             ||||||||||||||||||||
Sbjct  6142  TGATTTTGACACATGGTTTA  6123
```

==> **Region 14340 to 14378 HIV1 China 2016**

**HIV-1 clone XJ47 from China envelope glycoprotein (env) gene, partial cds**

**Sequence ID: EU184986.1Length: 697Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 43.7 bits(47) | 3.7 | 33/38(87%) | 1/38(2%) | Plus/Minus |

```
Query  14358  ATTGTGCAAACTTTAATGTTTTATTCTCTACAGTGTTC  14395
              || |||||||| |||| ||||||||||||| | |||||||
Sbjct  307    ATGGTGCAAATTTTATTGTTTTATTCTCGA-AGTGTTC 271
```

==> **Region 20373 to 20401 HIV1 USA 2004**

**Homo sapiens clone HIV1-H9-106 HIV-1 integration site**

**Sequence ID: AY516986.1Length: 124Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 42.8 bits(46) | 3.7 | 26/28(93%) | 0/28(0%) | Plus/Plus |

```
Query  20404  AATCACCTTTTGAATTAGAAGATTTTAT  20431
              ||||| ||||||||||||||||||| |||
Sbjct  82     AATCAACTTTTGAATTAGAAGATTATAT  109
```

==> **Region 20400 to 20430 HIV1 USA 2011**

**HIV-1 isolate JACH1853_A5 from USA envelope glycoprotein (env) gene, complete cds; and vpu protein (vpu), rev protein (rev), and tat protein (tat) genes, partial cds**

**Sequence ID: HQ217329.1Length: 2598Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 42.8 bits(46) | 3.7 | 28/30(93%) | 1/30(3%) | Plus/Minus |

```
Query  20431  TTCCTATGGACAGTACAGTTAAAAACTATT  20460
              || ||||||||||||||| ||||||||||||
Sbjct  2143   TTACTATGGACAGTACAG-TAAAAACTATT  2115
```

**Table2** - Synoptic table of 4 gene « EIE » motifs from HIV SIV strains in others areas than the "A" and "B" regions of the COVID-19 genome.

| Origines | HIV SIV | Gene | « Exogeneous Label | Informative | Elements » | Access | Homology | Bp identities | O R | S | Real location |
|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | type | | | | | | | F1ab | spike |
| 266. 21555. Orf1ab. | | | | | | | | | |
| 2015 Germany | SIV | POL | | **Simian immunodeficiency virus isolate D4 from Germany gag protein (gag) gene, complete cds; pol protein (pol) gene, partial cds; vif protein (vif), vpx protein (vpx), vpr protein (vpr), tat protein (tat), rev protein (rev), and envelope glycoprotein (env) gene...** | KM378564.1 | 100,00% | 20/20 | X | 8751 8770 |
| 2016 China | HIV1 | ENV | **HIV-1 clone XJ47 from China envelope glycoprotein (env) gene, partial cds** | | EU184986.1 | 87,00% | 33/38 | X | 14340 14378 |
| 2004 USA | HIV1 | INTEGRASE | **Homo sapiens clone HIV1-H9-106 HIV-1 integration site** | | AY516986.1 | 93,00% | 26/28 | X | 20373 20401 |
| 2011 USA | HIV1 | ENV | | **HIV-1 isolate JACH1853_A5 from USA envelope glycoprotein (env) gene, complete cds; and vpu protein (vpu), rev protein (rev), and tat protein (tat) genes, partial cds** | HQ217329.1 | 93,00% | 28/30 | X | 20400 20430 |

Detailed search of these 4 HIV SIV regions is available in supplementary materials « d »

## 3/ Results related to HIV1 Kenya 2008

(see also Discussion § 6 - Evidence for HIV1 and HIV2 sequences in this region and their compaction in 225bp portion of both COVID_19 and Bat coronavirus RaTG13 genomes).

This important HIV1 genome was particularly studied in an HIV vaccine strategy context by Canadian Professor Franck Plummer Lab. Team [9, 10].

This region, in addition to its hundred strong homologies with all the COVID_19 strains of 2020, shows only 2 other homologies with, on the one hand, **Bat coronavirus RaTG13,** and at a lower level, with **Rhinolophus affinis coronavirus isolate LYRa3 spike protein gene**.

The HIV1 Kenia 2008 fingerprint recall :

TGTTTTTATTACTTTTATTGCCACTATTCTCT

Here is the detail of these 3 homologies:

**Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome**
**Sequence ID: NC_045512.2Length: 29903Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 37.4 bits(40) | 8e-04 | 28/32(88%) | 1/32(3%) | Plus/Plus |

```
Query  1      TGTTTTTATTACTTTTATTGCCACTATTCTCT  32
              ||||||| ||  ||||||||||||||| |||||
Sbjct  21568  TGTTTTTCTTG-TTTTATTGCCACTAGTCTCT  21598
```

**Bat coronavirus RaTG13, complete genome**
**Sequence ID: MN996532.1Length: 29855Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 32.8 bits(35) | 0.032 | 27/32(84%) | 1/32(3%) | Plus/Plus |

```
Query  1      TGTTTTTATTACTTTTATTGCCACTATTCTCT  32
              ||||||| ||  ||||||||||||||| | |||
Sbjct  21550  TGTTTTTCTTG-TTTTATTGCCACTAGTTTCT  21580
```

It is observed that there is only one nucleotide of difference (C / T underlined in the sequences above) between the 2 respective sequences of **COVID_19** and **Bat coronavirus RaTG13.**

The first 28 nucleotides (out of 32) are therefore completely identical between these 2 sequences.

**Rhinolophus affinis coronavirus isolate LYRa3 spike protein gene, complete cds**
**Sequence ID: KF569997.1Length: 3913Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 25.6 bits(27) | 4.8 | 15/16(94%) | 0/16(0%) | Plus/Plus |

```
Query  15    TTATTGCCACTATTCT  30
             |||| ||||||||||||
Sbjct  3720  TTATGGCCACTATTCT  3735
```

Homologies from this 3rd sequence are non significant.

The NCBI description of this Bat coronavirus RaTG13 sequence referenced **Bat coronavirus RaTG13, complete genome Sequence ID: MN996532.1**

**please note the difference between collection date (24 jul 2013) and sequencing deposing dates (24 mar 2020).**

**Deposed date is : VRL 24-MAR-2020**

**Meanwhile, Bio Sample collection date date is : 24-Jul-2013**

| | |
|---|---|
| **strain** | Bat coronavirus isolate RaTG13 |
| **collection date** | 24-Jul-2013 |

# DISCUSSION :

Multiple results lead us to think that this genome of COVID-19 has unexpected HIV SIV « EIE »:
Here are the main reasons :
1 – A hight density of HIV SIV regions that are diverse both in their nature and in their age: indeed, a concentration of 12 signifiant HIV SIV sequences in only 744bp.
2 – Concentrations of HIV SIV regions « placed » in sequence and *« side by side »*.
3- Evaluation of the heterogeneity of the regions analyzed in this article.
4 – Invariability study of the 2 regions "A" and "B" for all COVID_19 available on NCBI.
5 - Evidence of the absence of 4 HIV1 HIV2 « EIE » from COVID_19 within SARS-2005 and MERS genomes.
6 - Evidence for HIV1 and HIV2 sequences in this region and their compaction in 225bp portion of both COVID_19 and Bat coronavirus RaTG13 genomes.
7 - Multiple traces of Plasmodium falciparum and Plasmodium vinckei.

8-The determining case of HIV1 Kenya 2008 absent from all coronaviruses other than COVID_19 and RaTG13.

9-First encouraging mutations in the 225b, « A » and « B » regions.

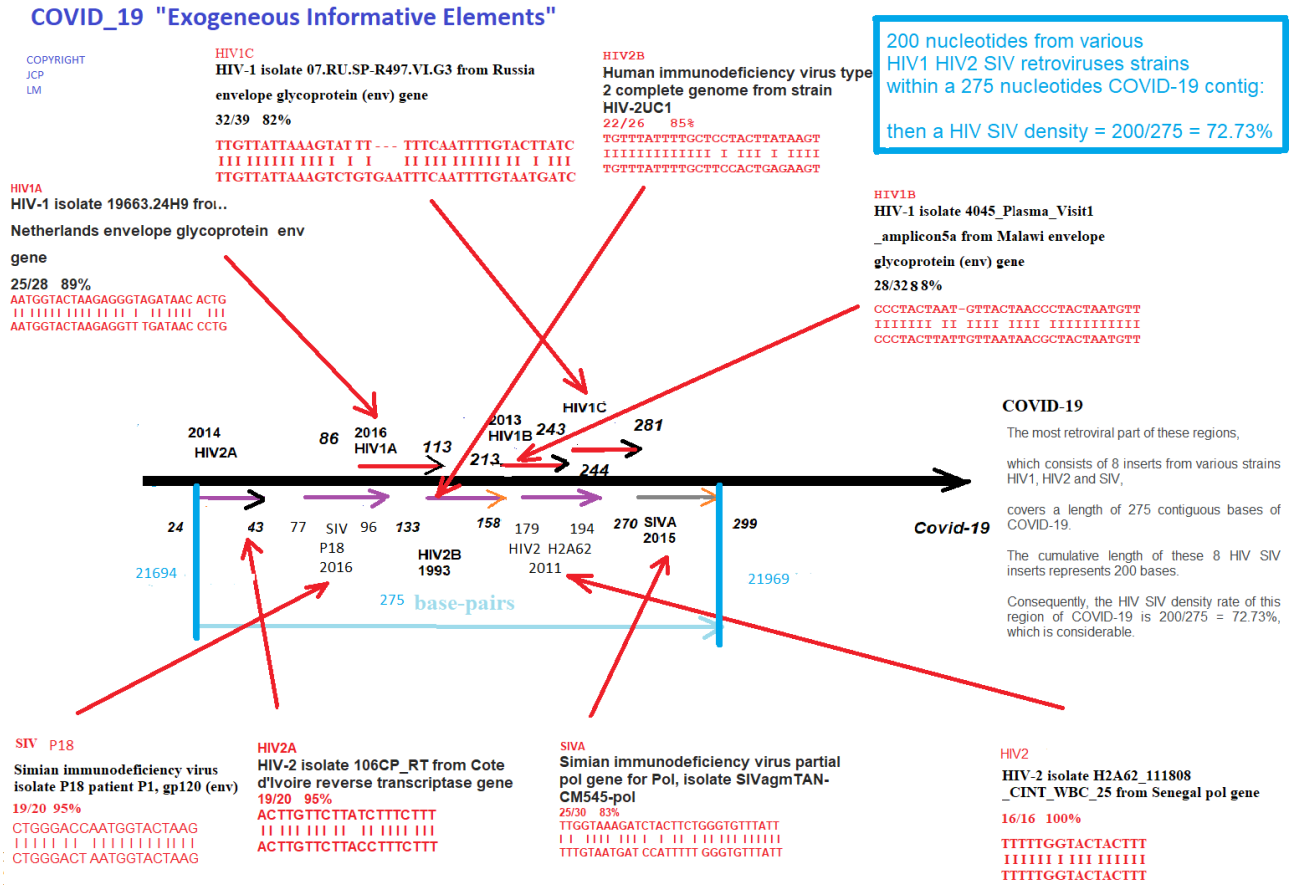
Now, we discuss these 9 different topics :

## 1- A high density of HIV SIV regions that are diverse both in their nature and in their age: indeed, a concentration of 12 significant HIV SIV « EIE » in only 744bp.

First, it is important to note that all the regions found here are included in one of the 2 main genes of COVID_19, so they are **« Informative Exogenous Elements ». A synthetic chart is in Figure1.**
Some significant results relating to this analysed region of 930 base pairs (600 + 330).
The entire genome has 29903 bases. The 12 regions are located between the bases 21225 and 21969, that is to say exactly 744bases.
This therefore represents an average space of 744/12 = 62 base pairs for each « EIE ».
Or as a % of the whole genome 744/29903 = 2.49% of the whole genome.
As the cumulative length of the 12 « EIE » is 305bp, we deduce that the average size of a insert motif is 337/12 = 25.4bases.
Finally, we deduce an occupancy rate of the 744bases space by « EIE » from HIV SIV of 25.4 / 62 = **40.99%**.
This percentage is considerable.

**So, to summarize: a contiguous region representing 2.49% of the whole COVID-19 genome is 40.99% made up of 12 diverse « EIE » originating from various strains of HIV SIV retroviruses.**

**Figure 1** – This summary chart demonstrating visually how 200b from various HIV SIV retroviruses strains within a




concentrated 275b COVID-19 contig have a density rate equal to 72.73%.

## 2- Concatenations of HIV SIV regions "placed" in sequence and *side by side*.

In Table 2, the reader will see that two very different « EIE » follow each other side by side in the RNA sequence of COVID-19:

The first, at addresses 20373 to 20401 comes from an HIV1 Integrase from a USA virus from 2004, while the second, at addresses 20400 to 20430 comes from an Envelope from another HIV1 virus from the USA from 2011 .

Even more surprising, in Table 1, we note the same phenomenon between, this time not 2 but 3 « EIE » from the radically different HIV SIV viruses:
Here are these 3 « EIE » concatenated with it seems *"a watchmaker's precision"*:

Malawi, year 2013.
HIV1 212-243 HIV-1 isolate
4045_Plasma_Visit1_amplicon9 Malawi envelope glycoprotein (approx) 88.00% 28/32
Addresses: 21883 21914

Russia, year 2010.
HIV1 242-280 HIV-1 isolate 07.RU.SP-R497.VI.F5 envelope glycoprotein Russia (env) gene 82.00% 32/39
Addresses: 21913 21951

Cameroon year 2015.
SIV 279-298 partial simian immunodeficiency virus pol gene for Pol, 83.00% 25/30
Addresses: 21950 21969

It will be observed that the cumulative length in COVID_19 of these 3 « EIE » is 126 bases for a number of HIV like bases of 120 bases. Then a total HIV/COVID_19 of 120/126 > 95%, which is remarkable.

## 3- Evaluation of the heterogeneity of the regions analyzed in this article.

This third approach will be more qualitative than the previous two, which were based on formal and quantitative measures.

In [11], we demonstrate the existence of a palindromic-like mirror structure that exists in real genomes and disappears totally in synthetic genome were authors select only main functional genes deleting amont if other non coding regions.
This biomathematic meta-organization is particularly based on characteristic proportions of Fibonacci numbers between DNA single strand nucleotides proportions TC / AG. We suggest that this meta-structure enhances the three-dimensional cohesion of the two DNA strands of the genome. We then generalize this study to the different synthetic genomes and synthetic cells published by the Craig Venter Institute on Mycoplasma Mycoides JCVI-syn1.0 (in 2010), JCVI-syn3.0 (in 2016) and JCVI-syn3A (in 2019). Finally, in the discussion section, we extend this study to synthetic genomes of E-Coli and Yeast chromosome XII.

Here, in the case of COVID-19 the situation is very different since this genome is natural and RNA based, although we suspect a small region to have been added to this genome. In addition, the regions which we believe have been added are themselves natural sequences since they come from different HIV SIV strains.
It will therefore be difficult for our method to demonstrate the partially synthetic nature of this genome, but perhaps we could detect a certain "heterogeneity"?

In [11], the reader will find out how our method of measuring the heterogeneity of a genome works. It uses observation based on the 2 possible partitions of TCAG nucleotide triplets: Partition 1 2 (beginning of the Fibonacci sequence : **1 2 3** 5 8 13 21 34 55 89...) or partition
2 1 (start of the Lucas sequence: **2 1 3** 4 7 11 18 29 47 76...).
We then discovered that these values were perfectly correlated in the case of natural genomes while they strongly dissociated for synthetic genomes [11].
It is this property which will be highlighted here (figs. 3 and 4).
In Fig.2, we verify this property for Fibonacci and Lucas numbers larger pairs such as 89 and 76 (Figure 2).
In these 2 figures, we study the region of COVID-19 between the bases 20000 and 23000. This region includes our 2 regions « A » of 600 then « B » of 330bp, and the region "Lyons Weiler" [4].
We indeed note a strong distance from the 2 curves for the region of 600bp, this would confirm the heterogeneity of this region.
Contrarly, the small region of 330bp seems homogeneous, this could result from the higher density of HIV SIV « EIE » in this region (8 footprints), particularly these 3 EIE glued side by side as specified above in (2).

Note: the 3 arrows in Figures 2 and 3 illustrate this detection of heterogeneity well: region A (600 bases) is heterogeneous because it mixes a coronavirus RNA code with fairly spaced HIV EIE.

Similarly we will show (Discussion, &7) that the "Lyons Weiler" region [4] is also heterogeneous, containing some plasmodium RNA sequences.
On the contrary region B (330 bases), because very dense in HIV EIE (density close to 3 EIE bases within 4 bases of COVID_19) appears, naturally, homogeneous.

Note: in Figures 2 and 3 below, the X axis corresponds to the nucleotic positions in the COVID_19 genome while the Y axis is the variation in the number of occurrences of type, for example (Figure 3) : do +1 when, starting from base n, the 3 following TCAG bases contain 2 bases T or C (blue legend Fib3TCAG / 2TC).

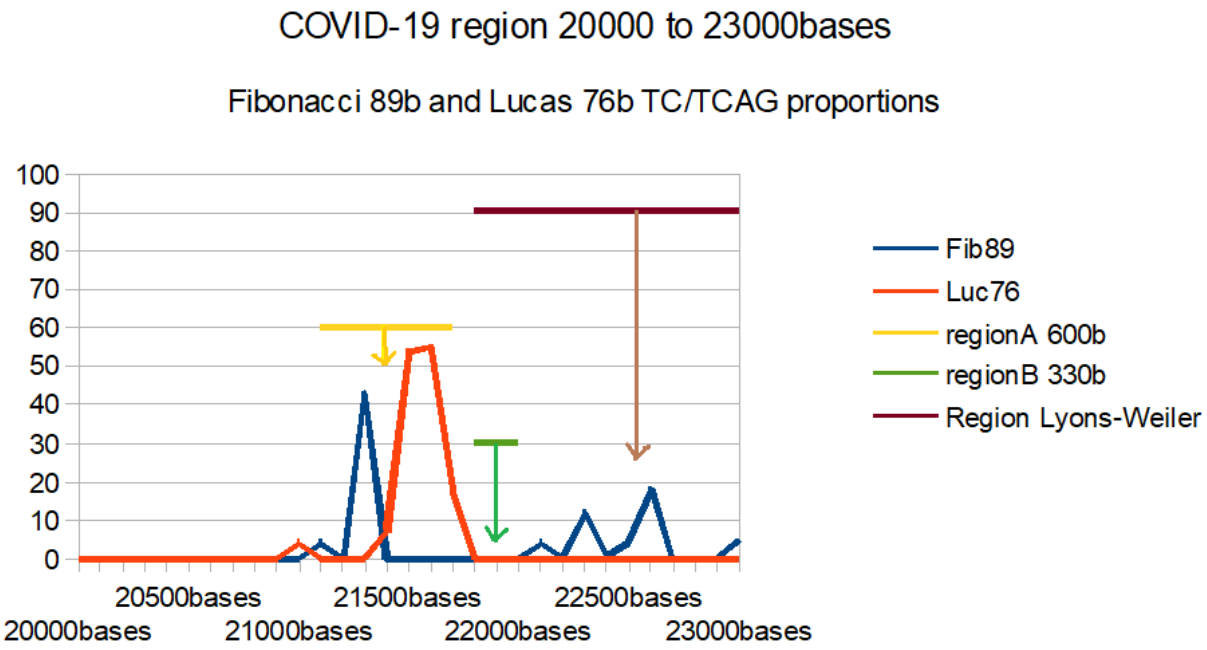



**Figure2** - The Luc76 / Fib89 curves strongly separate in the 2 regions "A" (600bp) and more moderately in "Lyons-weiler" but superimposed in the case of 330 bases region « B » (green).

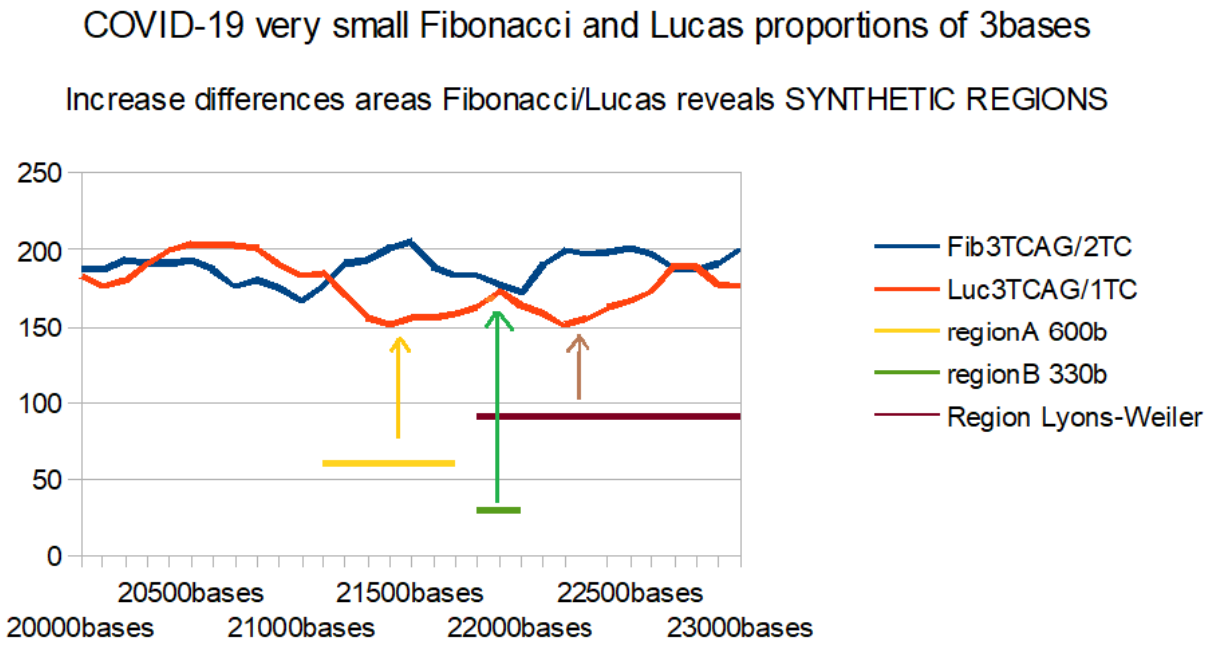


**Figure3** - The Luc3TCAG/2TC and Fib3TCAG/2TC curves strongly separate in the 2 regions "A" (600 bases) and in "Lyons-weiler", but superimposed in the case of 330 bases region B » (green).

## 4 – Conservation of the 2 regions "A" and "B" for all COVID_19 available at NCBI until 25 March 2020 :

BLAST research on regions "A" (600bases) and "B" (330bases) show us a very low variability of all (about hundreds available on NCBI BLASTn on 25 March 2020) strains of COVID_19 referenced by GenBank., for example from BLASTn queries in March 2020:

**We find 94/99 occurrences with 100% homology for region "A".**

Only 4 have some small mutations (1base, then <1%), and one has 25bases error (575/600)see supplementary materials « e ».

**We find 94/99 occurrences with 100% homology for region "B".**

Only 3 have some small mutations (1base, then <1%), another has 14bp differences (316/330bp), one (wuhan market, only 0 bases on 330bases) has a very poor homology (0%), see supplementary materials1 « e », particularly the details on this sequence: «Coronovirus sequence assembly isolated from 2019/2020 Wuhan outbreak patient».

Full details in supplementary materials1 « e ».

## 5 - Evidence of the absence of 4 HIV/SIV « Exogeneous Informative Elements » from COVID_19 within SARS-2005 and MERS genomes.

Here we wanted to find out if the 16 “EIE” discovered in the COVID-19 genome already existed in the human SARS genomes that appeared in 2003.

**Table 3 below summarizes this research. In particular, it appears that 14 of the 18 HIV SIV “EIE” existed - already - from the first human SARS genomes that appeared in China around 2003.**

However, **a novel long region of around 225 nucleotides**, less than 1% of the genome, appears to us to have been inserted: this region is completely absent in ALL SARS genomes, whereas it is present and 100% homologous for all COVID-19 genomes listed in NCBI.

This region is located (in the COVID-19 genome which served as a reference) between the addresses 21550 and 21772. It is therefore located between the end of region "A" (475 to 600 bases locations) and the start of region "B" (1 to 99 bases locations).

A remarkable fact is also observed: the HIV/SIV “EIE” which already existed in SARS have evolved a lot through numerous mutations. Thus, 4 “EIE” have very weak homologies (near 30%) between their SARS version and their COVID-19 version. These homologies gradually improve in more recent SARS (2015 or 2017 for example, right column in Table 4).

**The 4 « Exogeneous Informative Elements » added in COVID_19 are respectively:**

**HIV1 Kenia 2008**

**HIV2 Cap verde 2012**

**HIV2 Ivory Coast 2014**

**SIV Africa 2016.**

The reader will be able to note that these strains HIV1 HIV2 SIV are very recent and subsequent to the emergence of SARS. However, most of the other strains - which appear to us present in SARS - have dates after the emergence of the first SARS. This fact will have to be explained …

Full details in supplementary materials « f ».

**Table 3** – Comparing the 16 « EIE » from « A », « B » and remaining regions in COVID-19, HIV/SIV and SARS.

| HIV/SIV « Exogeneous Informative Elements » | Locations within « A » 600bases and « B » 330bases regions | Length nucleotides in COVID_19 | Length nucleotides in HIV-SIV « EIE » % HIV-SIV / COVID_19 | Length nucleotides in SARS genomes % SARS/COVID_19 |
|---|---|---|---|---|
| | | Region « A » | | |
| HIV2 2012 France | 66-81 | 16 unsignificant | 16 100% | 13 81% |

| | | | | | | |
|---|---|---|---|---|---|---|
| HIV1 2017 Sweden | 154-174 | 21 | 21 | 100% | 19 | 90% |
| HIV2 2012 Guinea | 236-253 | 18 | 17 | 94% | 11 | 61% |
| SIV 2016 Africa | 366-386 | 21 | 20 | 95% | 18 | 86% |
| **Start 225bases zone including 4 « Exogeneous Informative Elements »** | | | | | | |
| **HIV1 2008 Kenia** | **471-501** | **32** | **28** | **88%** | **0** | **0%** |
| **HIV2 2012 Cap verde** | **512-529** | **18** | **18** | **100%** | **0** | **0%** |
| **Region « B »** | | | | | | |
| **HIV2 2014 Cote d'ivoire** | **23-42** | **20** | **19** | **95%** | **0** | **0%** |
| **SIV 2016 Africa** | **77-96** | **20** | **19** | **95%** | **0** | **0%** |
| **End 225bases « EIE » zone including 4 « Exogeneous Informative Elements »** | | | | | | |
| HIV1 2016 Netherlands variant HIV1 USA 2011 | 85-112<br>85-108 | 28 | 25 | 89% | 13<br>9 | 46%<br>32% |
| HIV2 1993 côte ivoire | 132-157 | 26 | 22 | 85% | 20 | 77% |
| HIV2 2011 Sénégal | 179-194 | 16 unsignificant | 16 | 100% | 12 | 75% |
| HIV1 2013 Malawi | 212-243 | 32 | 28 | 88% | 22 | 69% |
| HIV1 2010 russia | 242-280 | 39 | 32 | 82% | 15 | 38% |
| SIV 2015 Cameroun. | 279-298 | 30 | 25 | 83% | 10 | 33% |
| others areas than the "A" and "B" regions | | | | | | |
| SIV 2015 Germany | 8751<br>8770 | 20 | 20 | 100% | 9 | 45% |
| HIV1 2016 China | 14340<br>14378 | 38 | 33 | 87% | 34 | 89% |
| HIV1 2004 USA | 20373<br>20401 | 28 | 26 | 93% | 28 | 100% |
| HIV1 2011 USA | 20400<br>20430 | 30 | 28 | 93% | 21 | 70% |

## The other case of MERS genome :

An analysis of the reference genome of the pathogenic RNA virus MERS ( **Middle East respiratory syndrome coronavirus, complete genome** NCBI Reference Sequence: NC_019843.3 , https://www.ncbi.nlm.nih.gov/nuccore/NC_019843.3?report=genbank ) shows that the end of our "A" region, all of the key 225 base regions, of the "B" region and of the "Lyons-weiler" region. 4 crucial regions of our article are totally ABSENT in MERS.

## 6 - Evidence for HIV/SIV sequences in this region and their compaction in 225 bases portion of both COVID_19 and Bat coronavirus RaTG13 genomes.

We now analyze the level of homologies between the 4 strains HIV/SIV of the 4 cases which are always present in COVID_19 but always absent in SARS. **Detailed results are in §Results (3).**

**The remarkable point is as follows: It is strange that the most significant "Bat" genome, Bat coronavirus RaTG13 genome** [12]**, is from 2020, just like COVID_19 ... In particular, for the HIV1 Kenia 2008 sequence[9, 10], there remains the one and only strain found in the "Bat" population, while for the 3 other EIE, the "Bat" strains are very numerous but with non-significant HIV/SIV homologies.**

**Table 4** – Comparing the 4 « EIE » from COVID-19, HIV/SIV and Bat coronavirus RaTG13 [12].

| HIV/SIV « Exogeneous Informative Elements » | Locations within « A » 600bases and « B » 330bases regions | Length nucleotides in COVID_19 | Length nucleotides in HIV/SIV « EIE» | | Length nucleotides in **Bat coronavirus RaTG13** genome | | |
|---|---|---|---|---|---|---|---|
| | | Region « A » | | | | | |
| **2008 Kenia HIV1** | **471-501** | **32** | **28** | **88%** | **27** | **84%** | **(note1)** |
| 2012 Cap verde HIV2 | 512-529 | 18 | 18 | 100,00% | 16 | 89% | (note2) |
| | | Region « B » | | | | | |
| 2014 Cote d'ivoire HIV2 | 23-42 | 20 | 19 | 95% | 15 | 79% | (note3) |
| 2016 Africa SIV | 77-96 | 20 | 19 | 95% | 10 | 53% | (note4) |

Note1

COVID_19 / HIV1 28/32 88%, Only both non COVID_19 strains:

Bat coronavirus RaTG13 and Rhinolophus affinis coronavirus isolate LYRa3 spike protein gene. No others Bat strains.

Note2
COVID_19 / HIV2 18/18 100%, Bat. 16/18. 89%, Sars urbani. 10/10
Various others Bat and sArs with VERY low homologies but all < 10

Note3
COVID_19 / HIV2 19/20 95%, Bat RaTG13. 15/17. 88%. well. Sars urbani. 9/9
Various others Bat and sArs but all <12

Note4

COVID_19 / SIV. 19/20. 95%, Bat coronavirus RaTG13 Hiv, Bat. 10/10. Bad homology.
Various Bat and sArs all <12

Full details in supplementary materials « g ».

**Zooming on the first HIV1 Kenia Homologies :**

Full BLASTn data is in Results.

Synthesis data : Comparing the 3 key regions « A », « B », and « Lyons-Weiler » region [4] in the cases of COVID-19, Bat RaTG13 coronavirus [12] and the best homologies for other Bat and SARS coronaviruses.

**Table 5** – Comparing the 3 key regions « A », « B », and « Lyons-Weiler » region [4] in the cases of COVID-19, Bat RaTG13 coronavirus [12] and the best homologies for other Bat and SARS coronaviruses.

| Coronavirus genome | Region « A » | | Region « B » | | Region « Lyons-weiler » | |
|---|---|---|---|---|---|---|
| COVID_19 | 600/600 | 100% | 330/330 | 100% | 1378/1378 | 100% |
| **Bat RaTG13** | **563/599** | **98%** | **309/330** | **94%** | **1209/1311** | **92%** |
| Other Bat | 518/605 | 86% (note1a) | 158/212 | 75% (Note1b) | 402/521 | 77% (Note1c) |
| Other SARS | 400/474 | 84% (note2a) | 144/177 | 73% (Note 2b) | 297/376 | 79% (Note2c) |

See supplementary materials « f » for details.

Note1a - **Bat SARS-like coronavirus isolate bat-SL-CoVZC45**
Note1b - **BtRs-BetaCoV/YN2013, complete genome**
Note 1c - **Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome**
Note2a - **SARS coronavirus GZ0402, complete genome**
Note 2b - **SARS coronavirus isolate CFB/SZ/94/03, complete genome**
Note2c - **SARS coronavirus SZ3, complete genome**

## 7 - Multiple traces of Plasmodium falciparum and Plasmodium vinckei.

In our analysis of suspected heterogeneous regions in the COVID-19 genome, Figures 2 and 3 highlight a second area located in the region referenced "Lyons Weiler" region [4]. We observe a peak in this region which would be located after our region of 330 bases referenced "B". This suspected region would be between 22,000 and 23,000 bases adresses within the COVID-19 genome. If we do not find traces of HIV SIV in this region, we discover 2 possible « EIE » from **Plasmodium Falciparum**, genome of Malaria agent disease. Here are the details:
Nota : adresses of Query are relative adresses within Lyons-Weiler region, then from location 21672 bases in reference COVID-19 genome.

**The first area is located from 446 to 481 bases in « Lyons-weiler region » (36 bases, 86% homology).**
**The second area is located from 969 to 997bp in « Lyons-weiler region » (29 bases, 90% homology).**

Reader find full details in supplementary materials2 « h ».

On the other hand, a complementary analysis reveals multiple traces of regions with strong homologies with Plasmodium vinckei in the 2 regions "A" (600 bases) and "B" (330 bases) of COVID_19.
Finally, we were able to verify the absence of these 2 traces of Plasmodium falciparum in the SARS strains. However, they are présent in Bat CORONAVIRUS strains.

## 8-The determining case of HIV1 Kenya 2008 absent from all coronaviruses other than COVID_19 and RaTG13.

Please, see BLASTn main data in RESULTS§ (ref 3).
The BLASTn analysis on April 10, 2020 option "SARS coronaviruses taxid 694009" reports 386 occurrences including 16 bats and 2 Rhinolophus, and 368 COVID_19.
Reader find full detailed data results in supplemetary materials2 ref « h ».

Then we have summarized these results in the following Table 6.

**Table 6** – Dates of collection then deposit of various « EIE » involved in the 225 bases region.

| Strain reference | Sequence ID: | 225 bases Homologies | | Date collect | Date sequencing |
|---|---|---|---|---|---|
| **Wuhan reference** | LR757998.1 | **225/225 100%** | | ***26 dec 2019*** | ***6 mar 2020*** |
| **Bat RaTG13** | MN996532.1 | **204/225** | **91%** | ***24 jul 2013*** | ***24mar2020*** |
| **Bat SARS-like coronavirus isolate bat-SL-CoVZC45** | MG772933.1 | **71/96 (note1)** | **32%** | **Feb 2017** | **5 feb 2020** |
| **Bat SARS-like coronavirus isolate bat-SL-CoVZXC21** | MG772934.1 | 45/54 | 45% | Jul 2015 | 5 feb 2020 |
| **Rhinolophus affinis coronavirus isolate LYRa3 spike protein gene** | KF569997.1 | 21/21 | 9% | 2011 | 29may2014 |
| **Rhinolophus affinis coronavirus isolate LYRa11** | KF569996.1 | 21/21 | 9% | 2011 | 29may2014 |
| **BtRf-BetaCoV/SX2013** | KJ473813.1 | 55/80 | 24% | 2013 | 7jul2017 |

| | | | | | |
|---|---|---|---|---|---|
| **BtRf-BetaCoV/HeB2013** | KJ473812.1 | 55/80 | 24% | 2013 | 7 jul 2017 |
| **Bat coronavirus (BtCoV/273/2005)** | DQ648856.1 | 55/80 | 24% | 2005 | 19 jul 2006 |
| **Bat SARS coronavirus Rf1** | DQ412042.1 | 55/80 | 24% | 2005 | 13 jul 2006 |
| **Bat SARS-like coronavirus isolate Rf4092** | KY417145.1 | 17/17 | 8% | 18 sep 2012 | 18 dec 2017 |
| **BtRf-BetaCoV/JL2012** | KJ473811.1 | 17/17 | 8% | 2012 | 7 jul 2017 |
| **Bat SARS-like coronavirus isolate Rs9401** | KY417152.1 | 26/30 | 12% | 16 oct 2015 | 18 dec 2017 |
| **Bat SARS-like coronavirus isolate Rs7327** | KY417151.1 | 26/30 | 12% | 24 oct 2014 | 18 dec 2017 |
| **Bat SARS-like coronavirus isolate Rs4084** | KY417144.1 | 26/30 | 12% | 18 sep 2012 | 18 dec 2017 |
| **Bat SARS-like coronavirus WIV1** | KF367457.1 | 26/30 | 12% | Sep 2012 | 6 nov 2013 |
| **Bat SARS-like coronavirus WIV1 spike protein (S) gene** | KC881007.1 | 26/30 | 12% | 18 sep 2012 | 22 nov 2013 |
| **Bat SARS-like coronavirus Rs3367** | KC881006.1 | 26/30 | 12% | 19 mar 2012 | 22 nov 2013 |
| **Bat SARS-like coronavirus RsSHC014** | KC881005.1 | 26/30 | 12% | 17 apr 2011 | 22 nov 2013 |

Note1 : in this case, a 225bases region is reduced to contiguous 96 bases.

## Location of the EIE HIV1 Kenya 2008 against the Spike gene:

Firstly, the EIE regions of HIV1 Kenya 2008 nonfunctional ( **Sequence ID:** EU875177.1 ) and of HIV1 Kenya real ( **Sequence ID:** FJ623481.1 ) are identical while the respective Gp120 genes are only 82% homologous: 494/603 (82%) .

**HIV-1 isolate 06KECst_005 from Kenya, complete genome**
**Sequence ID:** FJ623481.1**Length: 8766Number of Matches: 1**

Range 1: 5192 to 5794

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 595 bits(659) | 6e-168 | 494/603(82%) | 3/603(0%) | Plus/Plus |

In other hand, The HIV1 Kenya EIE nonfunctional region from COVID_19 genome is located overlapping between the end of the "orf1ab" gene and the start of the "S spike" gene :

**details COVID_19 genes : orf1ab Spike**
**266----------21555 21563----------25384**
**HIV1Kenya 2008 : 21542----------21572**

COVID_19 Wuhan market **ID:** LR757998.1 reference genome location of EIE Kenya 2008 HIV1 : 21542-21572 bases.

Spike gene location:
21563-25384 bases.

So, in terms of amino acids:
START address of HIV1 KENYA: 21 amino acids before SPIKE begins.

END address of HIV1 KENYA: 9 amino acids after the beginning of SPIKE.

Finally, how is this same question in the case of bat RaTG13 genome?

Locations of HIV1 Kenya within Bat RaTG13 **Sequence ID: MN996532.1**
is: 21550 **TGTTTTTCTTG-TTTTATTGCCACTAGTTTCT** 21580
(see RESULTS§ ref 3).

**Location of Spike gene within BatRaTG13 is:**
21545..25354
/gene="S"
/codon_start=1
/product="spike glycoprotein"
/protein_id="QHR63300.2"

So, in terms of amino acids:
START address of HIV1 KENYA: 6 amino acids after SPIKE begins.
END address of HIV1 KENYA: 36 amino acids after the beginning of SPIKE.

**So, unlike COVID_19 where HIV1 Kenya starts before the start of the SPIKE gene, here, in the case of bat RaTG13, HIV1 Kenya is entirely contained within the SPIKE gene.**

## 9- First encouraging mutations in the 225 bases, « A » and « B » regions.

We must recall here that the BLASTn analysis on April 10, 2020 option "SARS coronaviruses" reports 386 occurrences including 16 bats, 2 Rhinolophus, and 368 COVID_19. The same research running on 16 april 2020 reveals 523 strains sequences. The number of COVID_19 sequences available is therefore constantly changing principally due to USA new sequences deposits.

We were interested in the first cases of significant COVID_19 mutations in this key region of 225 bases (homologies of the order of 96%). we find 5 of them located in the BLASTn just in front of and near RaTG13, all come from the USA, taken and sequenced in April 2020, pathogenic.

A BLASTn analysis dated April 11, 2020 produces the following results:

386 sequences in total. whose:

351 strains with full 100% homology with 225 bases region.

**17 strains with mutations in 225 bases region.**

18 strains bat.

Now let's look at these 17 cases of mutations in the 220 bases region.

**Table 7** – Mutations in region 225 bases.

| Strain number | Strain reference | Mutations relatives adresses within 225 bases region | Homologies | HIV1/SIV « EIE » note1 | Collection and deposit dates |
|---|---|---|---|---|---|
| 1 USA | **SARS-CoV-2/WA-UW381/human/2020/USA, partial genome Sequence ID: MT263460.1** | 8 C/T | 224/225 99.6% | HIV1 Kenya 2008 | 30 mar 2020 6 apr 2020 |
| 2 USA | **SARS-CoV-2/WA-UW334/human/2020/USA, complete genome Sequence ID: MT263414.1** | 8 C/T | 224/225 99.6% | HIV1 Kenya 2008 | 24 mar 2020 06 apr 2020 |
| 3 USA | **ARS-CoV-2/WA-UW301/human/2020/USA, complete genome Sequence** | 81 C/T | 224/225 99.6% | | 23 mar 2020 06 apr 2020 |

| | | | | | |
|---|---|---|---|---|---|
| | **ID: MT263384.1** | | | | |
| 4 USA | **SARS-CoV-2/WA-UW270/human/2020/USA, partial genome Sequence ID: MT259262.1** | 79 C/T | 224/225 99.6% | | 13 mar 2020 06 apr 2020 |
| 5 USA | **SARS-CoV-2/WA-UW257/human/2020/USA, complete genome Sequence ID: MT259249.1** | 157 G/C | 224/225 99.6% | | 13 mar 2020 6 apr 2020 |
| 6 USA | **SARS-CoV-2/WA-UW231/human/2020/USA, complete genome Sequence ID: MT246488.1** | 8 C/T | 224/225 99.6% | HIV1 kenya 2008 | 14 mar 2020 06 apr 2020 |
| 7 USA | **SARS-CoV-2/WA-UW204/human/2020/USA, complete genome Sequence ID: MT246461.1** | 8 C/T | 224/225 99.6% | HIV1 kenya 2008 | 13 mar 2020 06 apr 2020 |
| 8 China | **SARS-CoV-2/KMS1/human/2020/CHN, complete genome Sequence ID: MT226610.1** | 217 T/A | 224/225 99.6% | SIV Africa 2016 | 20 jan 2020 06 apr 2020 |
| 9 Finland | **CoV-FIN-29-Jan-2020, partial genome Sequence ID: MT020781.2** | 140 C/T | 224/225 99.6% | | 29 jan 2020 17 mar 2020 |
| 10 China | **SARS-CoV-2/Yunnan-01/human/2020/CHN, complete genome Sequence ID: MT049951.1** | 77 T/A | 224/225 99.6% | | 17 jan 2020 06 apr 2020 |
| 11 USA | **2019-nCoV/USA-CA5/2020, complete genome Sequence ID: MT027064.1** | 140 C/T | 224/225 99.6% | | 24 mar 2020 06 apr 2020 |
| 12 USA | **SARS-CoV-2/WA-UW302/human/2020/USA, partial genome Sequence ID: MT263385.1** | 175-176 CA/NN 164-166 CCT/NNN | 220/225 97.7% | | 23 mar 2020 6 apr 2020 |
| 13 USA | **SARS-CoV-2/WA-UW356/human/2020/USA, complete genome Sequence ID: MT263436.1** | 188-196 TTCCATGCT/ NNNNNNNNN | 216/225 96% | HIV2 cote d'ivoire 2014 | 24 mar 2020 06 apr 2020 |
| 14 USA | **SARS-CoV-2/WA-UW351/human/2020/USA, complete genome Sequence ID: MT263431.1** | 189-197 TTCCATGCTA /NNNNNNNNN | 216/225 96% | HIV2 cote d'ivoire 2014 | 24 mar 2020 06 apr 2020 |
| 15 USA | **SARS-CoV-2/WA-UW287/human/2020/USA, complete genome Sequence ID: MT259277.1** | 189-197 TCCATGCTA/ NNNNNNNNN | 216/225 96% | HIV2 cote d'ivoire 2014 | 15 mar 2020 06 apr 2020 |
| 16 USA | **SARS-CoV-2/WA-UW306/human/2020/USA, partial genome Sequence ID: MT263389.1** | 145-191 46 del | 144/144 100% then 34/34 | | 23 mar 2020 06 apr 2020 |
| 17 China | **Wuhan seafood market pneumonia virus genome assembly, chromosome: whole_genome Sequence ID: LR757997.1** | 106-225 120 del | 1-105 100% | HIV2 cote d'ivoire 2014 and SIV Africa 2016 | 31 dec 2019 06 mar 20209 |

**17 COVID-19 different stains ===> 5 different « IEE » HIV/SIV**

Note1 : when the mutation is in HIV/SIV insert, we note the strain ref.

We observe that out of these 17 cases of mutations, the majority of them (13/17) concern the USA with dates posterior to the Chinese origin of the pandemic. Only 3 relate to China and one to Finland. There is probably the beginning if a mutations strategy of the genome to balance and integrate exogeneous HIV « EIE ».

On the other hand 9 of these 17 mutations directly affect an HIV / SIV region. The others affect the intermediate region separating the 2 and 2 HIV / SIV pools.

Thirdly, there are also deletions of whole « EIE » which is characteristic of RNA viruses.

It will also be noted that the majority of these strains come from recent samples (12/17 have dates of collection posterior or equal to March 2020). These dates would therefore correspond to a "mature" period of the COVID_19 genomes, which have now entered a phase of diversified mutations.

Finally, we observe the repetition of several mutations, proof if a robust mutations strategy process which eliminates the hypothesis of sequencing errors.

All detailed sequences and BLASTn are available in supplementary materials « k ».

**Table 8** – Comparing 225b region significative mutations § deletions % with whole genomes mutations § deletions %.

| Strain number | Strain reference | Mutations relatives adresses within 225 bases region | **Homologies region 225b / same region in reference genome LR757998.1 and mutations rate %** | **Homologies whole genomes / whole reference genome LR757998.1 and mutations rate %** | HIV1/SIV «EIE» | Collection and deposit dates |
|---|---|---|---|---|---|---|
| 12 USA | **SARS-CoV-2/WA-UW302/human/2020/USA, partial genome Sequence ID: MT263385.1** | 175-176 CA/NN 164-166 CCT/NNN | 220/225 **97.7% 2.222222%** | 29517/ 29598 = 81 **99.726333 % 0.273667%** | | 23 mar 2020 6 apr 2020 |
| 13 USA | **SARS-CoV-2/WA-UW356/human/2020/USA, complete genome Sequence ID: MT263436.1** | 188-196 TTCCATGCT/ NNNNNNN NN | 225-9 = 216 **96% 4.000000%** | 29828/ 29846 = 18 **99.939690 % 0.060309%** | HIV2 cote d'ivoire 2014 | 24 mar 2020 06 apr 2020 |
| 14 USA | **SARS-CoV-2/WA-UW351/human/2020/USA, complete genome Sequence ID: MT263431.1** | 189-197 TTCCATGCT A/NNNNNN NNN | 225-9 = 216 **96% 4.000000%** | 29834/ 29852 = 18 **99.939702 % 0.060297%** | HIV2 cote d'ivoire 2014 | 24 mar 2020 06 apr 2020 |
| 15 USA | **SARS-CoV-2/WA-UW287/human/2020/USA, complete genome Sequence ID: MT259277.1** | 189-197 TCCATGCTA /NNNNNNN NN | 225-9 = 216 **96% 4.000000%** | 29843/ 29866 = 23 **99.922989 % 0.077011%** | HIV2 cote d'ivoire 2014 | 15 mar 2020 06 apr 2020 |
| 16 USA | **SARS-CoV-2/WA-UW306/human/2020/USA, partial genome Sequence ID: MT263389.1** | 145-191 46 del | 225-179 = 46 **79.5555% 20.44444%** | 29517/ 29598 = 81 **99.726332 % 0.273667%** | | 23 mar 2020 06 apr 2020 |
| 17 China | **Wuhan seafood market pneumonia virus genome** | 106-225 120 del | 225-105 =120 | 19263/29388 = 10125 | HIV2 cote | 31 dec 2019 06 mar |

| | | | | |
|---|---|---|---|---|
| **assembly, chromosome: whole_genome Sequence ID: LR757997.1** | **46.6666% 53.333333%** | **65.547162 % 34.452838%** | d'ivoire 2014 and SIV Africa 2016 | 20209 |

In Table 8, results involving 6 significant genomes show a great average mutations level in each 225bases regions ( **13.5687%** ) than in their relating whole genomes ( **0.3496%** ). Then a ratio between average rate mutations region 225 bases and average rate mutations whole genome = **38.813,** due principally to the wuhan market hyper deleted genome **LR757997.1**

Note : last line ref17 China has a lot of deleted or « N » regions : 19263 TCAG nucleotides on 29470 length, then 10207 nucleotides deletions or undetermined nucleotides regions.

**Table 9** – Region « A » interesting mutations.

| Strain number | Strain reference | Mutations relatives adresses within 600 bases region | Homologies | HIV1/SIV « EIE » | Collection and deposit dates |
|---|---|---|---|---|---|
| 1 China Yunnan | **SARS-CoV-2/Yunnan-01/human/2020/CHN, complete genome Sequence ID: MT049951.1** | 547 T/A | 1-600 99% | | 17 jan 2020 6 apr 2020 |
| 2 China Wuhan | **Severe acute respiratory syndrome coronavirus 2 isolate WIV05, complete genome Sequence ID: MN996529.1** | 40 A/G | 1-600 99% | | 30 dec 2019 18 mar 2020 |
| 3 China Wuhan | **Severe acute respiratory syndrome coronavirus 2 isolate WIV02, complete genome Sequence ID: MN996527.1** | 219 G/A | 1-600 99% | | 30 dec 2019 18 mar 2020 |
| 4 USA 1st cases Minnesota | **Severe acute respiratory syndrome coronavirus 2 isolate USA/MN3-MDH3/2020, complete genome Sequence ID: MT188339.1** | 50 T/C 89 C/T | 1-600 98% | **HIV2 env 66-81 France 2012** | 9 mar 2020 11 mar 2020 |
| 5 USA Minnesota | **Severe acute respiratory syndrome coronavirus 2 isolate USA/MN1-MDH1/2020, complete genome Sequence ID: MT188341.1** | 287-289 ---/CTT | 1-600 99% | | 5 mar 2020 13 mar 2020 |
| 6 China Wuhan | **Wuhan seafood market pneumonia virus genome assembly, chromosome: whole_genome Sequence ID: LR757997.1** | Delete 576-600 | 1-575 100% | within region 225 bases | 31 dec 2019 6 mar 2020 |
| 7 Spain Valencia | **SARS-CoV-2/Valencia9/human/2020/ESP, partial genome Sequence ID: MT256917.1** | Delete 50-289<br>325 G/N | 1-49 100%<br>310/311 99% | **HIV2 env 66-81 France 2012**<br>**HIV1 154-174 Sweden 2017** | 2 mar 2020 6 apr 2020 |
| 8 Spain | **SARS-CoV-** | Delete | 1-49 100% | **HIV2 env** | 6 mar 2020 |

| | | | | | |
|---|---|---|---|---|---|
| Valencia | **2/Valencia10/human/2020/ESP, partial genome**<br>**Sequence ID: MT256918.1** | 50-289<br><br>324-325<br>AG/NN<br>327<br>G/N | 290-600<br>308/311 99% | **66-81**<br>**France**<br>**2012**<br><br>**HIV1 154-174**<br>**Sweden**<br>**2017** | 6 apr 2020 |

**8 COVID-19 different strains ===> 5 different « IEE » HIV/SIV**

Full details in supplementary materials2 « k »

**Table 10** – Region « B » interesting mutations.

| Strain number | Strain reference | Mutations relatives adresses within 330 bases region | Homologies | HIV1/SIV « EIE » | Collection and deposit dates |
|---|---|---|---|---|---|
| 1 China Yunnan | **SARS-CoV-2/KMS1/human/2020/CHN, complete genome**<br>**Sequence ID: MT226610.1** | 87<br>T/A | 1-330 99% | **SIV P18 77-96** | 20 jan 2020<br>6 apr 2020 |
| 2 Finland | **Severe acute respiratory syndrome coronavirus 2 isolate nCoV-FIN-29-Jan-2020, partial genome**<br>**Sequence ID: MT020781.2** | 10<br>C/T | 1-330 99% | | 29 jan 2020<br>17 mar 2020 |
| 3 USA Seatle WA | **SARS-CoV-2/WA-UW387/human/2020/USA, partial genome**<br>**Sequence ID: MT263466.1** | 291-293<br>TGT/--- | 1-330 99% | **SivagmTan 279 298** | 23 mar 2020<br>6 apr 2020 |
| 4 India Kerala state | **SARS-CoV-2/29/human/2020/IND, complete genome**<br>**Sequence ID: MT012098.1** | 294-296<br>TTA/--- | 1-330 99% | **SivagmTan 279 298** | 27 jan 2020<br>6 apr 2020 |
| 5 USA Seatle WA | **SARS-CoV-2/WA-UW302/human/2020/USA, partial genome**<br>**Sequence ID: MT263385.1** | 34-36<br>CCT/NNN<br>45-46<br>CA/NN | 1-330 99% | **HIV2 cote d'ivoire 24-43** | 23 mar 2020<br>6 apr 2020 |
| 6 USA Seatle WA | **SARS-CoV-2/WA-UW329/human/2020/USA, partial genome**<br>**Sequence ID: MT263409.** | 291-298<br>TGTTTATT/NNNNNNNN | 322/330 98% | **SivagmTan 279 298** | 22 mar 2020<br>6 apr 2020 |
| 7 USA Seatle WA | **SARS-CoV-2/WA-UW356/human/2020/USA, complete genome**<br>**Sequence ID: MT263436.1** | 58-66<br>TTCCATGCT/NNNNNNNNN | 321/330 98% | **SIV P18 77-96** | 24 mar 2020<br>6 apr 2020 |
| 8 Spain Valencia | **Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/Valencia10/human/2020/ESP, partial genome**<br>**Sequence ID: MT256918.1** | Delete 303-330 | 302/302 100% | | 6 mar 2020<br>6 apr 2020 |
| 9 USA Seatle WA | **SARS-CoV-2/WA-UW278/human/2020/USA, partial** | Delete 292-330 | 291/291 100% | **SivagmTan 279 298** | 16 mar 2020<br>6 apr 2020 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **genome**<br>**Sequence ID: MT259270.1** | | | | | |
| 10 USA Seatle WA | **SARS-CoV-2/WA-UW246/human/2020/USA, partial genome**<br>**Sequence ID: MT259238.1** | Delete 263-330 | 262/262 | 100% | **SivagmTan 279 298** | 16 mar 2020<br>6 apr 2020 |
| 11 USA Seatle WA | **ARS-CoV-2/WA-UW289/human/2020/USA, partial genome**<br>**Sequence ID: MT259279.1** | Delete 246-330 | 245/245 | | **Hiv1. Russia. 242 280**<br><br>**SivagmTan 279 298** | 15 mar 2020<br>6 apr 2020 |
| 12 USA Seatle WA | **SARS-CoV-2/WA-UW306/human/2020/USA, partial genome**<br>**Sequence ID: MT263389.1** | Delete 1-61<br><br>285-305 TTTGGGTGT TTATTACCAC AA/ 21N | 248/269 | 92% | **SIV 29-50 Tanzania**<br><br>**SivagmTan 279 298** | 23 mar 2020<br>6 apr 2020 |
| 13 China Wuhan | **Wuhan seafood market pneumonia virus genome assembly, chromosome: whole_genome**<br>**Sequence ID: LR757997.1** | Full region 1-330 deleted | 1-330 | 0% | ALL eight HIV/SIV « EIE » of region « B » are DELETED | 31 dec 2019<br>6 mar 2020 |

**13 COVID-19 different stains ===> 20 different « IEE » HIV/SIV**

Full details in supplementary materials2 « k »

Some conclusions on the geographical evolution of the genome:
In China, the strains seem to have changed very little in mutations (with the exception of Wuhan seafood market pneumonia virus genome assembly, chromosome: whole_genome Sequence ID: LR757997.1).
In Italy and in France, we find no remarkable mutation vis-à-vis the Chinese reference genome.
It is in Spain and the USA that we detect the most blatant traces of a notorious evolution of the genome:
In Spain, recent sequences (March 2020) demonstrate significant deletions and mutations in regions containing "EIE". According to the first results of analyzes [13], this genome would not have increased its pathogenicity and would seem to use new modes of transmission.
In the USA, the analysis of multiple sequences from the Seatle region (WA) and Minnesota shows a clear growing trees progressiveness in the mutations then successive deletions of the regions "A", "B" and 225 bases, thus :
Table7 (ref 1 to 7, then 11 to 13), we progress from simple mutations to longer mutations on 3 codons, they affect HIV / SIV "EIE".
Table9: also, there are grouped mutations (ref 4, 5) affecting "EIE" areas.
Table 10: here we illustrate at best a sort of "shedding" of "EIE" regions in which these genomes progress: thus, (ref 3 5 6 7), the mutations affect 2 or 3, then 8 consecutive bases.
Then (9 10 11 12), in addition to other new mutations, it is whole pieces, on several tens of bases of the genome which are deleted. The most remarkable point is that in all these cases, it is indeed "EIE" regions which are targeted.
These analysis are summarized by Figure8 in §Conclusion.

On the most recent date of April 23, 2020, we can check how other COVID_19 strains from Seatle WA have new deletions located in regions “A” and "B" of our article. It is deletions that are "shedding" in part of the EIE HIV / SIV located in region “A” and also in region “B”, particularly in the “side by side” EIE (see in Table 1: HIV1 Malawi 2013, HIV1 Russia 2010, SIV Cameroon 2015). There is the case particularly for:

Sequence ID: MT188341.1Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW386/2020, partial genome

Length: 29835 collected 5mar2020, sequenced13mar2020,

Sequence ID: **MT263466.1** Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW386/2020, partial genome

Length: 29634 collected 16mar2020, sequenced 15apr2020

Sequence ID: MT263385.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW302/2020, partial genome

Length: 29610 collected 23mar2020, sequenced 15apr2020

Sequence ID: MT293224.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1608/2020, complete genome

Length: 29847 collected 18mar2020, sequenced 15apr2020

Sequence ID: MT293213.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/WA-UW-1574/2020, complete genome

Length: 29887 collected 19mar2020, sequenced 15apr2020

These analysis are summarized by Figure9 in §Conclusion.

# CONCLUSIONS :

The question of the origin of COVID-19 remains an open question : see particularly [14-20] and [5]. To conclude, we will present first of all below some graphs and Tables of synthesis. This qualitative information clearly demonstrates the existence of tight regions of various “EIE” of HIV SIV in the genome of COVID-19.

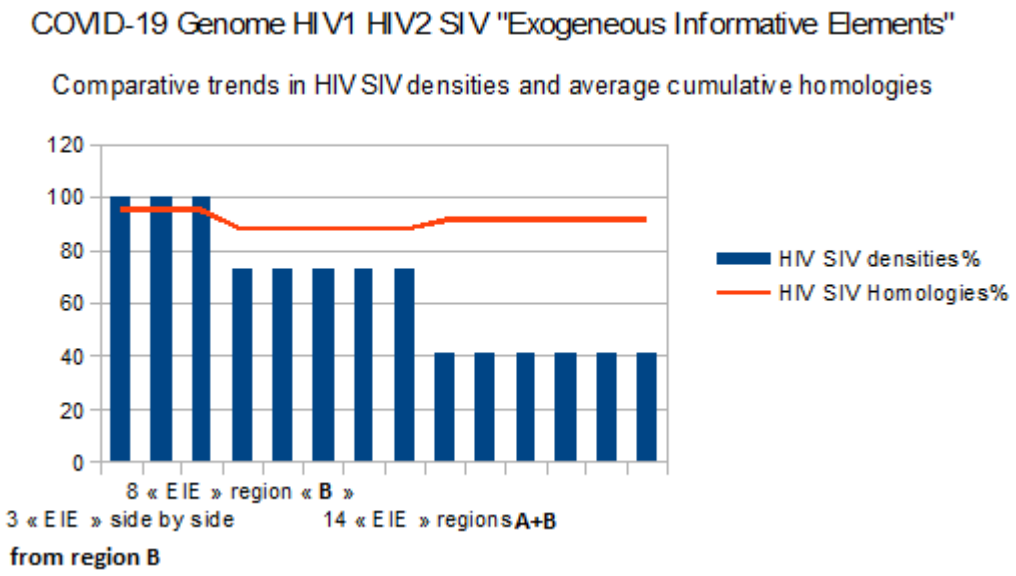



**Figure 4** – Comparative trends in HIV/ SIV « EIE » densities (blue) and average cumulative homologies (red) for 3 clusters : the 3 region B EIE side by side, 8 EIE from region B, and all 14 EIE (A+B cumulated regions).

**Table 11** - The 16 HIV SIV « EIE » according to their homologies with COVID-19 sorted by decreasing %.

| **HIV SIV strain** | **COVID-19 gene** | **Homology** |
|---|---|---|
| HIV2 Env France 2012 (unsignificant) | Orf1ab | **100,00%** |
| HIV1 Sweden 2017 (recombinant form in Sweden) | Orf1ab | **100,00%** |
| HIV2 Env Cap verde 2012 | S spike | **100,00%** |
| HIV2 Pol 2011 Senegal (unsignificant) | S spike | **100,00%** |
| SIV Pol 2015 Germany | Orf1ab | **100,00%** |
| SIV 2016 African Monkey | Orf1ab | **95,00%** |
| HIV2 RT Pol 2014 Cote d'ivoire | S spike | **95,00%** |
| SIV Env 2016 Africa | S spike | **95,00%** |
| HIV2Env 2012 Guinea | Orf1ab | 94,00% |
| HIV1 Integrase 2004 USA | Orf1ab | 93,00% |
| HIV1 Env 2011 USA | Orf1ab | 93,00% |
| HIV1 Env 2016 Netherlands | S spike | 89,00% |
| HIV1 Env 2008 Kenia | Orf1ab and S spike | 88,00% |
| HIV1 Env 2013 Malawi | S spike | 88,00% |
| HIV1 Env 2016 China | Orf1ab | 87,00% |
| HIV2 1993 Cote d'ivoire | S spike | 85,00% |
| SIV Pol 2013 CAmeroon | S spike | 83,00% |
| HIV1 Env 2010 Russia | S spike | 82,00% |
| **Average Homology %** | **9 Orf1ab and 10 S spike** | **92.61%** |

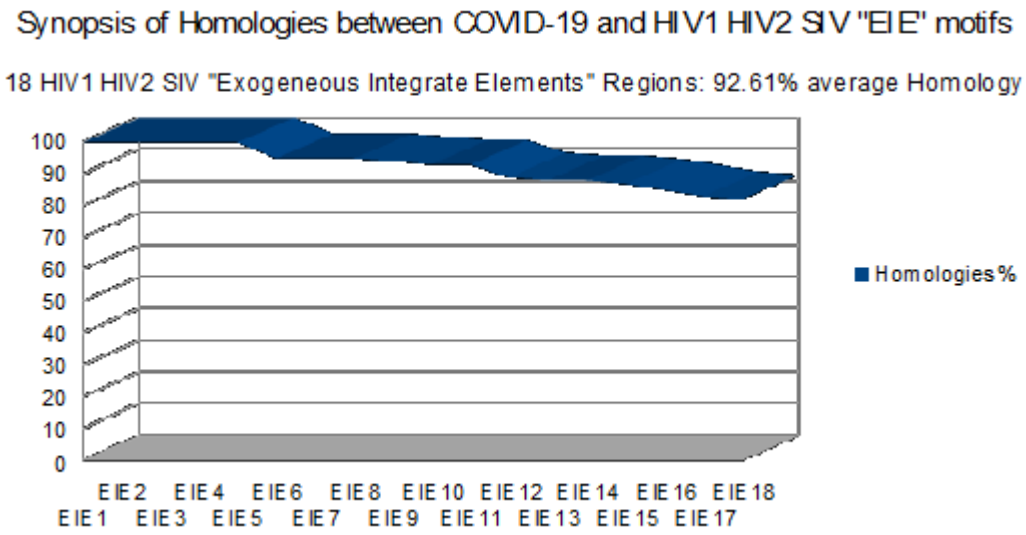



**Figure 5** - The 18 HIV SIV « EIE » motifs according to their homologies with COVID-19 sorted by decreasing %.

## Analysis of local and global cohesions and heterogeneities of the genomes of COVID_19, bat RaTG13 and SARS Urbani.

Now, we demonstrate how and why a new region including 4 HIV/SIV “EIE” radically distinguishes all COVID-19 strains from all SARS and Bat strains.

Then, we will be particularly interested in the Bat RaTG13 strain whose genomic proximity to COVID-19 will be analyzed with the greatest attention and precision.

The theoretical method used here makes it possible to evaluate the overall level of cohesion - then also of heterogeneity - of a sequence of nucleotides, and that whatever the scale due to the fractal nature of this numerical method.

**The "DNA Master Code" of DNA :**

Full details on this numerical method in [6-8], and recall Methods in supplementary Materials2 « m ».

Starting from the atomic masses of the C O N H S P bioatoms constituting RNA, DNA nucleotides and amino acid, a simple law of projection of these atomic masses leads to a UNIFICATION of GENOMICS and PROTEOMICS patterned images that can be calculated for any DNA/RNA codons sequence.

This numerical projection of atomic masses produces a whole numbers numerical code common to the triplets codons DNA, RNA, or amino acids.

A process of DIGITAL INTEGRATION at short, medium and very long distance then allows a globalization of genetic information by a principle which recalls an analogy with the HOLOGRAM. *« Thus, any codon radiates at long distance and vice versa ».*

The Master Code of this sequence then produces two signatures, one GENOMIC, the other for PROTEOMIC, materialized by 2 very strongly correlated curves. that is this level of coupling which will provide precious information on the COHESION or on the HETEROGENEITY of this nucleotide sequence. in particular the extreme regions (mini / maxi) would be associated with biological functions such as active sites, chromosomes breakpoints, etc.

Here we analyze the Master Code of 3 characteristic genomes COVID_19, bat RaTG13 and SARS Urbani.

We will study, for each of these 3 genomes, 5 successive amplitude scales and this according to the 3 reading frames of the codins and on the 2 main and complementary strands:

- wole genomes.

- bases 15,000 to 25,000.

-region including "A", "B", "lyons weiler".

- regions of 425 bases including 100, 225, 100 bases.

- 225 bases area.

Full deoails are available in Supplementary Materials2 « l ».

**Table 12** – Synthetic Genomics/Proteomic global Master Code coupling (%). Nota : we select in each case the best codons reading frame % coupling.

| Genome | **Selective Region 225 bases** | 100 bases , selective 220 bases region, 100 bases | Regions « A » to « Lyons-Weiler » (including « B ») | Region 15000 to 25000 | Full genome |
|---|---|---|---|---|---|
| Wuhan market **ID:** LR757998.1 | **69.47** | 88.30 | 78.37 | 92.31 | 86.23 |
| BatRaTG13 **ID:** MN996532.1 | **92.13** | 83.44 | 59.56 | 94.07 | 86.29 |

| | | | | | |
|---|---|---|---|---|---|
| SARS Urbani **ID:** MK062180.1 | **Absent** | 64.38 | 72.72 | 94.92 | 88.30 |

We must recall here both 225 bases regions within Wuhan market **ID:** LR757998.1 reference and bat RaTG13 genomes :

**Wuhan seafood market pneumonia virus genome assembly, chromosome: whole_genome**
**Sequence ID: LR757998.1 Length: 29866 Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 407 bits(450) | 7e-114 | 225/225(100%) | 0/225(0%) | Plus/Plus |

**Bat coronavirus RaTG13, complete genome**
**Sequence ID: MN996532.1 Length: 29855 Number of Matches: 1**

| Score | Expect | Identities | Gaps | Strand |
|---|---|---|---|---|
| 312 bits(345) | 4e-85 | 204/225(91%) | 0/225(0%) | Plus/Plus |

**The sequence SARS Urbani is totally absent** selecting 1000 SARS like genomes in BLAST..

**Homology of the 225 bases region between Wuhan market ID: LR757998.1 ref. and bat RaTG13 is very important : 204/225 bases (91% homology).**

Analysing the locations of the 4 HIV1 HIV2 « EIE » within the 225 bases region :
Wuhan market **ID: LR757998.1** start adress : 21543. Bat start adress : 21550.

Nucleotides and amino acids within Wuhan market **ID: LR757998.1** 225 bases region :

**HIV1 Kenya 2008**
**471 501 Nucleotides adresses within region « A » 600 bases**
**1 31 Nucleotides adresses within region 225 bases**
**1 10 Amino acids within region 225 bases**

HIV2 Cap verde 2012
512 529 Nucleotides adresses within region « A » 600 bases
42. 59 Nucleotides adresses within region 225 bases
14. 20 Amino acids within region 225 bases

HIV2 Cote d' ivoire 2014
66 85 Nucleotides adresses within region « B » 330 bases
195. 214. Nucleotides adresses within region 225 bases
65. 71 Amino acids within region 225 bases

SIV Africa 2016
76 97 Nucleotides adresses within region « B » 330 bases
205. 226 Nucleotides adresses within region 225 bases
68. 75 Amino acids within region 225 bases

Homologies between BatRaTG13[21549 on 225 bases] =Wuhan market **ID:** LR757998.1 ref [21542 on 225 bases]

| | |
|---|---|
| **1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1** | **Kenya HIV1** |
| **1** 1 1 1 1 1 1 1 1 1 **1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1** 1 | Cap verde HIV2 |
| 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 0 1 0 | |
| 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | |
| 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 | |
| 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 | |
| 0 1 1 0 1 1 1 1 1 1 1 1 1 **1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1** | 2 last HIV2 and SIV have a partial overlap. |
| **1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1** | |

Then, only 20 bases differences on 225 bases.

Nota : The regions in bold correspond to the relative positions of the 4 "EIEs" HIV1 Kenya 2008, HIV2 Cape Verde 2012, HIV2 Cote d (ivoire 2014 and SIV Africa 2016.

Wuhan market **ID:** **LR757998.1** ref region 225 bases
Frame1
TGTTTTTCTTGTTTTATTGCCACTAGTCTC
TAGTCAGTGTGTTAATCTTACAACCAGAAC
TCAATTACCCCCTGCATACACTAATTCTTT
CACACGTGGTGTTTATTACCCTGACAAAGT
TTTCAGATCCTCAGTTTTACATTCAACTCA
GGACTTGTTCTTACCTTTCTTTTCCAATGT
TACTTGGTTCCATGCTATACATGTCTCTGG
GACCAATGGTACTAA

bat RaTG13 region 225 bases
Frame1
TGTTTTTCTTGTTTTATTGCCACTAGTTTC
TAGTCAGTGTGTTAATCTAACAACTAGAAC
TCAGTTACCTCCTGCATACACCAACTCATC
CACCCGTGGTGTCTATTACCCTGACAAAGT
TTTCAGATCTTCAGTTTTACATTTAACTCA
GGATTTGTTTTTACCTTTCTTCTCCAATGT
GACCTGGTTCCATGCTATACATGTTTCAGG
GACCAATGGTATTAA

Wuhan market **ID:** LR757998.1 region 225 bases
FRAME1
=======
**CYS PHE SER CYS PHE ILE ALA THR SER LEU** **Kenya HIV1**
ARR SER VAL **CYS ARR SER TYR ASN GLN ASN** **Cap verde HIV2**
SER ILE THR PRO CYS ILE HIS ARR PHE PHE
HIS THR TRP CYS LEU LEU PRO ARR GLN SER
PHE GLN ILE LEU SER PHE THR PHE ASN SER
GLY LEU VAL LEU THR PHE LEU PHE GLN CYS
TYR LEU VAL PRO **CYS TYR THR CYS LEU TRP** **2 last HIV1 and SIV have a partial overlap**
**ASP GLN TRP TYR ARR**

bat RaTG13 region 225 bases
FRAME1
=======
**CYS PHE SER CYS PHE ILE ALA THR SER PHE** **Kenya HIV1**
ARR SER VAL **CYS ARR SER ASN ASN ARR ASN** **Cap verde HIV2**
SER VAL THR SER CYS ILE HIS GLN LEU ILE
HIS PRO TRP CYS LEU LEU PRO ARR GLN SER
PHE GLN ILE PHE SER PHE THR PHE ASN SER
GLY PHE VAL PHE THR PHE LEU LEU GLN CYS
ASP LEU VAL PRO **CYS TYR THR CYS PHE ARG** **2 last HIV1 and SIV have a partial overlap**
**ASP GLN TRP TYR ARR**

Nota : The best nucleotides and amino acids matchings must be analysed from the 3 codons and directions of codons reading frames.

In other words, in this above Table5 we see that apart from HIV1 KENYA the HIVs of the 225 bases region are more homologous in Wuhan market **ID:** **LR757998.1** than in batRATG13.

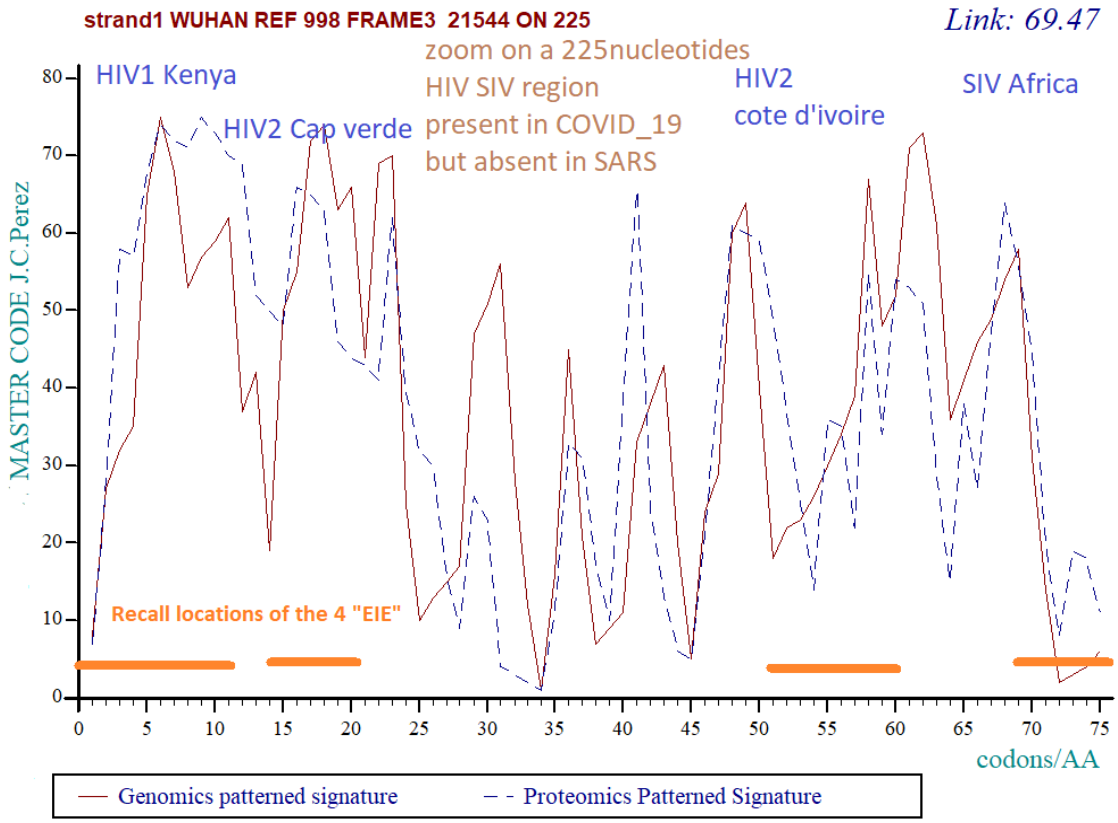



**Figure 6** – High level of HETEROGENEITY within the 225 bases region in Wuhan market **ID:** **LR757998.1** reference genome.

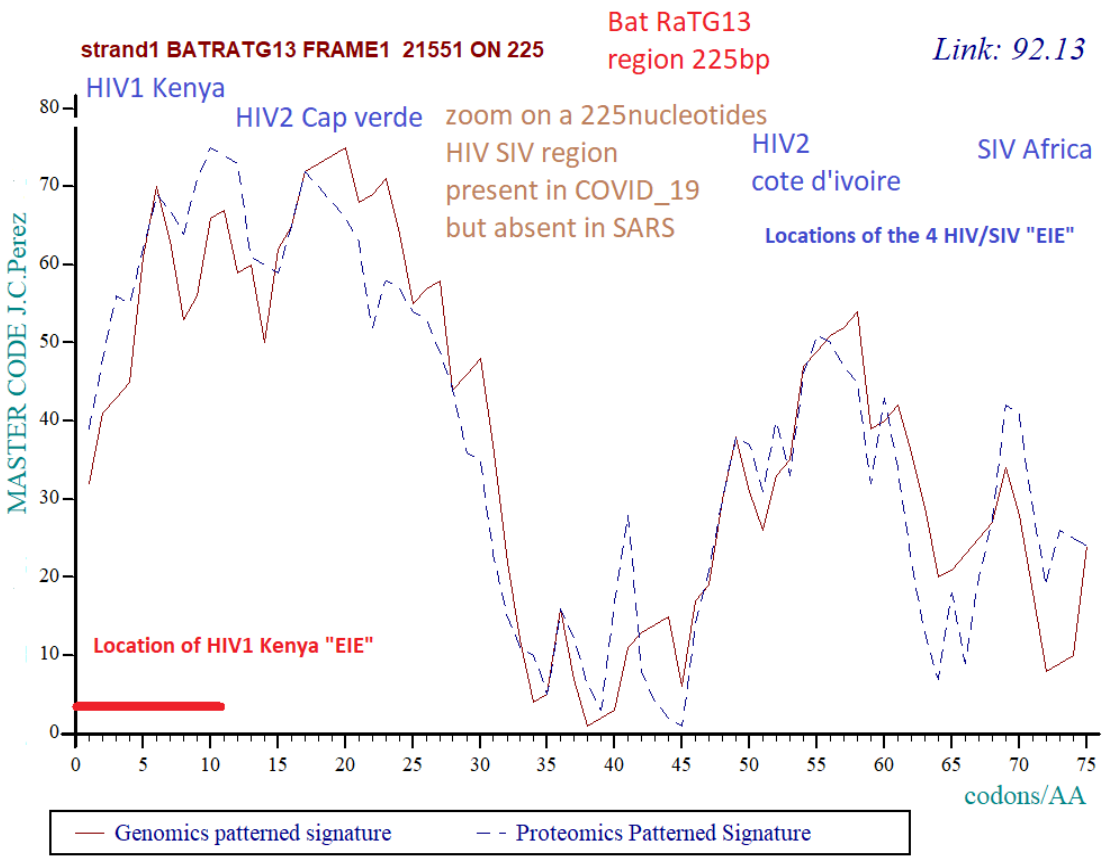



**Figure 7** – High level of COHESION in 225 bases bat RaTG13 region including the fingerprint of **Kenya HIV1** but not the 3 others HIV SIV signatures.

We will draw the reader's attention to the 2 figures 6 and 7 above: The first concerns the 225bp region of COVID-19 (Fig. 6), it appears chaotic and not very organized. On the contrary, the same analysis for the same 225bases region in bat RaTG13 (Fig. 7) shows a more "smoothed" and regular profile. Let us not forget that this sequence, although filed in 2020, was taken in 2013,then 7 years earlier.

**The path of a co-evolution hypothesis:**
HIV, SARS or COVID-19 are not the mitochondrial DNA for which we have been able to trace back to a common ancestor ..."mtDNA Eva" ancestry mother.
*Here SARS appeared to us AFTER HIV.*
Then COVID-19 appears to us AFTER SARS, therefore AFTER HIV.
It is therefore natural for the human logical mind to seek a possible trace of it **CHRONOLOGICALLY**.
This is what has just been done here.

*But what about if COVID-19 had appeared to humans BEFORE HIV?*
We would then "logically" look for traces of COVID-19 in HIV … And we would have found ...
Of course, HIV is a retrovirus while COVID-19 remains, in the current state of knowledge, an RNA virus.
But what about a possible evolution of these 2 viruses so different from an archaic virus which would have served as a **"matrix"** for evolutions towards these 2 viruses? An **"ANCESTOR RNA"**, in a way?
It would then be plausible to discover today some traces of signature which would be common to them ...
But one question will remain unanswered:

**«WHY HIV SIV « EIE » are found here *«side to side»* within a contiguous COVID-19 RNA sequence?»**

**And, more generally, who can prove that COVID_19 comes from Bat RaTG13?**

## Tracks and suggestions on the possible origins of COVID_19:

Gradually, we discovered a small region of the COVID_19 genome, 225 nucleotides long, which distinguishes the latter because it is completely absent in SARS but also in ALL bat genomes, except for one

batRaTG13 , this one even if which is presented in (Zhou P, 22020, Andersen K.G., 2020) as the certain origin of COVID_19.

==> We will now place ourselves in this scenario of a possible emergence of COVID_19 from batRaTG13:

-A first major fact, this region of 225 bases contains the fingerprint of 4 “EIE” of HIV: 2 come from HIV1 and 2 others from HIV2. This region is absent in SARS genomes.

-A second major fact: the first of these 4 “EIE” is EXCLUSIVELY only in batRaTG13 and in COVID_19. It is radically absent from all other SARS and bats strains . Meanwhile, 3 by 4 HIV “EIE” (excluding HIV1 Kenya) are partially presents in multiple bats strains genomes.

-A 3rd fact: indeed, it appears that batRaTG13 would be ANTERIOR to COVID_19: although its genome was only sequenced in 2020, this sample was taken in 2013 and then preserved according to the state of the art of the processes of conservation of biological samples.

-a 4th fact: the analysis of homologies between the 3 last “EIE” by the 4 HIV “EIE” in RaTG13 and COVID_19 clearly demonstrates that these 3 “EIE” have evolved and mutated in RaTG13 in order to adapt to this genome, which is not the case for COVID_19, most recent (2019). Contrarly, the first one (HIV1 Kenya) appears identical then not mutated between batRaTG13 (2013) and COVID_19 (2019). Why?

- a 5th fact: the "master code" method which allows, very precisely to quantify the degree of integration, cohesion, or heterogeneity of a DNA/RNA sequence confirms, by any other means, the 4th fact above (Figure 8).

-a 6th fact: the same analysis as that of 5) above on the same region of 225 bases in COVID_19, leads to a very chaotic and disorderly "master code" image (Figure 7), which would confirm the lack of adaptation of a genome recent.

-a 7th fact: the first of the 4 “EIE” (Kenya HIV1) appears almost identical between the 2 strains batRaTG13 and COVID_19, as if he had not needed to mutate in RaTG13.

==> Let us now put ourselves in the opposite scenario, that according to which COVID_19 does not come from batRaTG13: What would be the arguments or facts which would go in this direction?

-on the one hand, it remains surprising that the HIV1 Kenya fingerprint has similar homologies between 2 genomes so different over time: 2013 (but sequenced in 2020) for batRaTG13, and 2020 for COVID_19.

-on the oher hand, why does this region of 220 bases specific to COVID_19 and batRaTG13 appear to us so heterogeneous, even chaotic in COVID_19 unlike batRaTG13?

-thirtly, how to explain that we find multiple traces of the 3 HIV others than the fingerprint HIV1 Kenya in very many bat strains (for exemple, HIV2 Cap verde), for the HIV1 Kenya strain, only bat RaTG13 presents HIV homology to the exclusion of ALL other bat strains?

-Fourth, besides the progressive adaptation of the 4 HIV “EIE” observed in RaTG13, the best HIV homologies observed for COVID_18, could be explained by "humanization" common to both the HIV genomes and the human Coronaviruses. But who could explain how and why this so recent genome (2019-2020) was able to evolve so quickly to "humanize"?

-Finally, we can check in what context researchers have come to be interested in this strain HIV1 Kenya: that of the attempt to create a vaccine against HIV, based on exceptional properties of this HIV virus ( Land A.M. Et al ).

-To conclude, cumulating datas from Tables 8, 9 and 10, the following Figure 8 demonstrates how in different regions of the world genomes have evolved by significant mutations and large deletions in regions "A", "B", and 225 bases affecting or suppressing several "Exogeneous Informative Elements ".

We note in particular that the major part of these significant mutations § deletions affect the HIV / SIV regions "Exogenous Informative Elements" (red and green curves). What matters in the evolution of this virus, it is not the places but the duration: COVID_19 went around the world in both directions between December 31, 2019 and April 62020. The fact that after Wuhan downtown and region, the deletions appeared in Seattle can be explained either by the very high number of replicative cycles of the virus in the USA, or by a direct importation of the virus from China by the Chinese community present on the West Coast from the USA. Finally, In any case, it is the time of evolution of the genome that counts in the first place. Particularly, in [22], Hangping Yao et al. reports that, analysing different strains to outbreaks, the first referenced COVID_19 mutations around the world shows that New-york state and Europe strains are more pathogen killers than the USA first mutated genomes like principally Seattle area strains (WA Washington State) like cases reported here in Figures 8 and 9.

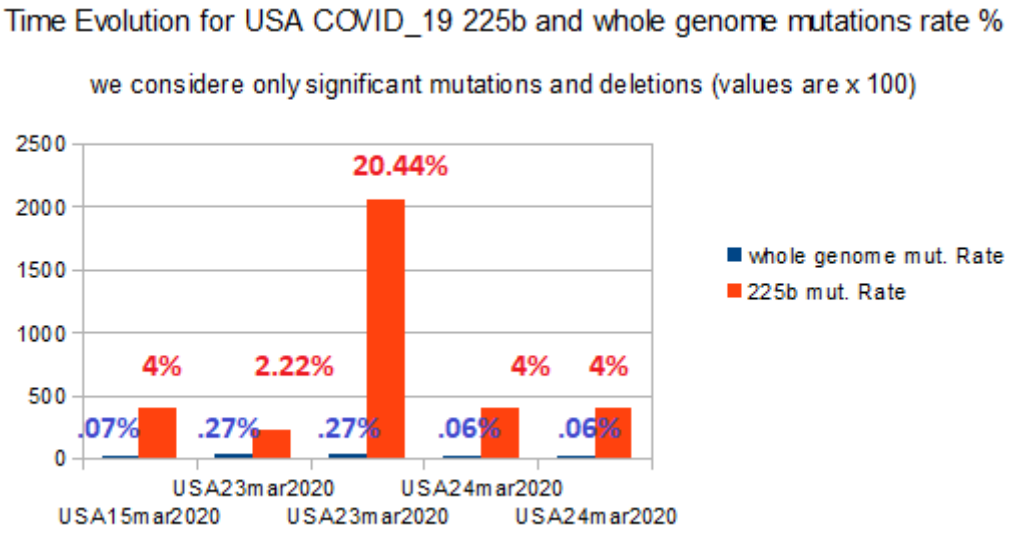



**Figure 8** – From Table 8, this chart show comparative time evolution between Seattle (WA) first mutations/deletions rates % at whole genome and 225b levels.

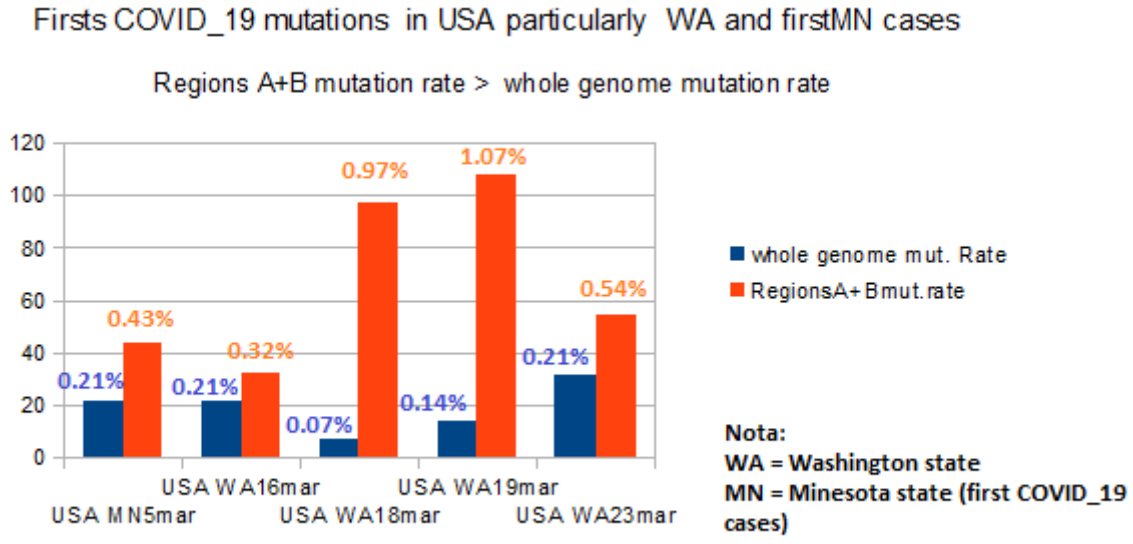


**Figure 9** – This other chart show comparative time evolution between (WA and Minesota strains first mutations) and mutations/deletions rates % at whole genome and in the case of region 930 bases = region « A » (600bases) + region « B » (330 bases).

## Final conclusions:

Despite the great length of these RNA virus genomes and the mutation-limiting mechanisms that characterize them, we predict that this 225 bases region will only mutate in order to increase its level of global integration vis-à-vis its genome.

The hyper-selective and discriminating region at the start of the 225 bases region corresponding to the Kenya HIV1 fingerprint could and should constitute a privileged target for possible tests or even therapies or vaccines.

But, be that as it may, if we were to retain only one thing from this study, it will be advice to the global community of those who are fighting this pandemic: ***"We cannot apply the current and proven tools and methods of fighting against viruses, hitherto NATURAL, because we do not know today how a new virus whose part of the genome is SYNTHETIC in nature will evolve and evolve. This was already true for SARS, and even more so for COVID_19."***

To summarize, the five main conclusions of this article are:

1 / Many Exogeneous Informative Elements "EIE" HIV / SIV (about 20 EIE) are present in the genome of COVID_19 but also partly in SARS.

2 / Some regions of COVID_19 contain a concentrated density of EIE. In particular, two regions contain, one 2 EIE "side by side", the other 3 EIE "side by side". Notably, the retroviral part of the densest region, which consists of 8 motifs from various strains HIV1, HIV2 and SIV, covers a length of 275 contiguous bases of COVID-19. The cumulative length of these 8 HIV SIV motifs represents 200 bases. Consequently, the HIV SIV EIE density rate of this region of COVID-19 is 200/275 = 72.73%, which is considerable.

3 / A small region of 225 bases distinguishes radically COVID_19 from all SARS genomes, it contains in particular among its 4 EIEs, a fragment coming from a nonfunctional 2008 KENYA HIV1 genome which we know was used in a vaccine strategy HIV.

4 / We report also the major differences between the genomes of COVID_19 and bat RaTG13 which one asserts the natural origin of covid_19.

5 / Finally, the first major mutations in the COVID_19 genome taken in March and deposited in April 2020 mainly concern patients from the state of Washinton (Seattle WA) in the USA. We then observe that these mutations tend to modify and delete EIE fragments located, very precisely in the dense EIE regions that we have discovered. In addition, these deletions would be associated with a decrease in the pathogenicity of the virus.

# REFERENCES :


1.WHO-SARS, https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.who.int/ith/diseases/sars/en/&ved=2ahUKEwiYufHk5tDoAhXU3oUKHSTwBuYQFjAWegQIBRAB&usg=AOvVaw0bFoEUPELafXU98baC4o2k

2.WHO-MERS, https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.who.int/emergencies/mers-cov/en/&ved=2ahUKEwjigPe059DoAhXEx4UKHU5xDDYQFjAMegQIBBAC&usg=AOvVaw1kaYVgLwAr9c7EyL7kGXQn

3.Perez, J.C, 2020/02/13, Wuhan nCoV-2019 SARS Coronaviruses Genomics Fractal Metastructures Evolution and Origins, DO -DOI: 10.20944/preprints202002.0025.v2, Researchgate : https://www.researchgate.net/publication/339331507_Wuhan_nCoV-2019_SARS_Coronaviruses_Genomics_Fractal_Metastructures_Evolution_and_Origins

4.Lyons Weiler J., 2020, 1-30-2020, On the origins of the 2019 ncov virus wuhan china, https://jameslyonsweiler.com/2020/01/30/on-the-origins-of-the-2019-ncov-virus-wuhan-china/

5.Perez J.C, (2020). "WUHAN COVID-19 SYNTHETIC ORIGINS AND EVOLUTION." International Journal of Research - Granthaalayah, 8(2), 285-324. https://doi.org/10.5281/zenodo.3724003.

6.Perez J.C, **Codex biogenesis** - Les 13 codes de l'ADN (French Edition) [Jean-Claude ... 2009); Language: French; **ISBN**-10: 2874340448; **ISBN**-13: 978-2874340444 ..

7.Perez J.C, Deciphering Hidden DNA Meta-Codes -The Great Unification & Master Code of Biology. J Glycomics

Lipidomics 5:131, 2015, doi: 10.4172/2153- 0637.1000131 https://www.omicsonline.org/openaccess/deciphering-hidden-dna-metacodesthe-great-unification--mastercode-ofbiology-2153-0637- 1000131.php?aid=55261
8.Perez, J.C. Six Fractal Codes of Biological Life:perspectives in Exobiology, Cancers Basic Research and Artificial Intelligence Biomimetism Decisions Making. *Preprints* **2018**, 2018090139 (doi: 10.20944/preprints201809.0139.v1). https://www.preprints.org/manuscript/201809.0139/v1
9.Land A.M. Et al, Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya., J Virol. 2008 Aug;82(16):8172-82. doi: 10.1128/JVI.01115-08. Epub 2008 Jun 11., DOI: 10.1128/JVI.01115-08 https://www.ncbi.nlm.nih.gov/pubmed/18550667
10.Venkatesan P, Franck Alla Plummer, The Lancet Infectious diseases, April 2020, DOI: https://doi.org/10.1016/S1473-3099(20)30188-2 , https://www.thelancet.com/pdfs/journals/laninf/PIIS1473-3099(20)30188-2.pdf
11.*Perez, J. Epigenetics Theoretical Limits of Synthetic Genomes: The Cases of Artificials Caulobacter* (*C. eth-2.0), Mycoplasma Mycoides (JCVI-Syn 1.0, JCVI-Syn 3.0 and JCVI_3A), E-coli and YEAST chr XII*. *Preprints* **2019**, 2019070120 (doi:10.20944/preprints201907.0120.v1).https://www.preprints.org/manuscript/201907.0120/v1
12.Zhou, P et al, 2020, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (7798), 270-273 (2020), DOI: 10.1038/s41586-020-2012-7
13.FISABIO, 2020, http://fisabio.san.gva.es/web/fisabio/noticia/-/asset_publisher/1vZL/content/secuenciacion-coronavirus.
14.Andersen, K.G., Rambaut, A., Lipkin, W.I. et al. The proximal origin of SARS-CoV-2. Nat Med (2020). https://doi.org/10.1038/s41591-020-0820-9
15.Prashant Pradhan et al, Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag,https://www.biorxiv.org/content/10.1101/2020.01.30.927871v1
16.Yuanchen Ma et al., 2020-2-27, ACE2 shedding and furin abundance in target organs may influence the efficiency of SARS-CoV-2 , http://www.chinaxiv.org/abs/202002.00082
17.Xiaolu Tang, Changcheng Wu, Xiang Li, Yuhe Song, Xinmin Yao, Xinkai Wu, Yuange Duan, Hong Zhang, Yirong Wang, Zhaohui Qian, Jie Cui, Jian Lu, On the origin and continuing evolution of SARS-CoV-2, *National Science Review*, , nwaa036, https://doi.org/10.1093/nsr/nwaa036
18.Lu, R et al., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2820%2930251-8/fulltext
19.Wei Ji, et al, Homologous recombination within the spike glycoprotein of the newly identified coronavirus 2019-nCoV may boost cross-species transmission from snake to human, https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/jmv.2568220.
20.Peng Zhou et al, Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin, BioRxiv, January 2020, https://doi.org/10.1101/2020.01.22.914952
21.Leoz M, Feyertag F, Kfutwah A, Mauclère P, Lachenal G, et al. (2015) The Two-Phase Emergence of Non Pandemic HIV-1 Group O in Cameroon. PLOS Pathogens 11(8): e1005029. https://doi.org/10.1371/journal.ppat.1005029
22.Hangping Yao, et al., Patient-derived mutations impact pathogenicity of SARS-CoV-2 medRxiv 2020.04.14.20060160; doi: https://doi.org/10.1101/2020.04.14.20060160