May 2022 : Peculiar Evolution of the Monkeypox Virus Genomes

Article · May 2022

CITATION
1

READS
6,611

2 authors:

jean-claude Perez
116 PUBLICATIONS **524** CITATIONS
SEE PROFILE

Valère Lounnas
European Molecular Biology Laboratory (Heidelberg)
13 PUBLICATIONS **16** CITATIONS
SEE PROFILE

May 2022 :
Peculiar Evolution of the Monkeypox Virus Genomes

Jean Claude Perez, PhD Maths§Computer Science Bordeaux University; Retired (IBM European Research center on Artificial Intelligence Montpellier France);
Bordeaux metropole France; https://orcid.org/0000-0001-6446-2042 France
and Luc Montagnier Foundation Scientific Council, Quai Gustave-Ador 62 1207 Genève, Switzerland, jeanclaudeperez2@gmail.com

Valère Lounnas, PhD, EMBL Heidelberg alumni, Meyerhofstraße 1, 69117 Heidelberg, Germany

KEYWORDS

Monkeypox virus, Biomathematics, Master code, Evolution, Genomics, Proteomics.

ABSTRACT

We compare the evolution of 14 monkeypox virus genomes til that of May 2022 that is currently spreading across humans in numerous countries outside Africa. Our aim was to discover mutations or other viral evolutions (recombination) that may explain the sudden impact of this very low-level circulating epidemic or alert on a potential peculiar pathogenic character.

We have evidenced the presence of a large number of T bases in succession, at the level of the polymerase, between the DNA-dependent RNA polymerase subunit rpo132 and the cowpox A-type inclusion protein, progressively rising from the absence of a characteristically long pattern of T-bases in succession (≤ 10) in the early genomes of 1971, up to 19 T-bases in the Israel 2018 strain of reference, and 30 T bases thereafter in the 2022 strains. We find a complementary match for this long T bases sequence only in the simian hemorrhagic encephalitis virus, at the very 3' end of the genome after the stop codon, with a long succession of 28 A bases. More strikingly, we find that the corresponding 10 phenyl-alanine aa chain is reported as matching uniquely *(E≤0.001)* a hypothetical protein element in *Plasmodium falciparum,* Y*ersinia pestis, Escherichia coli and Penicillium nordicum.* We wonder about the possibility that this region of the monkeypox genome may potentially code for a not yet identified polypeptides with a functional role situated right upstream this long T-repeat.

INTRODUCTION

Monkeypox is a zoonotic disease caused by the monkeypox virus, an orthopoxvirus closely related to variola virus, the causative agent of smallpox. Monkeypox was first discovered in 1958 in monkeys, although they are not the source of the virus. Human cases were first described in 1970. There are 2 strains of monkeypox: the West African and Central African strains.

Several cases of monkeypox have been identified in various geographically countries. In May 2022 cases were reported in Australia, Austria, Belgium, Canada, Denmark, France, Germany, Greece, Israel, Italy, the Netherlands, Portugal, Spain, Sweden, Switzerland and the U.K (NCBI, 2022), (Antwerpen M, et al, 2022), (Isidro et al, 2022).

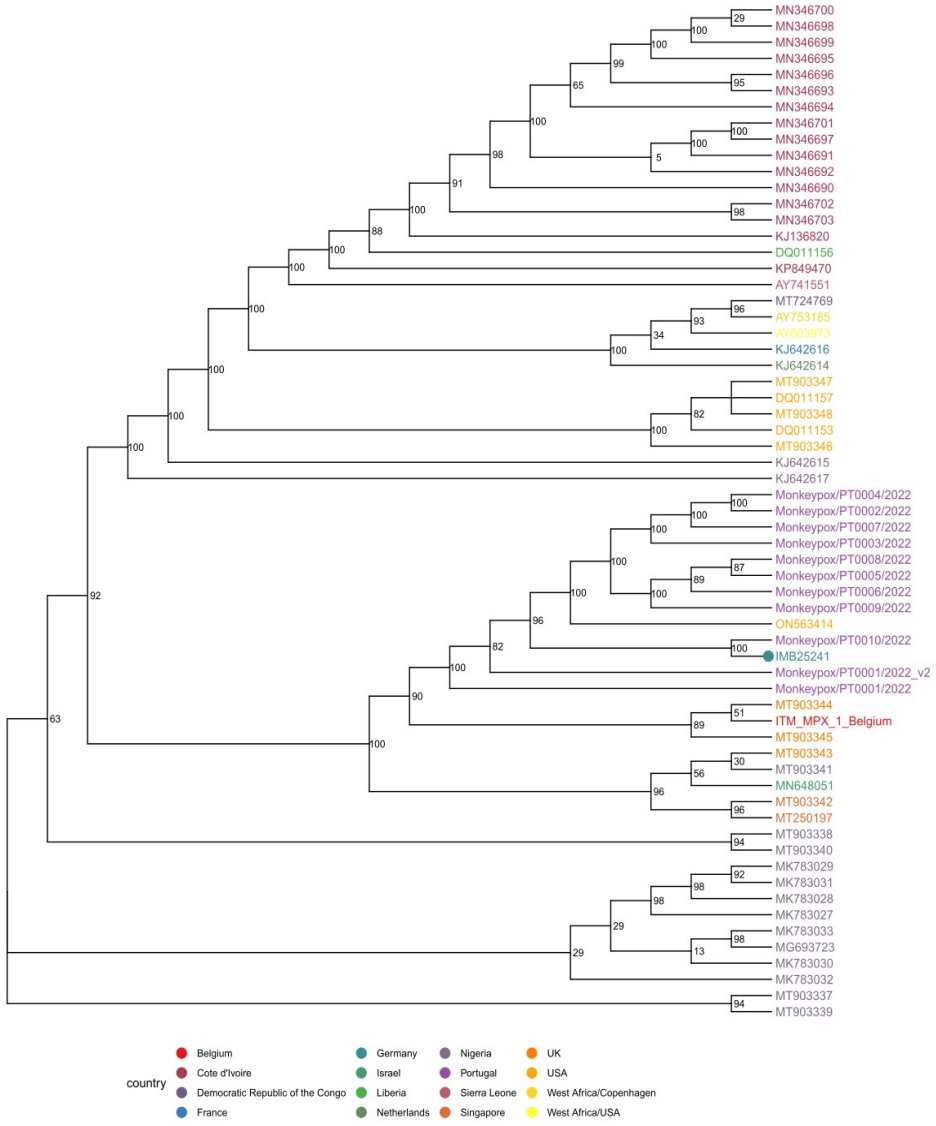

Figure1 – Monkeypox tree (from https://virological.org/t/first-german-genome-sequence-of-monkeypox-virus-associated-to-multi-country-outbreak-in-may-2022/812)
Nextrain reference tree https://nextstrain.org/monkeypox?s=03

Monkeypox is classified as a zoonotic disease where transmission of the virus is usually due to animal-human contact. Genetically, monkeypox viruses cluster into two groups: the Congo basin and the west African clade.

Monkeypox virus
Monkeypox virus Zaire-96-I-16

This particular outbreak has been identified as due to a virus from the west African clade which is often associated with milder disease and, in this case, human-to-human spread is suspected. The first referenced human to human strain was located in Israel in 2018: a case of monkeypox in a man who returned from Nigeria to Israel in 2018 (Erez et al, 2018).

MATERIALS and METHODS

Monkeypox strains analyzed :

We analyzed 14 monkeypox whole genomes:

Gabon 1988 alias 2015 KJ642619.1
https://www.ncbi.nlm.nih.gov/nuccore/KJ642619.1

Cameroun 1990 alias 2015 KJ642618.1
https://www.ncbi.nlm.nih.gov/nuccore/KJ642618.1

Liberia 1970 DQ011156.1
https://www.ncbi.nlm.nih.gov/nuccore/DQ011156.1

Nigeria 1971) alias 2015 KJ642617.1
https://www.ncbi.nlm.nih.gov/nuccore/KJ642617.1

2018 Israel MN648051.1
https://www.ncbi.nlm.nih.gov/nuccore/MN648051.1

Zaire 2009 alias 2020 NC_003310.1
https://www.ncbi.nlm.nih.gov/nuccore/NC_003310.1

Rivers state 2020 MT903340.1
https://www.ncbi.nlm.nih.gov/nuccore/MT903340.1

UK 2020 MT903344.1
https://www.ncbi.nlm.nih.gov/nuccore/MT903344

USA 2022 ON563414.1
https://www.ncbi.nlm.nih.gov/nuccore/ON563414.1?report=GenBank&s=03

German 2022 ON568298.1
https://www.ncbi.nlm.nih.gov/nuccore/ON568298

Singaore 2020 MT903342.1
https://www.ncbi.nlm.nih.gov/nuccore/MT903342.1?report=genbank

Nigeria 2018 MG693723.1
https://www.ncbi.nlm.nih.gov/nucleotide/MG693723.1?report=genbank&log$=nuclalign&blast_rank=1&RID=98T6WWFV016

UK 2020 MT903345.1
https://www.ncbi.nlm.nih.gov/nucleotide/MT903345.1?report=genbank&log$=nuclalign&blast_rank=1&RID=98TT3F4E013

France 2022 ON602722.1
https://www.ncbi.nlm.nih.gov/nuccore/ON602722.1?report=genbank

Biomathematics methods, The Master Code analysis :

The "Master Code" method (Perez, 2009), (Perez, 2015) and (Perez§Montagnier, 2021) allows, from the atomic masses common only to DNA, RNA and amino acids numerical values, to highlight a META-CODE which would unify the 3 codes of DNA, RNA and amino acid sequences.

Particularly, the Master code coupling curves measures the level of correlation unifying the expression of 2 Genomics (DNA) and Proteomics (amino acids) for any sequence, coding for a protein, or not.

In (Perez, 2017a) we analyzed all types of Prions in the early 2000s mad cow disease (plants, yeast, humans, cows, sheep, etc.). We had then highlighted a "signature" or sort of invariant which would be common to all Prions: a typical signature of the Master code taking the characteristic form of a "W" (or even of an “M” symmetrically). We had extended this type of analysis to amyloids implicated in Alzheimer's disease (Perez, 2017b).

RESULTS

Table 1 – Evolution of the « T » bases contiguous region for the 14 analysed genomes.

Name	Genbank ID	Start T location	Number of T
Gabon1988 (2015)	KJ642619.1		0
Cameroun1990 (2015)	KJ642618.1		0
Liberia1970	DQ011156.1		0
ZAire2009	NC_003310.1		0
Nigeria1971 (2015)	KJ642617.1	133245	27
Israel2018	MN648051.1	133298	19
Rivers state 2020	MT903340.1	133081	25
UK2020A	MT903344.1	133081	27
Singapore2020	MT903342.1	133093	28
Nigeria2018	MG693723.1	126745	29
UK2020B	MT903345.1	133100	28
France2022	**ON602722.1**	**132972**	**19**
USA2022	**ON563414.1**	**133094**	**30**
Germany2022	**ON568298.1**	**133201**	**30**

The last 3 cases analyzed date from May 2022. It is of note that the 2022 French genome is limited to a succession of 19 T. But in fact this sequence may also accept C bases substituted for T as both ttt and ttc codons are translated in phenyl-alanine residue. In that respect the length of the French sequence is actually equivalent to 21T. Sequencing errors are possible but not to that extent over 8 nucleotides. So the difference of the French sequence raises some question as it is obviously not the same as the other strains in that respect. It is also the case for the Italian sequence (ON622721 from https://www.ncbi.nlm.nih.gov/nuccore/ON622721.1/).

DISCUSSION

This is by chance that we have discovered the presence of a 30-T long sequence in the middle of the USA2022 monkeypox genome, between the DNA-dependent RNA polymerase subunit rpo132 and the cowpox A-type inclusion protein, before a gene complement region that may become coding under circumstances that need to be specified by expert in the field.

For instance, if we look at the monkeypox strain Gabon-1988 we can identify in this region a sequence of nucleotide coding straightforwardly for a 42-aa long polypeptide that may constitute a small protein.

Number of codons : 42
MGYLRSFYKRFHVPDHVQPSYVSPSLYRVYQSSLSEGDRTP.

Monkeypox virus strain Gabon-1988, complete genome

GenBank: KJ642619.1

gene 131162..134656 /product="DNA-dependent RNA polymerase subunit rpo132"

```
                          tatt gtttagtaga tactcatcaa gattatcaag ataagctaat
134701 tcactaaaca tattatcgga ttcggtattg ttactcgaga atagagttcg ttatgctcct
134761 gatattcgga aatctgtgga gtttcaggtt ttggtggaag tgtaactgct acttggtggg
134821 atactgaagg atatttcaga gagttgtgga tgttcgggtt cgacatccac cgatggtgtc
134881 acgccactaa tcggttcggt aacgtctgtg gatggaggtg ctacttctac agaacctgta
134941 gcctcagttg tcaacggaga tacatcttca atgcgcggaa atgtataatt tggtaatggt
135001 ttctcatgtg gatcttaaga agaagaggta agatatctac gaaagatacc gatcacgttc
135061 tagttctctt ttgtagaact ttaacttttt ctttctcagc atctagttga tattccgacc
135121 tcttcacgtt tcacatgggt tacctccgca gtttttacaa gcgatttcac gttccagatc
135181 acgttcagcc ttcatacgtc tctccctctc tctatcgagt ttatcagagc agtctttctg
135241 aaggcgatcg aactccataa atttctccaa cgctttgatt gtttccatag atttccgaag
135301 tttagcttct aggacggcga ttctttttttc tttcgaattc acggggtaca accgtttcca
135361 ttaccaccat ctctacgttt cttttctaga tcggcaatct ttctcaacat ttcatcccca
135421 tgccttttca ttcctcgagt ctatcgtcgt cgaaatatcg ttccagctcc ttttcgacct
135481 caataacttt agcacgttgt ctcatcaagc tctctcttgt agtactatca tttttatctg
135541 attccctggc acgtttaaga tcttcatgta attgagtcag ctcttgacac aatctcttaa
135601 ctaacttcct ctcttgcttc ttcgtcatag tacttacaat cactatggga tccattgtta
135661 ccacgtctgt actcggcgag ctcacgttta agagattcaa tttccagttt gtacattgat
135721 ttcattatta cgtccgcagt cgttcaactg tatttcaaga tctgagattc tagattgtaa
135781 tctctgtagc atttccacgg cattcactca gttgtctttc aagatctgag attctagatt
135841 ggagtctgct aatctctgta agatttcctc ctccgctctc gatgcagtcg gtcaacttat
135901 tctctagttc tctaatacgt gaacgcagtg catcaacttc ttgtgtgtct tcttgattgc
135961 gtgtgcattc atcgagtcta gattcgagat ctctaacgtg tcgtcgttct tcctcaagtt
```

gene complement(135770..137860) /product="cowpox A-type inclusion protein"

Figure 2a – Genome sequence extract of monkeypox strain Gabon-1988 potentially coding for a small protein after the DNA-dependent RNA polymerase subunit rpo132 and before the gene complement.

Number of codons : 42
MGYLRSFYKRFHVPDHVQPSYVSPSLYRVYQSSLSEGDRTP.

Monkeypox virus isolate MPXV_USA_2022_MA001, complete genome

GenBank: ON563414.3

Gene 128941..132435 /note="A25R RNA polymerase subunit (RPO132) (Cop-A24R) RNA polymerase, 132 kDa subunit similar to Vaccinia virus strain Copenhagen A24R"

gene complement(133217..133444) /note="A-type inclusion protein (Cop-A25L); A26L" "MDPIVIVSTMTKKQERKLVKRLRQELTQLHEDLKRVRESDKNDSTTRESLMKQRAKVIEVEKELERYFDDNRLEE"

```
                                attg tttagtagat actcatcaag ataagctaat tcactaaaca
   132481 tattatcgga ttcggtattg ttactcgaga atagagttcg ttatgctcct gatattcgga
   132541 aatctgtgga gttttcaggtt ttggtggaag tgtaactgct acttggtggg atactgaagg
   132601 atatttcaga gagttgtgga tgttcgggtt cgacatccac cgatggtgtc acgccactaa
   132661 tcggttcggt aacgtctgtg gatggaggtg ctacttctac agaacctgta gcctcagttg
   132721 tcaacggaga tacatattca atgcgcggaa atgtataatt tggtaatggt ttctcatgtg
   132781 gatcttaaga agaagaggta agatatctac gaaagatacc gatcacgttt ctagttctct
   132841 tttgtagaac tttaactttt tctttctcag catctagttg atattccgac ctcttcacgt
   132901 ttcgcatggg ttacctccgc agtttttaca agcgatttca cgttccagat cacgttcagc
   132961 cttcatacgt ctctccctct ctctatcgag tttatcagag cagtctttct gaaggcgatc
   133021 gaactccata aatttctcca acgctttgat tgtttccata gatttccgaa gtttagcttc
   133081 taggacggcg attctttttt ttttttttt ttttttttt ttcgaattca cggggtacaa
   133141 ccgtttccat taccaccatc tctatgtttc ttttctagat cggcaatctt tctcaacatt
   133201 tcatccccat acctttttca
                           agt

gene complement →          agtt tcctcgagtc tattgtcgtc gaaatatcgt tccagctcct
   133261 tttcgacctc aataacttta gcacgttgtt tcatcaagct ctctcttgta gtactatcat
   133321 ttttatctga ttccctgaca cgtttaagat cttcatgtaa ttgagtcagc tcttgacgca
   133381 atctcttaac taacttcctc tcttgcttct tcgtcatagt acttacaatc actatgggat
   133441 ccat
```

Figure 2b – Genome sequence extract of monkeypox strain USA2022 potentially coding for a small protein after the DNA-dependent RNA polymerase subunit rpo132 and before the gene complement.

Number of codons : 42
MGYLRSFYKRFHVPDHVQPSYVSPSLYRVYQSSLSEGDRTP.

This growing pattern of T-bases in succession follows a conserved nucleotide sequence that is conserved and may code for a small protein. The functional role of this pattern at the viral genome level is unknown to us.

While it long repeat are common finding at the terminaison of a genome, as for instance at the end of the monkey encephlitis virus, it is almost never encountered fully inside a sequence.

Simian hemorrhagic encephalitis virus isolate Sukhumi, complete genome
Sequence ID: NC_038293.1Length: 15370Number of Matches: 1

- See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 15336 to 15370GenBankGraphicsNext MatchPrevious Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
55.4 bits(60)	1e-04	33/35(94%)	0/35(0%)	Plus/Minus

```
Query  133098  ttttttttttttttttttttttttttCGAATTCAC  133132
              ||||||||||||||||||||||||||  |||||||
Sbjct  15370   TTTTTTTTTTTTTTTTTTTTTTTTTTTAATTCAC  15336
```

Why it is located in this region ?

Its presence at the end of what seems to be a potential protein may indicate a possible genome regulation role.

May it have another functional role ?

Also remarkable, although there is no evidence this nucleotide sequence is in a genome section that may be translated in aa, we find that a sequence of 30 T-bases codes for a polypeptide chain of 10 phenyl-alanine residues in succession, and that a Blast search for this unorthodox protein sequence surprisingly retrieves a signal with an expectation value significantly beyond randomness *(E≤0.001)* for a match with an identical polypeptide reported as a hypothetical protein in *Plasmodium falciparum, Yersinia pestis, Escherichia coli and Penicillium nordicum !*

However, the question of the functional role remains open as we note (Figure3) this T-base long repeat is located at a peculiar position of the genome predicted to have a marked functional role according to the Master code (44000 aa/ 132000 nt).

An analysis zooming on the small section of 100 bases both sides of the 30-T sequence shows its new functionality (Figure 3) or for the 19-T one in Figure 4.

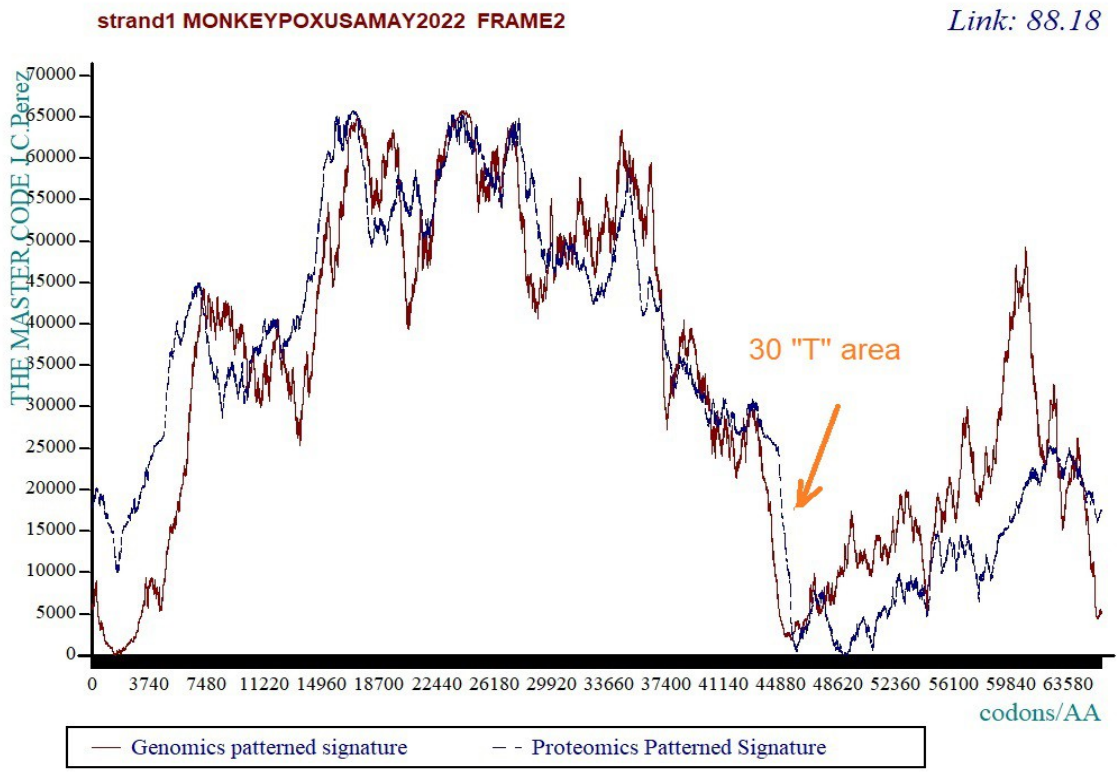

Figure3a – Master code analysis of the whole USA2022 Monkeypox genome. The region 44000 amino acids where there is the 30 T bases insert.appears to be highly functional.

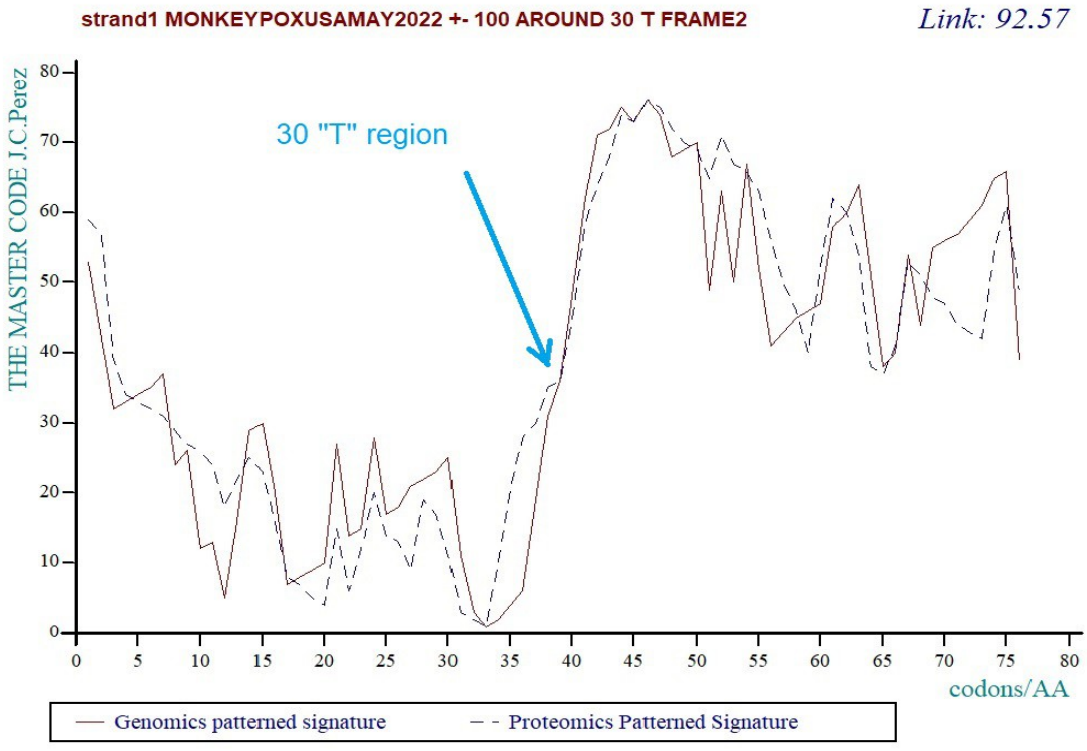

Figure3b – 100 bases upload and download the 30 T bases region in USA2022.

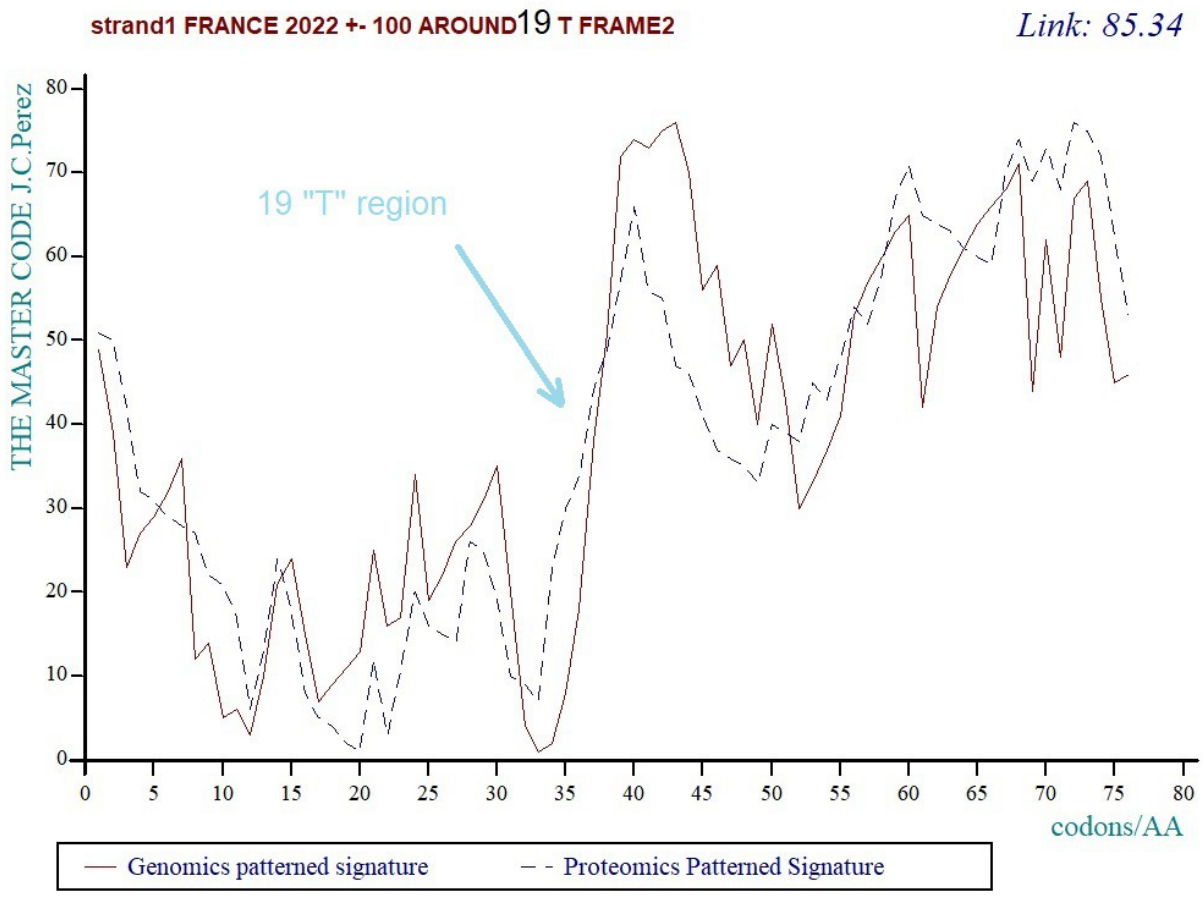

Figure4 - 100 bases upload and download the 19 T bases region in FRANCE2022.

CONCLUSIONS

The objective was here to present a genome characteristic that may partly explain the sudden propagation of the monkeypox virus in the form we observe in May 2022 in quite a number of countries.

The role of the peculiar 30-T base long sequence right in the middle of the virus genome is still to be determined.

REFERENCES

(Antwerpen M, et al, 2022), Markus H. Antwerpen, Daniel Lang, Sabine Zange, Mathias C. Walter* and Roman Wölfel

Bundeswehr Institute of Microbiology, Munich, Germany, First German genome sequence of Monkeypox virus associated to multi-country outbreak in May 2022,,

https://virological.org/t/first-german-genome-sequence-of-monkeypox-virus-associated-to-multi-country-outbreak-in-may-2022/812

(Erez et al, 2018) Erez, Noam et al. “Diagnosis of Imported Monkeypox, Israel, 2018.” Emerging infectious diseases vol. 25,5 (2019): 980-983. doi:10.3201/eid2505.190076

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6478227/

(NCBI, 2022), NCBI Insights , https://ncbiinsights.ncbi.nlm.nih.gov/2022/05/26/monkeypox-virus-genome/amp/

(Isidro J et al, 2022), Joana Isidro1, Vítor Borges1, Miguel Pinto1, Rita Ferreira1, Daniel Sobral1, Alexandra Nunes1, João Dourado Santos1, Maria José Borrego3, Sofia Núncio2, Ana Pelerito2, Rita Cordeiro2, João Paulo Gomes, First draft genome sequence of Monkeypox virus associated with the suspected multi-country outbreak, May 2022 (confirmed case in Portugal), https://virological.org/t/first-draft-genome-sequence-of-monkeypox-virus-associated-with-the-suspected-multi-country-outbreak-may-2022-confirmed-case-in-portugal/799

(Perez, 2009), 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 https://www.amazon.fr/Codex-Biogenesis-13-codes-lADN/dp/2874340448

(Perez, 2015), Deciphering Hidden DNA Meta-Codes -The Great Unification & Master Code of Biology, journal of Glycémies abd Lipidomics,

https://www.longdom.org/abstract/deciphering-hidden-dna-metacodes-the-great-unification-amp-master-code-of-biology-11590.html , ISSN: 2153-0637, DOI: 10.4172/2153-0637.1000131

(Perez, 2017a), Perez, Jean-claude. “The Master Code of Biology: from Prions and Prions-like Invariants to the Self-assembly Thesis.” Biomedical Journal of Scientific and Technical Research 1 (2017): 001-002.

(Perez, 2017b), : Jean Claude Perez. The Master Code of Biology: Self-assembly of two identical Peptides beta A4 1-43 Amyloid In Alzheimer’s Diseases. Biomed J Sci & Tech Res 1(4)- 2017. BJSTR.MS.ID.000394. DOI: 10.26717/BJSTR.2017.01.000394

(Perez§Montagnier, 2020), Perez, J. C., & Montagnier, L. . (2020). COVID-19, SARS AND BATS CORONAVIRUSES GENOMES PECULIAR HOMOLOGOUS RNA SEQUENCES. International Journal of Research -GRANTHAALAYAH, 8(7), 217–263. https://doi.org/10.29121/granthaalayah.v8.i7.2020.678

(Perez ; 2021a), Jean-Claude Perez, (2021). SARS-COV2 VARIANTS AND VACCINES MRNA SPIKES FIBONACCI NUMERICAL UA/CG METASTRUCTURES. International Journal of Research -GRANTHAALAYAH, 9(6), 349–396. https://doi.org/10.29121/granthaalayah.v9.i6.2021.4040

(Perez, 2021b), Perez, J. C. (2021). THE INDIA MUTATIONS AND B.1.617 DELTA VARIANTS: IS THERE A GLOBAL "STRATEGY" FOR MUTATIONS AND EVOLUTION OF VARIANTS OF THE SARS-COV2 GENOME?. International Journal of Research -GRANTHAALAYAH, 9(6), 418–459. https://doi.org/10.29121/granthaalayah.v9.i6.2021.4039

(Perez§Montagnier, 2021), Perez and Montagnier (2021) - Perez, J. C., & Montagnier, L. . (2021). SIX FRACTAL CODES OF LIFE FROM BIOATOMS ATOMIC MASS TO CHROMOSOMES NUMERICAL STANDING WAVES: THREE BREAKTHOUGHS IN ASTROBIOLOGY, CANCERS AND ARTIFICIAL INTELLIGENCE. International Journal of Research -GRANTHAALAYAH, 9(9), 133–191. DOI: https://doi.org/10.29121/granthaalayah.v9.i9.2021.4191

(Perez et al, 2021c), Jean Claude Perez, Valère Lounnas, Montagnier Montagnier (2021)

THE OMICRON VARIANT BREAKS THE EVOLUTIONARY LINEAGE OF SARS-COV2 VARIANTS. International Journal of Research -GRANTHAALAYAH, 9(12), 108. 10.29121/granthaalayah.v9.i12.2021.4418