Research Article
Open Access
CrossMark

Peculiar evolution of the Monkeypox virus genomes

Abstract

We compare the evolution of 14 genomes of monkeypox viruses including that of May 2022 that is currently spreading among humans in numerous countries outside Africa. Our aim was to discover mutations and other viral evolutions (recombination) of the virus genome that may explain the sudden impact of this epidemic circulating at very low-level and alert on its potential pathogenic character. We have evidenced the presence of a succession of a large number of T bases 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 found in succession (≤ 10) in the early genomes of 1971, up to the 19 T-base sequence in the Israel 2018 reference strain and the 30 T bases thereafter in the 2022 strains. We find a complementary match for this long sequence of T bases only in the simian hemorrhagic encephalitis virus, at the 3' end of the genome with a long succession of 28 A-bases after the stop codon. 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, Yersinia pestis, Escherichia coli and Penicillium nordicum. We wonder whether this region of the monkeypox genome situated right upstream this long T-repeat may potentially code for a not yet identified polypeptide sequences with a functional role.

Keywords: Monkeypox virus, biomathematics, master code, evolution, genomics, proteomics

Volume 7 Issue 1 - 2022

Jean Claude Perez
PhD Math's & Computer Science, Bordeaux University, France

Correspondence: Jean Claude Perez, PhD Math's & Computer Science, Bordeaux University, France, Tel 33 (0)5 40 00 27 88, Email jeanclaudeperez@gmail.com

Received: September 19, 2022 | **Published:** November 24, 2022

Introduction

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

Several cases of monkeypox viruses have been identified in a number of geographically distinct 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 (Figure 1).[1,2]

Nextstrain reference tree https://nextstrain.org/monkeypox?s=03

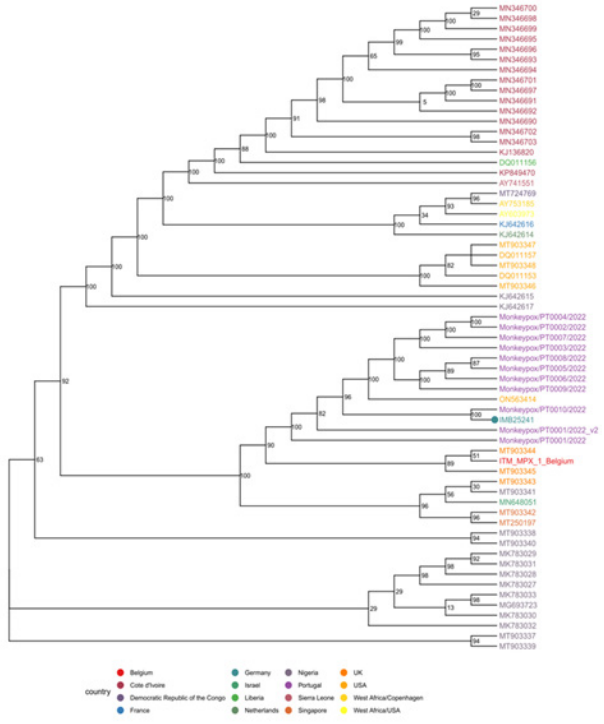

Figure 1 Monkeypox viruses tree (from https://virological.or g/t/first-german-genome- sequence-ofmonkeypox-virus-associated-to-multi-country-outbreak-in-may-2022/812)

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

Monkeypox virus Zaire-96-I-16

This particular outbreak has been identified as due to a virus from the West African clade that is often associated with a milder disease and, in this case, human-to-human spread is suspected. The first human to human strain referenced was identified in Israel in 2018:in a man who returned from Nigeria to Israel in 2018 Erez.[3]

Materials and methods

Monkeypox strains analyzed:

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

Singapore 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=98T3F4E013

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,[4] and Perez §Montagnier,[5] is a META-CODE based on the atomic masses common only to DNA, RNA and amino acids to highlight a It allows us to unify the 3 codes of DNA, RNA and amino acid sequences.

Specifically, our Master Code coupling curves is a measurement of measures the level of correlation unifying any pair of genomic sequences (DNA double strand) and its proteomics (amino acids) translated sequence, whether or not it may for a protein.

In a previous article Perez,[6] we have analyzed all types of prions in the early 2000s mad cow disease (present in plants, yeast, humans, cows, sheep, etc.). We then had then highlighted a possible "signature", a sort of invariant characteristic common to all prions. The typical signature of the Master Code unifying correlation take the shape of a "W" (or an "M" symmetrically). We had extended this type of analysis to amyloids implicated in the Alzheimer disease, Perez,[7]

Results

Table 1 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 bases. But in fact this sequence may also contain C bases substituted for T as both ttt and ttc codons are translated in phenyl-alanine residues. In that respect the length of the French sequence is actually equivalent to 21T. Sequencing errors are possible but not to the extent it would cover a range of 8 nucleotides. So the difference observed in 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 (ON622721https://www.ncbi.nlm.nih.gov/nuccore/ON622721.1/).

Table 1 Evolution of the T-bases contiguous region for the 14 genomes analyzed

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**

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 experts in the field.

For instance, if we look at the monkeypox strain Gabon-1988 we can identify in this region a sequence of nucleotides coding straightforwardly for a 42-aa long polypeptide chain that may constitute a small protein (Figure 2a, 2b).

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 ttaactttct ctttctcagc atctagttga tattccgacc
  135121 tcttcacgtt tcacatggt 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 ctctttgcac aatctcttaa
  135601 ctaacttcct ctctttgcttc ttcgtcatag tacttacaat cactatggga tccattgtta
  135661 ccacgtctgt actcggcgag ctcacgtttta 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 the 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

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 gtttcaggtt 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 tcttttctcag 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 attcttttttt ttttttttttt ttttttttttt 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 the 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 T-repeats repeat are common findings finding at the termination terminaison of a genome, as for instance at the end of the monkey encephalitis encephlitis virus, it is almost never encountered fully inside a whole genome 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 tttttttttttttttttttttttttttCGAATTCAC 133132

||||||||||||||||||||||||| |||||||

Sbjct 15370 TTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTCAC 15336

Why it is this peculiar nucleotide sequence located in this region of the genome?

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 Figure 3 this long T-base 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 genome sections of 100 bases framing both sides of the 30-T sequence shows it is a new functionality Figure 3 as is the case for the 19-T sequence in Figure 4.

Figure 3a According to the Master Ccode analysis of the whole USA2022 Monkeypox genome. The region of 44000 amino acids where there is the 30 T- bases insert .appears to be highly functional.

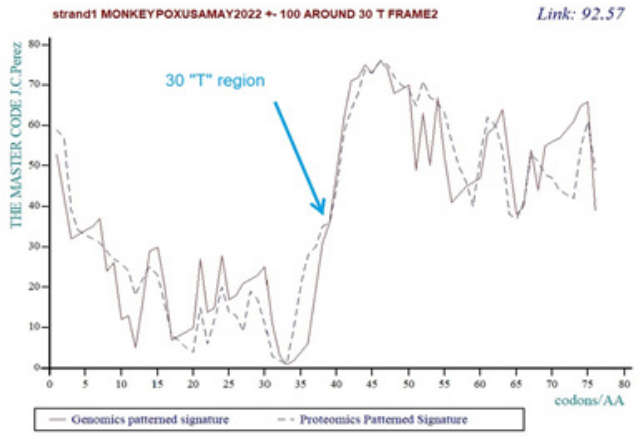

Figure 3b 100 bases upstream upload and downstream download the 30 T- bases region in the USA2022 strain.

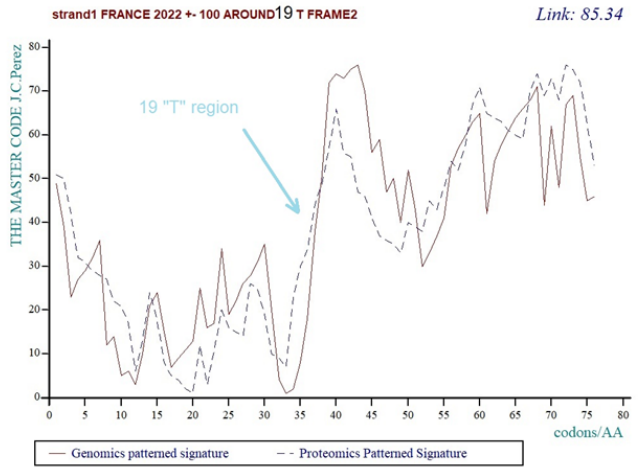

Figure 4 100 bases upstream upload and downstream download the 19 T-bases region in the FRANCE2022 strain.

Conclusions

The objective here was here to present how a new type of theoretical analysis helps identify a genome characteristic that would have otherwise remained unseen with the already established methods of mathematical genome analysis. Our findings may partly explain the sudden propagation of the monkeypox virus in the form we observed observe in quite a number of countries in May 2022. The role of the peculiar 30-T base long repeat sequence right in the middle of the virus genome is still to be determined experimentally. This work is an incentive for experimental investigations, for instance using a knockout genome (removing the T-repeat) among other possibilities.

Acknowledgments

None.

Conflicts of interest

Authors declare that there is no conflict of interest.

References

1. Markus H Antwerpen, Daniel Lang, Sabine Zange, et al. Bundeswehr Institute of Microbiology, Munich, Germany, First German genome sequence of Monkeypox virus associated to multi-country outbreak in May; 2022.
2. Isidro J, Borges V, Pinto M, et al. *First draft genome sequence of Monkeypox virus associated with the suspected multi-country outbreak*; 2022.
3. Erez, Noam, Hagit Achdout, Elad Milrot, et al. Diagnosis of Imported Monkeypox, Israel. *Emerg infect Dis*. 2019;25(5)980–983.
4. Perez JC. Deciphering Hidden DNA Meta-Codes -The Great Unification & Master Code of Biology. *J Glycomics Lipidomics*. 2015;5:131.
5. Perez JC. SARS-COV2 variants and vaccines mrna spikes fibonacci numerical UA/CG metastructures. *International Journal of Research -granthaalayah*. 2021;9(6):349–396.
6. Perez JC. The Master Code of Biology: from Prions and Prions-like Invariants to the Self-assembly Thesis. *Biomed J Sci & Tech Res*. 2017;1(4).
7. Perez JC. 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.* 2017;1(4):1191–1195.
8. Perez JC. SARS-COV2 variants and vaccines mrna spikes fibonacci numerical UA/CG metastructures. *International Journal of Research-Granthaalayah*. 2021;9(6):349–396.
9. Perez JC. The india mutations and b.1.617 delta variants: is there a global & quot;strategy" for mutations and evolution of variants of the sars-cov2 genome. *International Journal of Research-Granthaalayah*. 2021;9(6):418–459.
10. Perez JC, Montagnier L. 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*. 2021;9(9):133–191.
11. Perez JC, Lounnas V, Montagnier L. The omicron variant breaks the evolutionary lineage of sars-cov2 variants. *International Journal of Research-Granthaalayah*. 2021;9(12):108.