1 **Supplemental data**

2

3 **1. Amplification methodology**

4

5 DNA was extracted from skin swabs using QIAamP DNAmini kit ®. Then, in order to enrich
6 viral DNA, extracted DNA was digested using exonuclease V (Plasmid Safe, Epicentre ®
7 Biotechnologies, Madison, USA) and two restriction enzymes (Not1 and SalI ref company).
8 After precipitation, the purified DNA was randomly amplified using the Rolling Circle
9 Amplification (RCA) method with the REPLI-g Mini Kit (Qiagen, Hilden, Germany). PCRs
10 were performed on the amplified DNA preparation, in order to detect MCPyV genomes.
11 (Table S). Briefly, 1 µL of amplified DNA template obtained by RCA, was subjected to PCR
12 (Piko® Thermal Cyclers, Finnzymes, Finland) using the Phusion High-Fidelity DNA
13 polymerase (Finnzymes, Finland). Three different specific primer pairs (A, B and C)
14 amplifying MCPyV fragments of different sizes (Table S). PCRs were performed for 40
15 cycles (98°C for 30 sec, 98°C for 5 sec, annealing at 52°C for A/ 50°C for B and C) with an
16 extension of 72°C for 15 sec and another of 72°C for 1 min. Amplicons were analyzed by
17 agarose gel electrophoresis and directly sequenced.

18

19 **2. HTS methodology**

20

21 All raw reads were first submitted to a quality control procedure consisting of execution of
22 several programs belonging to the FASTX-Toolkit software (1). As the average quality of
23 data was judged acceptable, a light filtering policy was adopted: only low complexity
24 sequences and PCR duplicates were removed for samples. All remaining reads were then
25 mapped against the MCC350 reference genome using BWA 0.5.9rc1 software. We achieved

26 best results by setting BWA: i) to allow up to 3 mismatches, ii) by disabling usage of seeds in
27 order to gain sensibility, iii) and by letting BWA trim low-quality bases from the 3'-ends.
28 Resulting mapping files were then used to called consensus with samtools (2).
29

30 1. **Gordon, A. a. H., G. J. .** 2010. Fastx-toolkit. FASTQ/A short-reads pre-processing
31 tools.
32 2. **Li, H., and R. Durbin.** 2009. Fast and accurate short read alignment with Burrows-
33 Wheeler transform. Bioinformatics **25:**1754-1760.
34
35
36
37

38

38 **Legends**

39

40 **Table S. Sequences of the 3 primer pairs used for amplification of three fragments of**
41 **MCPyV genome from DNA extracted from skin swabs** : 3 primer pairs called A, B and C
42 were used to obtain a colinearized sequence of 1468 bp from 3 different PCRs.

43

44 **Fig. S1. Plot of reads coverage of the 3 complete MCPyV sequences obtained by HTS**
45 For each position of each sample, the raw number of reads covering this position (from 1 to
46 5387 bp) is plotted on the x-axis.

47

48 **Fig. S2. Phylogenetic tree generated with the Neighbor-Joining method (fragment A).**
49 The phylogenetic analysis was derived by the neighbor-joining method on 56 sequences of
50 436 bp of MCPyV from the fragment A, part of gene coding LT-Ag (position of the MCC350
51 prototype: 2136-2590). The phylogeny was derived by the NJ method by using the
52 "Hasegawa-Kishino-Yano" (HKY) model in the PAUP program version 4.0b10 (Sinauer
53 Associates, Sunderland, MA, USA). Reliability of the inferred tree was evaluated by
54 bootstrap analysis on 1,000 replicates. Branch lengths are drawn to scale, with the bar
55 indicating 0,001-nt replacement per site. The 22 new sequences generated in this study are in
56 bold (GenBank accession numbers: KF266906-KF266928).
57 The unexpected presence of the 5 African strains in the large Caucasian clade is linked to the
58 fact that, considering this short 436 bp fragment, no specific mutations are observed leading
59 to the existence of specific North America/Europe *vs* Africa, in contrast to the situation for the
60 Amerindian, Asian and Oceanian strains.

61

62 **Fig. S3. Phylogenetic tree generated with the Neighbor-Joining method (fragment B).**

63 The phylogenetic analysis was derived by the neighbor-joining method on 52 sequences of
64 590 bp of MCPyV from the fragment B, part of gene coding VP1 (position of the MCC350
65 prototype: 3380-3973). The phylogeny was derived by the NJ method by using the "General
66 Time-Reversible" (GTR) model in the PAUP program version 4.0b10 (Sinauer Associates,
67 Sunderland, MA, USA). Reliability of the inferred tree was evaluated by bootstrap analysis
68 on 1,000 replicates. Branch lengths are drawn to scale, with the bar indicating 0,001-nt
69 replacement per site. The 17 new sequences generated in this study are in bold (GenBank
70 accession numbers: KF266929-KF266945).
71
72
73

Table S. Sequences of the 3 primer pairs used for amplification of three fragments of MCPyV genome from DNA extracted from skin swabs : 3 primer pairs called A, B and C were used to obtain a colinearized sequence of 1468 bp from 3 different PCRs.

Primers	Region amplified on MCPyV genome	Primer forward	Primer reverse	amplicon size[b] (bp)
A	2136-2590 [a] (LT Ag)[*]	CAGAAAACCATTTGCCAA	GGTTATTTATTCCTTGCC	436
B	3380-3973 [a] (VP1)[*]	GATTATCTTTTCCTTCCAT	GAAAAGAGTTCATGATTA	574
C	4827-5303 [a] (VP2 and NCCR)[*]	TTATTAGAGAGCCTATAC	TTATTGCTGCAGGGTTTC	458

[a] : position of the primers according to the numeration of the prototypic strain 350 (Feng et al., 2008).
[b] : size without primer pairs

[*]: LT Ag for large T antigen; VP1 for viral protein 1; VP2 for viral protein 2 and NCCR for non-coding control region

Fig. S1. Plot of reads coverage of the 3 complete MCPyV sequences obtained by HTS

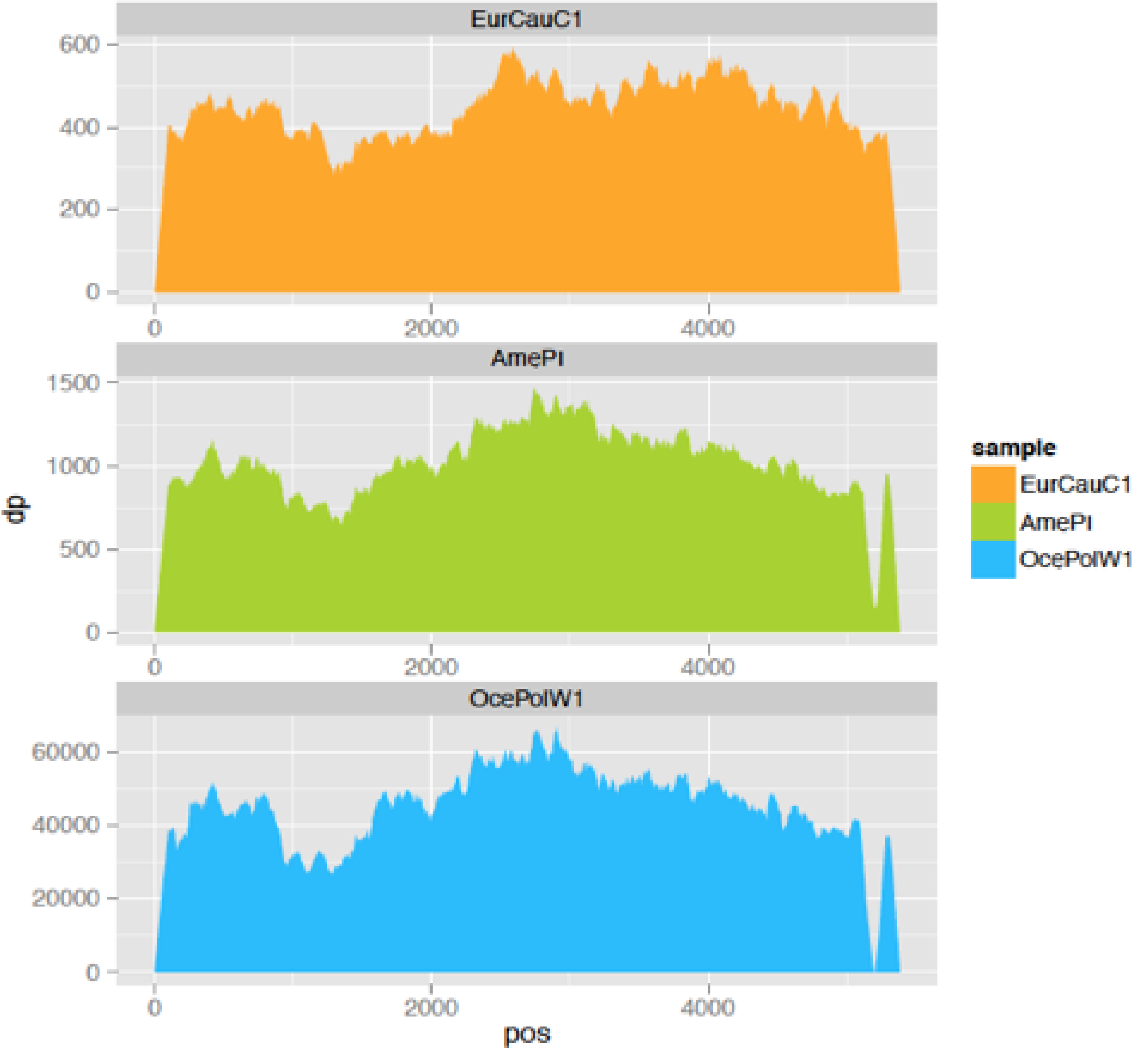

Fig. S2. Phylogenetic tree generated with the Neighbor-Joining method (fragment A).

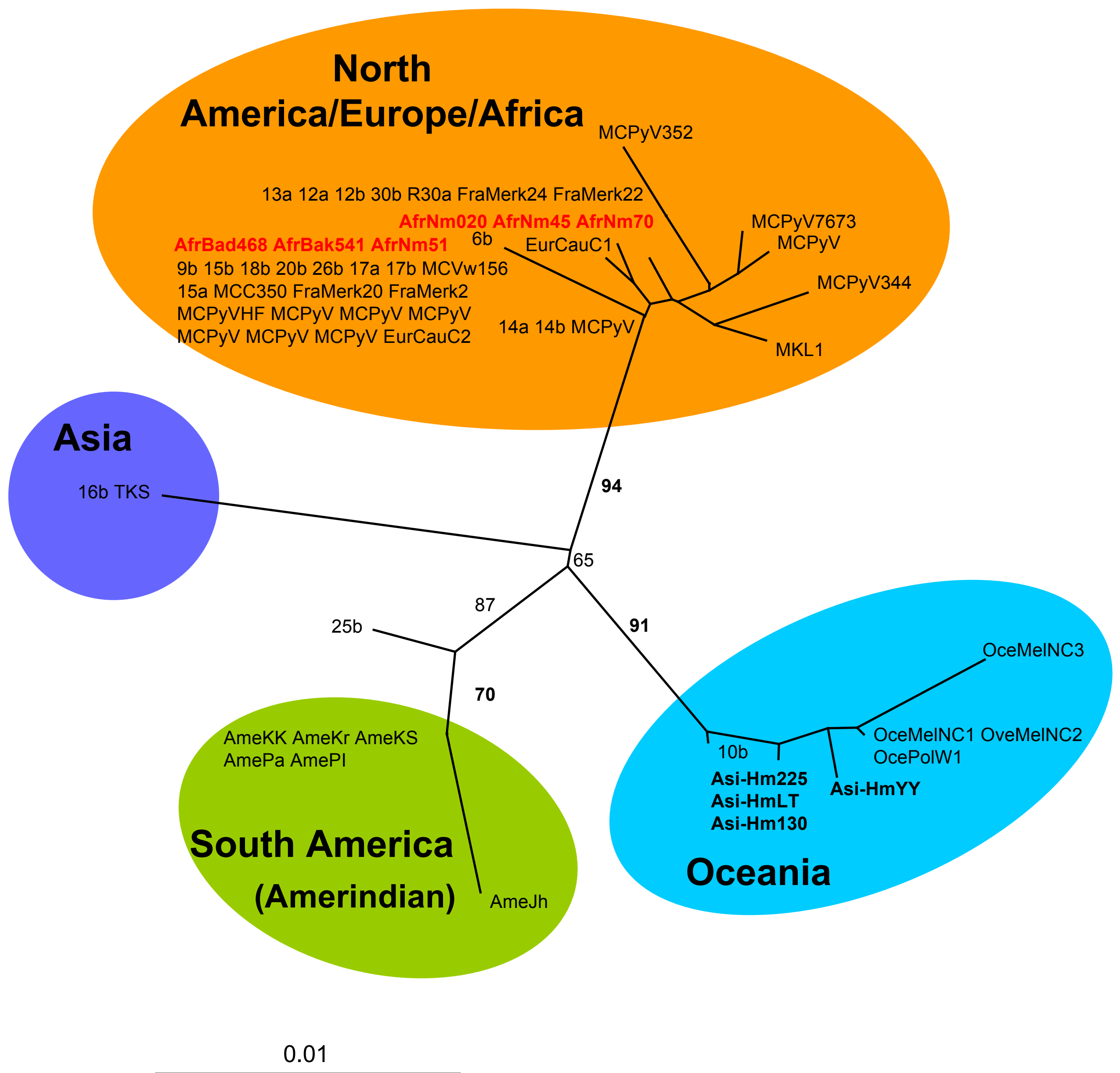

Fig. S3. Phylogenetic tree generated with the Neighbor-Joining method (fragment B).