Electronic Supplement

Materials and Methods

In-silico reconstruction of archetype-like sequences from samples E4 and F1

There were no archetype-like sequences among analyzed clones from the 5th urine sample from patient E. However for patient E, mrNCRR architecture for variant V15.1 consisted of a complete archetype-like sequence with a single insertion of segment P60 - Q19. Since the sequence of this repeat was identical to the segment it repeated, it was deleted *in silico* to artificially recreate the archetype-like sequence from which mrNCRR V15 was presumably derived. This *in silico* construct, "E_2.1_reconstruct", mapped to lineage Ib-2 (Fig. 3).

Similarly, there were no archetype-like sequences among analyzed clones from the urine sample from patient F. However, none of the clones from patient F had a complete R segment. The most complete *in silico* archetype recreation for patient F, [P1-68, Q1-39, R20-42, R52-63, S1-63], was reconstructed *in silico* by first removing the P20-Q34 sequence insert from variant mrNCRR V4.1 (duplicated sequences were identical) and then extending the R segments from this isolate with slightly longer R segments from the other variants that had identical sequences in the portions of their R segments that overlapped with those in variant V4.1. This *in silico* reconstruct with gaps showed highest similarity to genogroup 1b-1 (not shown).

Figure legends

Electronic Supplement Fig. 1. BKV phylogenetic analysis based on complete BKV sequences from which NCRRs had been deleted *in silico.*

The **c**omplete BKV viral sequences of all isolates analyzed in Zheng et al (1) and 3 sequences from Sharma et al were downloaded from the EMBL database. Isolates with non-archetype-like NCRR were excluded from further analysis. The remaining sequences were aligned with the Sequencher program and used to generate a data files of the complete sequence of each isolate minus its NCRR sequence (as in Zheng et al., (1)). Sequences in each data set were aligned, the data bootstrapped 1000 times, and a nearest-

neighbor phylogenetic tree constructed using ClustalX. The tree was visualized using njplot (large image) with WW sequence used as the reference outgroup. An unrooted trees was also visualized using the unrooted program (inserts at the bottom left of each panel). For visual clarity only the names of the genotypes and sub-genotypes appear for the unrooted tree. [Nomenclature convention: Isolates are designated by their genotype and subgentoype according to Zheng et al (1), the isolate name and then the EMBL access number]

.

Electronic Supplement Fig. 2. BKV phylogenetic analysis based on the NCRRs that had been excised *in silico* from full length BKV genomic sequences analyzed for Electronic Supplement Fig.1.

Non-archetype NCRR sequences were excised *in silico* from the complete BKV viral sequences of all isolates with archetype NCRRs analyzed in Electronic Supplement Fig. 1. The archetype NCRR sequences that remained were aligned with the Sequencher program and used to generate a data file that was aligned, bootstrapped 1000 times, and used to construct a nearest-neighbor phylogenetic tree with ClustalX. The tree was visualized using njplot (large image). The WW NCRR sequence was used as the reference outgroup. [Nomenclature convention as in Electronic Supplement Fig. 1]. An unrooted tree was also visualized using the unrooted program (inserts at the bottom left of each panel). For visual clarity only the names of the genotypes and sub-genotypes appear for the unrooted tree.

1. **Zheng, H. Y., Y. Nishimoto, Q. Chen, M. Hasegawa, S. Zhong, H. Ikegaya, N. Ohno, C. Sugimoto, T. Takasaka, T. Kitamura, and Y. Yogo.** 2007. Relationships between BK virus lineages and human populations. Microbes Infect **9:**204-13.

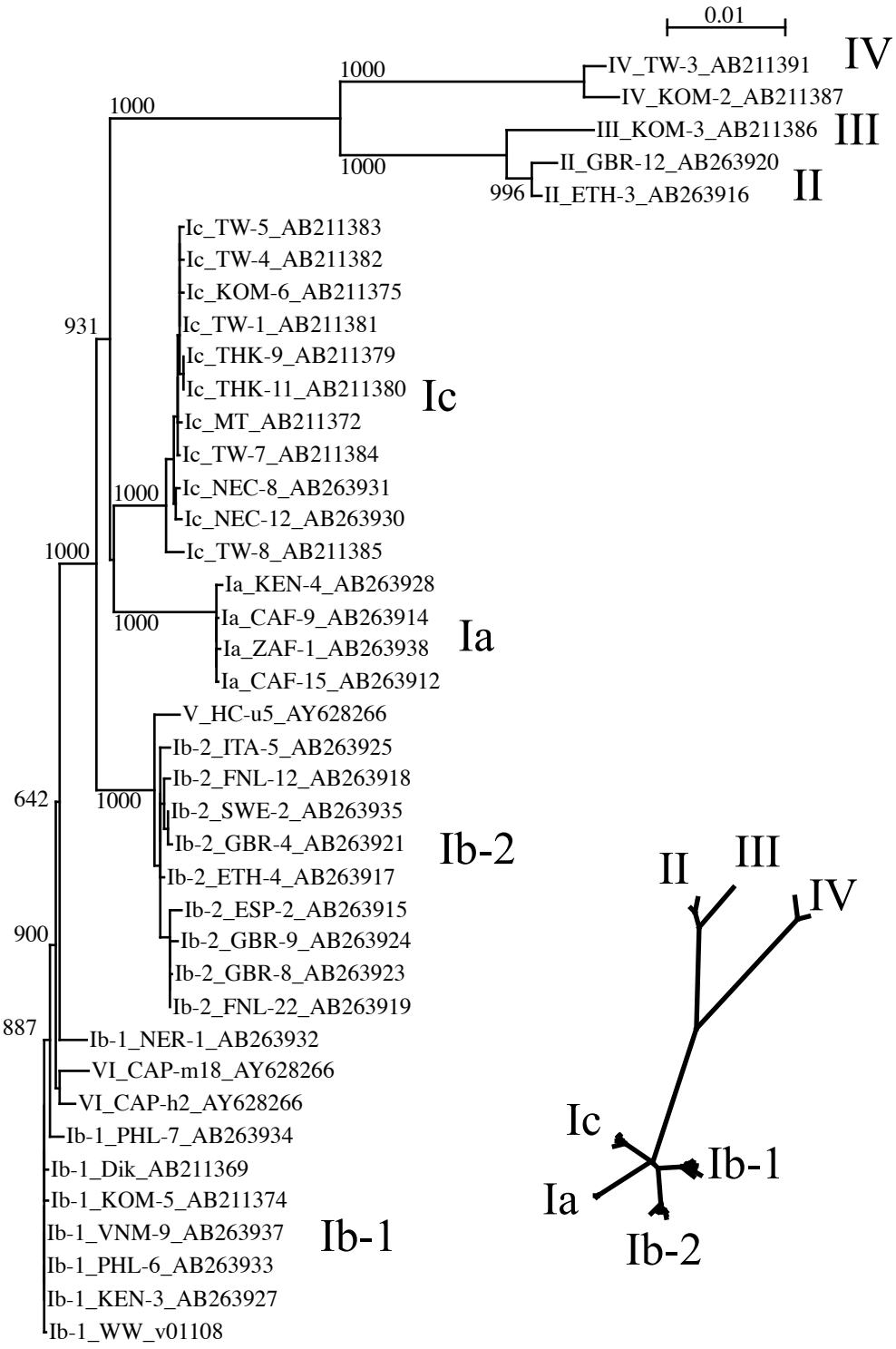
0.01
IV
III
II
Ic
Ia
Ib-2
Ib-1
IV_TW-3_AB211391
IV_KOM-2_AB211387
III_KOM-3_AB211386
II_GBR-12_AB263920
II_ETH-3_AB263916
Ic_TW-5_AB211383
Ic_TW-4_AB211382
Ic_KOM-6_AB211375
Ic_TW-1_AB211381
Ic_THK-9_AB211379
Ic_THK-11_AB211380
Ic_MT_AB211372
Ic_TW-7_AB211384
Ic_NEC-8_AB263931
Ic_NEC-12_AB263930
Ic_TW-8_AB211385
Ia_KEN-4_AB263928
Ia_CAF-9_AB263914
Ia_ZAF-1_AB263938
Ia_CAF-15_AB263912
V_HC-u5_AY628266
Ib-2_ITA-5_AB263925
Ib-2_FNL-12_AB263918
Ib-2_SWE-2_AB263935
Ib-2_GBR-4_AB263921
Ib-2_ETH-4_AB263917
Ib-2_ESP-2_AB263915
Ib-2_GBR-9_AB263924
Ib-2_GBR-8_AB263923
Ib-2_FNL-22_AB263919
Ib-1_NER-1_AB263932
VI_CAP-m18_AY628266
VI_CAP-h2_AY628266
Ib-1_PHL-7_AB263934
Ib-1_Dik_AB211369
Ib-1_KOM-5_AB211374
Ib-1_VNM-9_AB263937
Ib-1_PHL-6_AB263933
Ib-1_KEN-3_AB263927
Ib-1_WW_v01108
1000
1000
1000
996
931
1000
1000
1000
642
900
887

Fig. ES 1 PVAN

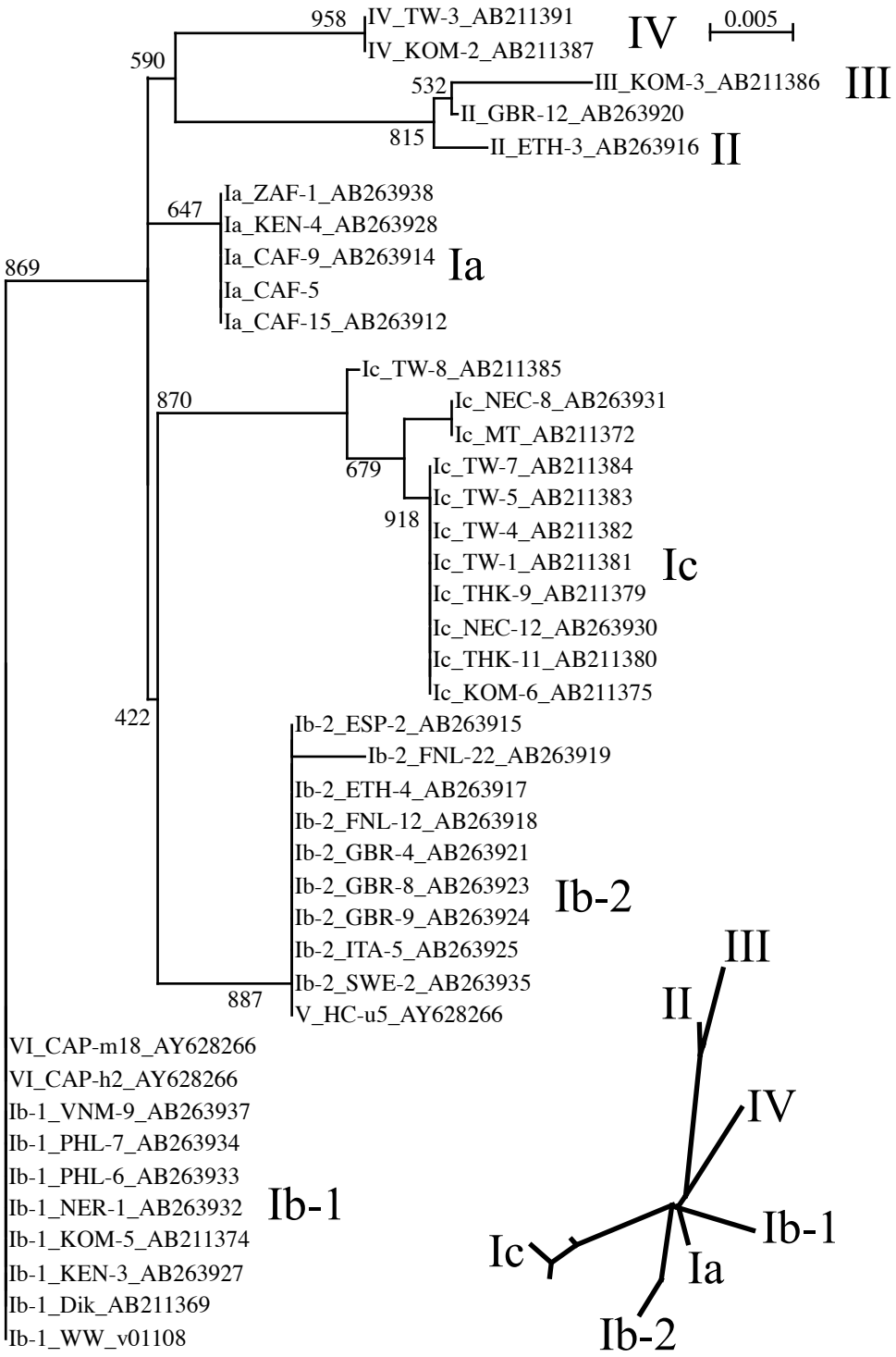
958
IV_TW-3_AB211391
IV_KOM-2_AB211387
IV
0.005
590
532
III_KOM-3_AB211386
III
II_GBR-12_AB263920
815
II_ETH-3_AB263916
II
647
Ia_ZAF-1_AB263938
Ia_KEN-4_AB263928
Ia_CAF-9_AB263914
Ia
869
Ia_CAF-5
Ia_CAF-15_AB263912
Ic_TW-8_AB211385
870
Ic_NEC-8_AB263931
Ic_MT_AB211372
679
Ic_TW-7_AB211384
Ic_TW-5_AB211383
918
Ic_TW-4_AB211382
Ic_TW-1_AB211381
Ic
Ic_THK-9_AB211379
Ic_NEC-12_AB263930
Ic_THK-11_AB211380
Ic_KOM-6_AB211375
422
Ib-2_ESP-2_AB263915
Ib-2_FNL-22_AB263919
Ib-2_ETH-4_AB263917
Ib-2_FNL-12_AB263918
Ib-2_GBR-4_AB263921
Ib-2_GBR-8_AB263923
Ib-2
Ib-2_GBR-9_AB263924
Ib-2_ITA-5_AB263925
Ib-2_SWE-2_AB263935
887
V_HC-u5_AY628266
VI_CAP-m18_AY628266
VI_CAP-h2_AY628266
Ib-1_VNM-9_AB263937
Ib-1_PHL-7_AB263934
Ib-1_PHL-6_AB263933
Ib-1_NER-1_AB263932
Ib-1
Ib-1_KOM-5_AB211374
Ib-1_KEN-3_AB263927
Ib-1_Dik_AB211369
Ib-1_WW_v01108
III
II
IV
Ic
Ia
Ib-1
Ib-2

Fig. ES 2 PVAN