SCIENTIFIC REPORTS

Corrected: Publisher Correction

OPEN

Discovery and genetic characterization of diverse smacoviruses in Zambian non-human primates

Received: 24 January 2019
Accepted: 7 March 2019
Published online: 08 April 2019

Paulina D. Anindita[1], Michihito Sasaki[1], Gabriel Gonzalez[2], Wallaya Phongphaew[1], Michael Carr[3,4], Bernard M. Hang'ombe[5,7,8], Aaron S. Mweene[6,7,8], Kimihito Ito[2,3], Yasuko Orba[1] & Hirofumi Sawa[1,3,7,8,9]

The *Smacoviridae* has recently been classified as a family of small circular single-stranded DNA viruses. An increasing number of smacovirus genomes have been identified exclusively in faecal matter of various vertebrate species and from insect body parts. However, the genetic diversity and host range of smacoviruses remains to be fully elucidated. Herein, we report the genetic characterization of eleven circular replication-associated protein (*Rep*) encoding single-stranded (CRESS) DNA viruses detected in the faeces of Zambian non-human primates. Based on pairwise genome-wide and amino acid identities with reference smacovirus species, ten of the identified CRESS DNA viruses are assigned to the genera *Porprismacovirus* and *Huchismacovirus* of the family *Smacoviridae*, which bidirectionally encode two major open reading frames (ORFs): *Rep* and capsid protein (*CP*) characteristic of a type IV genome organization. The remaining unclassified CRESS DNA virus was related to smacoviruses but possessed a genome harbouring a unidirectionally oriented *CP* and *Rep*, assigned as a type V genome organization. Moreover, phylogenetic and recombination analyses provided evidence for recombination events encompassing the 3′-end of the *Rep* ORF in the unclassified CRESS DNA virus. Our findings increase the knowledge of the known genetic diversity of smacoviruses and highlight African non-human primates as carrier animals.

Recent advances in high-throughput sequencing technologies have allowed metagenomic analyses to discover an ever-increasing genetic diversity of viral genomes from vertebrates, invertebrates, prokaryotic and environmental samples[1]. Small circular replication-associated protein (*Rep*) encoding single-stranded (CRESS) DNA viruses have been discovered in a diversity of prokaryotes and eukaryotes[2]. With the increasing diversity of CRESS DNA viruses, they have been classified into six viral families: *Genomoviridae*, *Geminiviridae*, *Nanoviridae*, *Circoviridae*, *Bacilladnaviridae* and *Smacoviridae*. The family *Smacoviridae* was recently assigned as a new viral family by the International Committee on Taxonomy of Viruses (ICTV)[3], which is further classified into six genera. Smacoviruses have 2.3–2.9 kb genomes, containing two major open reading frames (ORFs), encoding *Rep* and the capsid protein (*CP*). Rep possesses DNA helicase activity and initiates viral replication by a rolling circle replication (RCR)[4]. In comparison with *Rep*, the *CP* ORF is more divergent among smacoviruses. The genetic diversity observed within smacoviruses might be due to high mutation rates[5] and intra-familial recombination events in their genomes[6].

[1]Division of Molecular Pathobiology, Research Center for Zoonosis Control, Hokkaido University, Sapporo, 001-0020, Japan. [2]Division of Bioinformatics, Research Center for Zoonosis Control, Hokkaido University, Sapporo, 001-0020, Japan. [3]Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, Sapporo, 001-0020, Japan. [4]National Virus Reference Laboratory, School of Medicine, University College Dublin, Belfield, Dublin, 4, Ireland. [5]Department of Paraclinical Studies, School of Veterinary Medicine, University of Zambia, PO Box 32379, Lusaka, 10101, Zambia. [6]Department of Disease Control, School of Veterinary Medicine, University of Zambia, PO Box 32379, Lusaka, 10101, Zambia. [7]Africa Centre of Excellence for Infectious Diseases of Humans and Animals, University of Zambia, P.O. Box 32379, Lusaka, 10101, Zambia. [8]Global Virus Network Affiliate, University of Zambia, P.O. Box 32379, Lusaka, 10101, Zambia. [9]Global Virus Network, 801 W. Baltimore St., Baltimore, MD, 21201, USA. Correspondence and requests for materials should be addressed to M.S. (email: m-sasaki@czc.hokudai.ac.jp)

Sample	Non-human primate species (common name)	Corresponding primer	Virus	Abbreviation	Accession number
ZM09#51	*Papio cynocephalus* (yellow baboon)	ssDNAV-3	*Papio cynocephalus* associated smacovirus 3/ZM09-51	PcSmV3-ZM09-51	LC386205
ZM09#64	*P. kindae* (Kinda yellow baboon)	ssDNAV-1	*Papio kindae* associated smacovirus-like virus 1/ZM09-64	PkSmV1-ZM09-64	LC386203
ZM09#71	*P. cynocephalus* (yellow baboon)	ssDNAV-2	*Papio cynocephalus* associated smacovirus 2/ZM09-71	PcSmV2-ZM09-71	LC386204
ZM09#72	*P. cynocephalus* (yellow baboon)	ssDNAV-6	*Papio cynocephalus* associated smacovirus 6/ZM09-72	PcSmV6-ZM09-72	LC386201
ZM09#74	*P. cynocephalus* (yellow baboon)	ssDNAV-3	*Papio cynocephalus* associated smacovirus 3/ZM09-74	PcSmV3-ZM09-74	LC386197
ZM09#76	*P. kindae* (Kinda yellow baboon)	ssDNAV-3	*Papio kindae* associated smacovirus 3/ZM09-76	PkSmV3-ZM09-76	LC386198
ZM09#83	*Chlorocebus cynosuros* (malbrouck)	ssDNAV-5	*Chlorocebus cynosuros* associated smacovirus 5/ZM09-83	CcSmV5-ZM09-83	LC386200
ZM09#86	*C. cynosuros* (malbrouck)	ssDNAV-1	*Chlorocebus cynosuros* associated smacovirus 1/ZM09-86	CcSmV1-ZM09-86	LC386195
ZM09#95	*C. cynosuros* (malbrouck)	ssDNAV-4	*Chlorocebus cynosuros* associated smacovirus 4/ZM09-95	CcSmV4-ZM09-95	LC386199
ZM09#96	*C. cynosuros* (malbrouck)	ssDNAV-1	*Chlorocebus cynosuros* associated smacovirus 1/ZM09-96	CcSmV1-ZM09-96	LC386196
ZM09#96	*C. cynosuros* (malbrouck)	ssDNAV-6	*Chlorocebus cynosuros* associated smacovirus 6/ZM09-96	CcSmV6-ZM09-96	LC386202

Table 1. Sample information and results of PCR screening.

Smacoviruses, previously known as "stool-associated circular viruses", have been detected in faecal samples obtained from healthy and diarrheic animal species, including cattle[7], sheep[7], pigs[8,9], rats[10], chickens[11], camels[12], non-human primates[13] and humans[13–15], as well as insect species such as dragonflies[16] and blow flies[17] but not from environmental samples. Despite the lack of evidence for a direct causal relationship, smacoviruses were identified in the faecal virome derived from human patients with diarrhea in France as well as in central and south American children with unexplained gastrointestinal disease negative for known pathogens[13,15]. It remains, however, to be established whether smacovirus infect human cells, causes overt disease or not in humans and animals.

In this study, sequence reads related to CRESS DNA virus genomes were initially discovered in faecal samples of Zambian NHPs through viral metagenomic analysis. We subsequently determined whole genome sequences of eleven CRESS DNA viruses and characterized ten of them as new smacovirus species. This study extends the known genetic diversity of smacoviruses and the species range of NHPs which harbour these ssDNA viruses.

Results

Identification of CRESS DNA viruses in Zambian NHP species.

Fifty faecal samples from NHPs consisting of 25 malbroucks (*Chlorocebus cynosuros*) and 25 baboons (*Papio spp.*) in Zambia were suspended, pooled and subjected to metagenomic analysis. Among a total of 63,587,648 sequence reads generated, 1,381,545 reads were assigned to ssDNA viruses by BLASTx by comparison of the translated nucleotide sequences from the samples with the viral protein database[18]. Six contigs, ranging in length from 0.8–2.0 kb, related to members of *Smacoviridae* were generated by *de novo* assembly.

To examine the prevalence of the smacovirus-like genomes, six different pairs of primer sets were designed and used to screen fifty faecal samples from NHPs (Supplementary Table S1). Ten (20%) of the NHP faecal samples were positive for smacovirus-like genomes. Nine faecal samples were positive for a single smacovirus-like genome, whereas one malbrouck faecal sample (ZM09#96) harboured two different smacovirus-like genomes (Table 1). Speciation of NHPs was confirmed by sequencing of mitochondrial *cytochrome b* (*cytb*) (Table 1).

The complete, circular genomes from eleven smacovirus-like genomes were then amplified by inverse PCR, cloned into plasmid vectors and then sequenced bidirectionally by a primer walking strategy employing Sanger sequencing. As a result, we found that the complete circular genome sizes ranged from 2488 to 2766 nucleotides, which is within the known range of previously reported CRESS DNA virus genomes. BLASTx analysis showed that these CRESS DNA virus genomes were related to known viruses from the family *Smacoviridae*.

Classification and genome organization of Zambian NHP CRESS DNA viruses.

The smacovirus-like CRESS DNA viruses from Zambian NHPs each contained two large ORFs which showed sequence similarity to *CP* and *Rep* of previously described smacoviruses. Following the CRESS DNA virus classification scheme proposed by Rosario *et al.*[4], the genomes of ten CRESS DNA viruses belonged to the ambisense type IV organization whereas the genome of one CRESS DNA virus (isolate PkSmV1-ZM09–64) contained a

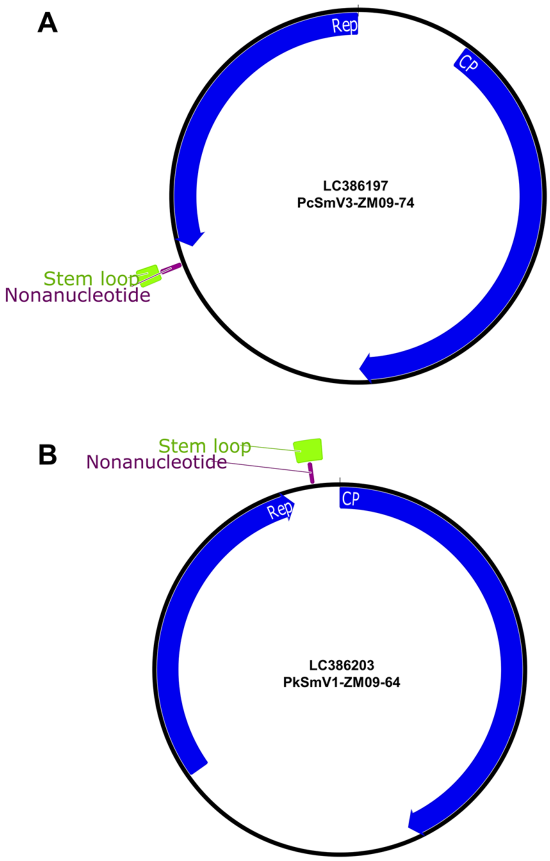

Figure 1. Genome organization of the representative smacovirus carrying a ambisense type IV genome (**A**) and CRESS DNA virus carrying a unisense type V genome (**B**) from Zambian NHPs showing ORFs (*Rep*, rolling replication-associated protein; *CP*, capsid protein), predicted stem loop and nonanucleotide.

unisense type V organization in which the *CP* and *Rep* ORFs were in the same orientation similar to the previously described porcine smacovirus-related CRESS DNA virus, PigSCV (JQ023166)[13,19] (Fig. 1). The predicted stem loop structures were located near the 3′-end of the *Rep* ORF with homology to the degenerate NAGTNTTAC nonanucleotide sequence motif which are also shared by all other reported smacoviruses[6] (Table 2). This motif has been identified as the putative origin of RCR of smacoviruses during the replication cycle[7,13].

To further characterize the identified viruses as CRESS DNA viruses, we also searched for amino acid motifs within the encoded Rep that play important functional roles in viral genome replication. All Rep proteins of the identified CRESS DNA viruses harboured an RCR domain (motifs I, II and III) and a helicase super family 3 domain (Walker A, B and C) as illustrated in Table 2. Interestingly, a variation of amino acid residues within the RCR motifs I and II was found throughout the Rep proteins of the identified CRESS DNA viruses compared to the previously described consensus motifs[7]. Notably, ten of eleven CRESS DNA viruses encoded Rep proteins possessing 5 amino acid residues within RCR motif I whereas the Rep of isolate PkSmV1-ZM09-64 had 6 amino acid residues (Table 2). In addition, the *Rep* ORF of PkSmV1-ZM09-64 encoded a leucine residue at the beginning of the Walker B motif which is unseen in smacoviruses, which possess isoleucine, valine or tryptophan at the first residue (Table 2).

The pairwise nucleotide sequence identities were calculated to determine the genetic distances between the identified CRESS DNA virus genomes and previously described smacoviruses (Table 3). The isolate PkSmV1-ZM09-64, carrying a unisense type V genome, was not assigned to a virus family and excluded from this analysis due to the inversion of the replication-associated protein ORF (Fig. 1). All of the identified genomes except PkSmV1-ZM09-64 had <77% genome-wide pairwise nucleotide sequence identity. Based on the smacovirus species demarcation threshold of 77% genome-wide pairwise nucleotide identity[6], we grouped these CRESS DNA viruses into five smacovirus species (species 2–6 in Table 3). Species 3 and 6 were only identified from *C. cynosuros*, whereas species 4 and 5 were identified from 2 different NHP species.

Pairwise amino acid sequence identity was calculated for the Rep proteins of all smacoviruses from Zambian NHPs (Zm-SmVs) and that of known smacoviruses (Table 3). Among the Zm-SmVs species, four species belonged to the genus *Porprismacovirus* while a single species was most closely related to the genus *Huchismacovirus* following the smacovirus genus demarcation threshold of 40% pairwise amino acid sequence identity of Rep[6]. PcSmV6-ZM09-72 and CcSmV6-ZM09-96 were assigned to the genus *Porprismacovirus* as they were classified as closely-related species to other Porprismacoviruses (CcSmV4-ZM09-95 and CcSmV5-ZM09-83, respectively).

Virus	Genome length (nt)	Rep (aa)	CP (aa)	Nonanucleotide motif	RCR motif I	RCR motif II	RCR motif III	Walker A	Walker B	Walker C
PkSmV1-ZM09-64	2544	259	368	AGGTCTTAC	ATVWVD	HFQFRG	YVYK	SRGNWGKT	LVDTP	VLCN
CcSmV1-ZM09-86	2691	284	369	AATGATTAC	ATIDR	HWQCRF	YETK	PTGNIGKS	IIDIP	VMTN
CcSmV1-ZM09-96	2659	268	182	AAAAATTAC	VTMGR	HWQVRF	YEAK	ETGNRGKS	VIDIP	VLTN
PcSmV2-ZM09-71	2766	271	359	AATGATTAC	CTVPN	HFQLRC	YVEK	RVGGRGKT	VIDMP	VITN
PcSmV3-ZM09-51	2540	269	312	TAGTGTTAC	MTAPR	HWQCRI	YEAK	ETGNVGKS	IIDVP	VMTN
PcSmV3-ZM09-74	2488	245	330	TAGTATTAC	GTAPR	HWQIRI	YEAK	ETGNVGKS	IIDVP	VMTN
PkSmV3-ZM09-76	2542	269	330	TAGTGTTAC	MTAPR	HWQCRI	YEAK	ETGNRGKS	IIDVP	VMTN
CcSmV4-ZM09-95	2656	260	271	TAGTATTAC	LTIPR	HWQVVV	YCRK	VRGGHGKT	WIDLP	VTTN
CcSmV5-ZM09-83	2617	261	354	TAGTATTAC	GTISA	HYQYTV	YCKK	GGGHGKT	WIDLP	VTTN
PcSmV6-ZM09-72	2613	269	353	TAGTATTAC	GTISA	HFQYYI	YCKK	VGGAHGKT	WIDIP	VTTN
CcSmV6-ZM09-96	2537	269	354	TAGTATTAC	GTISA	HFQYCI	YCKK	VGGAHGKT	WIDLP	VTTN

Table 2. Genome features of Zambian non-human primate CRESS DNA viruses.

Species	Isolate	Closest genome sequence	Identity (%)	Closest genome sequence of known smacovirus species	Identity (%)	Closest Rep sequence of known smacovirus species	Identity (%)	Genus
1	(LC386203) PkSmV1-ZM09-64	*		*		(KT862223) Bovine faeces associated smacovirus 1	42.73	unassigned
2	(LC386204) PcSmV2-ZM09-71	(LC386198) PkSmV3-ZM09-76	56.58	(KP233194) Lemur smacovirus	55.46	(KT862223) Bovine faeces associated smacovirus 1	46.86	*Huchismacovirus*
3	(LC386195) CcSmV1-ZM09-86	(LC386196) CcSmV1-ZM09-96	82.29	(KT600068) Human feces smacovirus 2	64.69	(KT600068) Human feces smacovirus 2	75.18	*Porprismacovirus*
3	(LC386196) CcSmV1-ZM09-96	(LC386195) CcSmV1-ZM09-86	82.29	(KT600068) Human feces smacovirus 2	62.44	(GQ351274) Chimpanzee stool associated circular ssDNA virus	64.48	*Porprismacovirus*
4	(LC386205) PcSmV3-ZM09-51	(LC386198) PkSmV3-ZM09-76	99.80	(KJ577810) Porcine stool associated circular virus 1	67.15	(JX274036) Porcine stool associated circular virus	81.07	*Porprismacovirus*
4	(LC386197) PcSmV3-ZM09-74	(LC386198) PkSmV3-ZM09-76	92.86	(KJ577810) Porcine stool associated circular virus 1	67.84	(JX274036) Porcine stool associated circular virus	83.95	*Porprismacovirus*
4	(LC386198) PkSmV3-ZM09-76	(LC386205) PcSmV3-ZM09-51	99.80	(KJ577810) Porcine stool associated circular virus 1	67.32	(JX274036) Porcine stool associated circular virus	81.48	*Porprismacovirus*
5	(LC386199) CcSmV4-ZM09-95	(LC386201) PcSmV6-ZM09-72	79.62	(KT600068) Human feces smacovirus 2	55.70	(KT862221) Sheep faeces associated smacovirus 2	42.08	*Porprismacovirus*
5	(LC386201) PcSmV6-ZM09-72	(LC386199) CcSmV4-ZM09-95	79.62	(KT600068) Human feces smacovirus 2	55.12	(KT862218) Bovine faeces associated smacovirus 2	39.24	*Porprismacovirus*
6	(LC386200) CcSmV5-ZM09-83	(LC386202) CcSmV6-ZM09-96	80.14	(KT600068) Human feces smacovirus 2	55.00	(KT862218) Bovine faeces associated smacovirus 2	40.92	*Porprismacovirus*
6	(LC386202) CcSmV6-ZM09-96	(LC386200) CcSmV5-ZM09-83	80.14	(KP233194) Lemur smacovirus	56.65	(KT862218) Bovine faeces associated smacovirus 2	39.66	*Porprismacovirus*

Table 3. Pairwise sequence identity among Zambian non-human primate CRESS DNA viruses and known smacoviruses. *Not comparable due to the inversion of the replication-associated protein ORF.

Phylogenetic relationships among Zambian CRESS DNA viruses and known smacoviruses. Phylogenetic trees were constructed based on the analyses of the whole genome sequences (Fig. 2), and, separately, of the amino acid sequences of CP (Fig. 3) and Rep (Fig. 4) using a maximum likelihood (ML) estimation coupled with a Bayesian inference. The genome-wide phylogenetic tree revealed that ZM-SmVs, shown in red color in the tree, segregated into a cluster of previously reported smacoviruses (Fig. 2). The phylogenetic tree of Rep supported the genus assignment of the ZM-SmVs: PcSmV3-ZM09-74, PkSmV3-ZM09-76, PcSmV3-ZM09-51, CcSmV1-ZM09-86 and CcSmV1-ZM09-96 formed a cluster with members of *Porprismacovirus* and PcSmV2-ZM09-71 shared a common ancestor with members of *Huchismacovirus* (Fig. 4). CcSmV4-ZM09-95, CcSmV5-ZM09-83, PcSmV6-ZM09-72 and CcSmV6-ZM09-96 were distinct from known smacoviruses with low amino acid identities for their Rep proteins (Table 3).

Even though a clear and unambiguous congruence between the CP and Rep phylogenies was not apparent, PcSmV6-ZM09-72, CcSmV5-ZM09-83, CcSmV4-ZM09-95, and CcSmV6-ZM09-96 were clustered together in

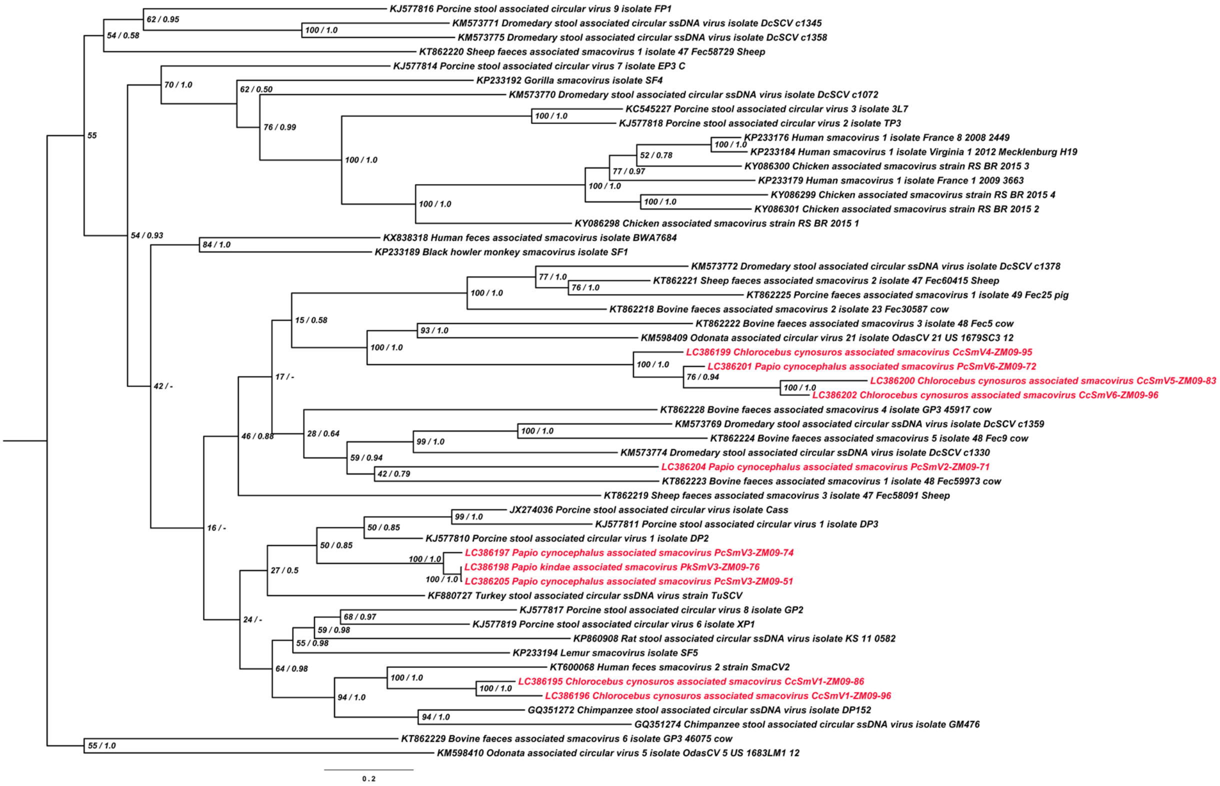
KJ577816 Porcine stool associated circular virus 9 isolate FP1
62 / 0.95
KM573771 Dromedary stool associated circular ssDNA virus isolate DcSCV c1345
100 / 1.0
54 / 0.58
KM573775 Dromedary stool associated circular ssDNA virus isolate DcSCV c1358
KT862220 Sheep faeces associated smacovirus 1 isolate 47 Fec58729 Sheep
KJ577814 Porcine stool associated circular virus 7 isolate EP3 C
KP233192 Gorilla smacovirus isolate SF4
70 / 1.0
62 / 0.50
KM573770 Dromedary stool associated circular ssDNA virus isolate DcSCV c1072
KC545227 Porcine stool associated circular virus 3 isolate 3L7
100 / 1.0
KJ577818 Porcine stool associated circular virus 2 isolate TP3
76 / 0.99
KP233176 Human smacovirus 1 isolate France 8 2008 2449
100 / 1.0
KP233184 Human smacovirus 1 isolate Virginia 1 2012 Mecklenburg H19
55
52 / 0.78
KY086300 Chicken associated smacovirus strain RS BR 2015 3
100 / 1.0
77 / 0.97
KP233179 Human smacovirus 1 isolate France 1 2009 3663
100 / 1.0
KY086299 Chicken associated smacovirus strain RS BR 2015 4
100 / 1.0
100 / 1.0
KY086301 Chicken associated smacovirus strain RS BR 2015 2
KY086298 Chicken associated smacovirus strain RS BR 2015 1
54 / 0.93
KX838318 Human feces associated smacovirus isolate BWA7684
84 / 1.0
KP233189 Black howler monkey smacovirus isolate SF1
KM573772 Dromedary stool associated circular ssDNA virus isolate DcSCV c1378
77 / 1.0
KT862221 Sheep faeces associated smacovirus 2 isolate 47 Fec60415 Sheep
100 / 1.0
76 / 1.0
KT862225 Porcine faeces associated smacovirus 1 isolate 49 Fec25 pig
KT862218 Bovine faeces associated smacovirus 2 isolate 23 Fec30587 cow
15 / 0.58
KT862222 Bovine faeces associated smacovirus 3 isolate 48 Fec5 cow
93 / 1.0
KM598409 Odonata associated circular virus 21 isolate OdasCV 21 US 1679SC3 12
100 / 1.0
LC386199 Chlorocebus cynosuros associated smacovirus CcSmV4-ZM09-95
100 / 1.0
LC386201 Papio cynocephalus associated smacovirus PcSmV6-ZM09-72
17 / -
76 / 0.94
LC386200 Chlorocebus cynosuros associated smacovirus CcSmV5-ZM09-83
42 / -
100 / 1.0
LC386202 Chlorocebus cynosuros associated smacovirus CcSmV6-ZM09-96
KT862228 Bovine faeces associated smacovirus 4 isolate GP3 45917 cow
KM573769 Dromedary stool associated circular ssDNA virus isolate DcSCV c1359
46 / 0.88
28 / 0.64
100 / 1.0
99 / 1.0
KT862224 Bovine faeces associated smacovirus 5 isolate 48 Fec9 cow
59 / 0.94
KM573774 Dromedary stool associated circular ssDNA virus isolate DcSCV c1330
LC386204 Papio cynocephalus associated smacovirus PcSmV2-ZM09-71
42 / 0.79
KT862223 Bovine faeces associated smacovirus 1 isolate 48 Fec59973 cow
KT862219 Sheep faeces associated smacovirus 3 isolate 47 Fec58091 Sheep
99 / 1.0
JX274036 Porcine stool associated circular virus isolate Cass
16 / -
50 / 0.85
KJ577811 Porcine stool associated circular virus 1 isolate DP3
KJ577810 Porcine stool associated circular virus 1 isolate DP2
50 / 0.85
100 / 1.0
LC386197 Papio cynocephalus associated smacovirus PcSmV3-ZM09-74
LC386198 Papio kindae associated smacovirus PkSmV3-ZM09-76
27 / 0.5
100 / 1.0
LC386205 Papio cynocephalus associated smacovirus PcSmV3-ZM09-51
KF880727 Turkey stool associated circular ssDNA virus strain TuSCV
KJ577817 Porcine stool associated circular virus 8 isolate GP2
24 / -
68 / 0.97
KJ577819 Porcine stool associated circular virus 6 isolate XP1
59 / 0.98
KP860908 Rat stool associated circular ssDNA virus isolate KS 11 0582
55 / 0.98
KP233194 Lemur smacovirus isolate SF5
64 / 0.98
KT600068 Human feces smacovirus 2 strain SmaCV2
100 / 1.0
LC386195 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-86
100 / 1.0
94 / 1.0
LC386196 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-96
GQ351272 Chimpanzee stool associated circular ssDNA virus isolate DP152
94 / 1.0
GQ351274 Chimpanzee stool associated circular ssDNA virus isolate GM476
KT862229 Bovine faeces associated smacovirus 6 isolate GP3 46075 cow
55 / 1.0
KM598410 Odonata associated circular virus 5 isolate OdasCV 5 US 1683LM1 12
0.2

Figure 2. Phylogenetic tree constructed based on the whole genome nucleotide sequences of Zambian non-human primate CRESS DNA viruses and reference smacoviruses. Tree topology following the maximum likelihood (ML) approach is shown and annotated with the support of the posterior probability from the Bayesian inference approach. Smacoviruses identified in the study are shown in red color. The sequence of PkSmV1-ZM09-64 was not included in the analysis due to the unisense genome organization of the ORFs (*Rep* and *CP*).

both CP and Rep trees, forming their own group (Figs 3 and 4) suggesting that they likely shared a common ancestor. PcSmV3-ZM09-51, PcSmV3-ZM09-74, and PkSmV3-ZM09-76 also consistently clustered together with porcine stool associated circular virus 1 (DP2, KJ577810) throughout the CP and Rep trees (Figs 3 and 4). In contrast, PkSmV1-ZM09-64 clustered together with CcSmV1-ZM09-86, CcSmV1-ZM09-96, human feces smacovirus 2 (SmaCV2, KT600068) and chimpanzee stool associated circular ssDNA virus (DP152, GQ351272) in the CP phylogeny (Fig. 3), whereas PkSmV1-ZM09-64 was located outside of the cluster formed by these smacoviruses in the Rep tree and was most closely related to a previously described bovine smacovirus (Fec59973, KT862223) and more distantly related to human and avian smacoviruses (Fig. 4). This phylogenetic incongruence revealed that the *CP* ORFs of PkSmV1-ZM09-64, CcSmV1-ZM09-86 and CcSmV1-ZM09-96 were derived from a common ancestor; however, the ancestor of the *Rep* gene of PkSmV1-ZM09-64 was different from that of CcSmV1-ZM09-86 and CcSmV1-ZM09-96. These findings suggested the possibility of recombination event(s) that may account for the discordant phylogenies. In addition, PcSmV2-ZM09-71 did not cluster with other ZM-SmVs throughout the constructed trees (Figs 2, 3 and 4). Taken together, these results indicated that all ZM-SmVs discovered in the study have distinct evolutionary histories and PkSmV1-ZM09-64 may have arisen from recombination.

Recombination analysis of smacovirus genomes. Detection of the phylogenetic incongruence between the CP and Rep phylogenies prompted us to investigate whether potential recombination sites existed in the ZM-SmV genomes. This recombination analysis revealed a region with multiple recombination breakpoints (i.e. a recombination hot spot) adjacent to the 3′-end of the *Rep* ORF (Fig. 5), which, interestingly, has also been inferred by another study[13]. These results corroborate prior studies indicating that smacoviruses increase their genetic diversity through recombination events. Two cold spots, where a recombination event is less likely to occur, were observed at the 5′-end of the *CP* ORF and the 3′-end of the *Rep* ORF. The presence of these cold spots implies functional conservation which is noteworthy in viruses as diverse as the ssDNA smacoviruses and suggests importance for these regions in the viral life cycle.

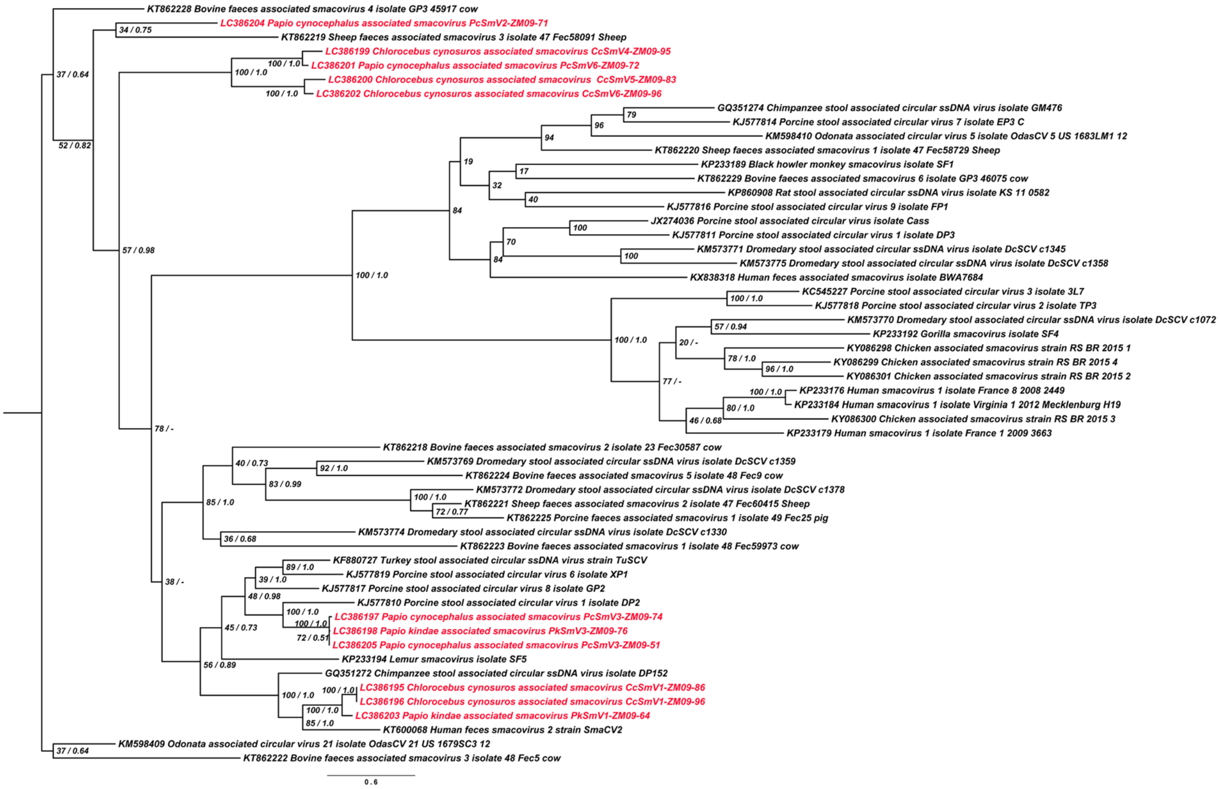
KT862228 Bovine faeces associated smacovirus 4 isolate GP3 45917 cow
LC386204 Papio cynocephalus associated smacovirus PcSmV2-ZM09-71
34 / 0.75
KT862219 Sheep faeces associated smacovirus 3 isolate 47 Fec58091 Sheep
LC386199 Chlorocebus cynosuros associated smacovirus CcSmV4-ZM09-95
100 / 1.0
LC386201 Papio cynocephalus associated smacovirus PcSmV6-ZM09-72
37 / 0.64
100 / 1.0
LC386200 Chlorocebus cynosuros associated smacovirus CcSmV5-ZM09-83
100 / 1.0
LC386202 Chlorocebus cynosuros associated smacovirus CcSmV6-ZM09-96
GQ351274 Chimpanzee stool associated circular ssDNA virus isolate GM476
79
KJ577814 Porcine stool associated circular virus 7 isolate EP3 C
96
KM598410 Odonata associated circular virus 5 isolate OdasCV 5 US 1683LM1 12
94
KT862220 Sheep faeces associated smacovirus 1 isolate 47 Fec58729 Sheep
52 / 0.82
19
KP233189 Black howler monkey smacovirus isolate SF1
17
KT862229 Bovine faeces associated smacovirus 6 isolate GP3 46075 cow
32
KP860908 Rat stool associated circular ssDNA virus isolate KS 11 0582
40
KJ577816 Porcine stool associated circular virus 9 isolate FP1
84
JX274036 Porcine stool associated circular virus isolate Cass
100
KJ577811 Porcine stool associated circular virus 1 isolate DP3
70
57 / 0.98
KM573771 Dromedary stool associated circular ssDNA virus isolate DcSCV c1345
100
84
KM573775 Dromedary stool associated circular ssDNA virus isolate DcSCV c1358
100 / 1.0
KX838318 Human feces associated smacovirus isolate BWA7684
KC545227 Porcine stool associated circular virus 3 isolate 3L7
100 / 1.0
KJ577818 Porcine stool associated circular virus 2 isolate TP3
KM573770 Dromedary stool associated circular ssDNA virus isolate DcSCV c1072
57 / 0.94
KP233192 Gorilla smacovirus isolate SF4
100 / 1.0
20 / -
KY086298 Chicken associated smacovirus strain RS BR 2015 1
78 / 1.0
KY086299 Chicken associated smacovirus strain RS BR 2015 4
96 / 1.0
KY086301 Chicken associated smacovirus strain RS BR 2015 2
77 / -
100 / 1.0
KP233176 Human smacovirus 1 isolate France 8 2008 2449
KP233184 Human smacovirus 1 isolate Virginia 1 2012 Mecklenburg H19
80 / 1.0
KY086300 Chicken associated smacovirus strain RS BR 2015 3
46 / 0.68
KP233179 Human smacovirus 1 isolate France 1 2009 3663
78 / -
KT862218 Bovine faeces associated smacovirus 2 isolate 23 Fec30587 cow
40 / 0.73
KM573769 Dromedary stool associated circular ssDNA virus isolate DcSCV c1359
92 / 1.0
KT862224 Bovine faeces associated smacovirus 5 isolate 48 Fec9 cow
83 / 0.99
KM573772 Dromedary stool associated circular ssDNA virus isolate DcSCV c1378
100 / 1.0
85 / 1.0
KT862221 Sheep faeces associated smacovirus 2 isolate 47 Fec60415 Sheep
72 / 0.77
KT862225 Porcine faeces associated smacovirus 1 isolate 49 Fec25 pig
KM573774 Dromedary stool associated circular ssDNA virus isolate DcSCV c1330
36 / 0.68
KT862223 Bovine faeces associated smacovirus 1 isolate 48 Fec59973 cow
KF880727 Turkey stool associated circular ssDNA virus strain TuSCV
89 / 1.0
KJ577819 Porcine stool associated circular virus 6 isolate XP1
39 / 1.0
38 / -
KJ577817 Porcine stool associated circular virus 8 isolate GP2
48 / 0.98
KJ577810 Porcine stool associated circular virus 1 isolate DP2
100 / 1.0
LC386197 Papio cynocephalus associated smacovirus PcSmV3-ZM09-74
100 / 1.0
LC386198 Papio kindae associated smacovirus PkSmV3-ZM09-76
45 / 0.73
72 / 0.51
LC386205 Papio cynocephalus associated smacovirus PcSmV3-ZM09-51
KP233194 Lemur smacovirus isolate SF5
56 / 0.89
GQ351272 Chimpanzee stool associated circular ssDNA virus isolate DP152
100 / 1.0
LC386195 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-86
100 / 1.0
LC386196 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-96
100 / 1.0
LC386203 Papio kindae associated smacovirus PkSmV1-ZM09-64
85 / 1.0
KT600068 Human feces smacovirus 2 strain SmaCV2
KM598409 Odonata associated circular virus 21 isolate OdasCV 21 US 1679SC3 12
37 / 0.64
KT862222 Bovine faeces associated smacovirus 3 isolate 48 Fec5 cow
0.6

Figure 3. Phylogenetic tree constructed based on the CP amino acid sequences of Zambian non-human primate CRESS DNA viruses and reference smacoviruses. Tree topology following the ML approach is shown and annotated with the support of the posterior probability from the Bayesian inference approach. Smacoviruses identified in this study are shown in red color.

Discussion

In the present study, sub-genomic fragments of CRESS DNA virus were initially detected in the faeces of NHPs in Zambia by metagenomic analysis. The complete genomes of eleven CRESS DNA viruses were then subsequently recovered by inverse PCR in 20% of faecal samples indicative of a high prevalence in Zambian NHPs. Although, a single CRESS DNA virus was found in nine individuals, two were found in a single Zambian malbrouck suggestive of a co-infection event (Table 1), a necessary precondition for viral recombination and emergence of new strains. Based on the genome-wide pairwise sequence identity analysis and degree of sequence divergence, these newly identified viruses could be tentatively classified as novel smacovirus species and await formal classification by the ICTV[6]. ZM-SmVs from both malbroucks and baboons formed distinct clusters within the genus *Porprismacovirus* on the phylogenetic trees, suggesting that they evolved from different ancestral progenitors further exemplifying the extent of the viral diversity.

Despite the high prevalence in Zambian NHPs, the detected ZM-SmVs showed phylogenetic divergence and there was no evidence for spread of specific ZM-SmV strains in the species of monkey and baboon NHPs we studied, which raises the question whether ZM-SmVs infect and transmit among NHPs and potentially also other species. Detection of these viruses in the NHP faecal matter suggests at least two distinct hypotheses with respect to their origin. First is that they might productively infect the NHPs; however, smacoviruses have not been identified in animal tissues[6]. Second is that they may represent ssDNA viruses ultimately derived from plants, insects or mammals, which comprise the NHP diet or from a resident microorganism of the NHP gut[11,13,20,21]. Indeed, a recent study has described high sequence similarities between smacoviral genomes and spacer sequences of a faecal archaeon, *Candidatus* Methanomassiliicoccus intestinalis, indicating a tropism of smacoviruses for archaea[22]. To date, and in common with a growing number of ssDNA and other uncultured viruses, isolates of infectious smacoviruses have not been reported. Taken together, the precise origins of the ZM-SmVs reported here remain to be established and further studies including attempts at isolation of smacoviruses are needed to characterize smacovirus infection in detail.

There was no clear congruence between the CP and Rep phylogenies for the identified CRESS DNA viruses. Specifically, PkSmV1-ZM09-64 showed clearly different phylogenetic relationships in both the CP and Rep trees. We also detected a potential recombination hot spot of breakpoints in the genome of smacovirus at the 3′-end of the *Rep* ORF providing further evidence of the importance of recombination events during the evolution of smacoviruses[13,23,24]. A recent study has also reported that the Rep of these viruses is chimeric and likely derives from recombination events that lead to intra-host lineage diversification[24]. Interestingly, the recombination analysis showed the breakpoint hot spot extended into the intergenic region between the *CP* and *Rep* ORFs. This observation has been seen in diverse ssRNA[25] and dsDNA viruses[26] and supposed the existence of "functionally

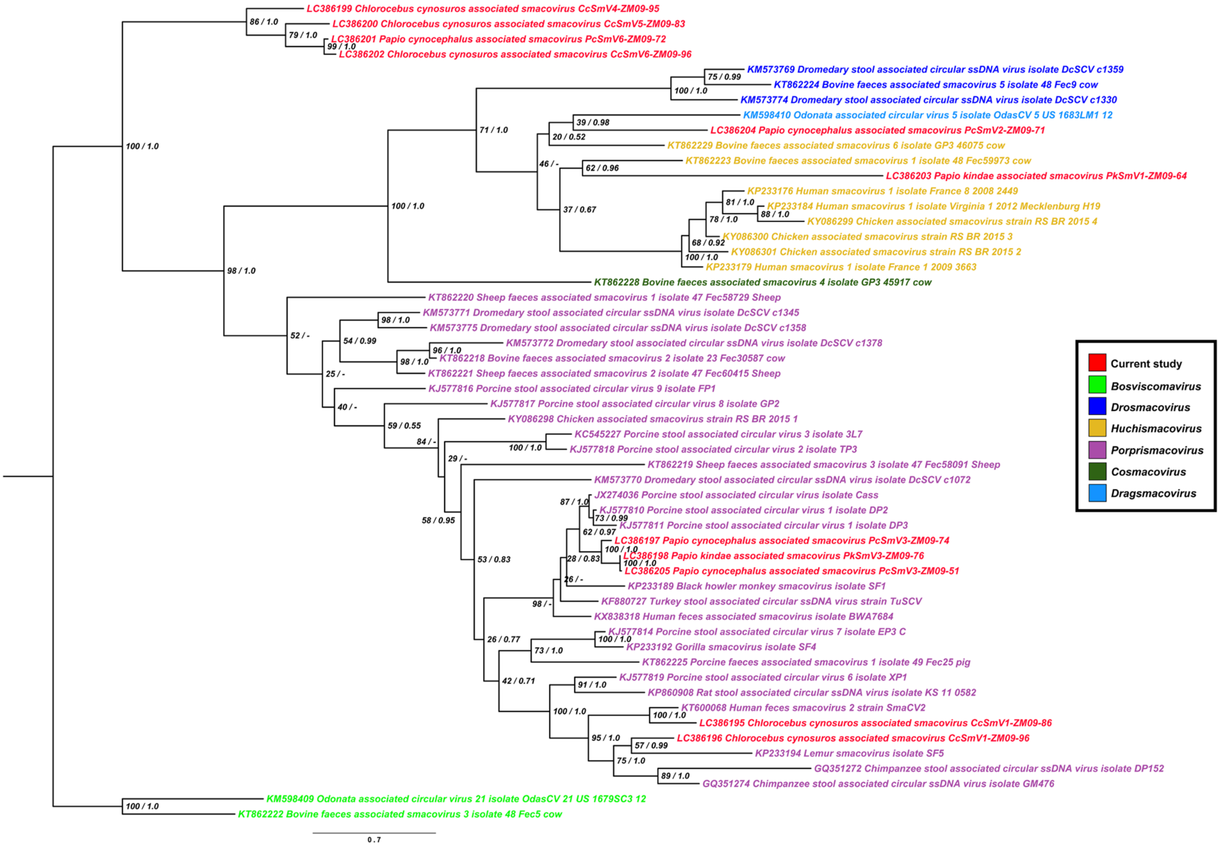
LC386199 Chlorocebus cynosuros associated smacovirus CcSmV4-ZM09-95
86 / 1.0
LC386200 Chlorocebus cynosuros associated smacovirus CcSmV5-ZM09-83
79 / 1.0
LC386201 Papio cynocephalus associated smacovirus PcSmV6-ZM09-72
99 / 1.0
LC386202 Chlorocebus cynosuros associated smacovirus CcSmV6-ZM09-96
KM573769 Dromedary stool associated circular ssDNA virus isolate DcSCV c1359
75 / 0.99
KT862224 Bovine faeces associated smacovirus 5 isolate 48 Fec9 cow
100 / 1.0
KM573774 Dromedary stool associated circular ssDNA virus isolate DcSCV c1330
KM598410 Odonata associated circular virus 5 isolate OdasCV 5 US 1683LM1 12
39 / 0.98
LC386204 Papio cynocephalus associated smacovirus PcSmV2-ZM09-71
71 / 1.0
20 / 0.52
KT862229 Bovine faeces associated smacovirus 6 isolate GP3 46075 cow
100 / 1.0
KT862223 Bovine faeces associated smacovirus 1 isolate 48 Fec59973 cow
46 / -
62 / 0.96
LC386203 Papio kindae associated smacovirus PkSmV1-ZM09-64
KP233176 Human smacovirus 1 isolate France 8 2008 2449
81 / 1.0
KP233184 Human smacovirus 1 isolate Virginia 1 2012 Mecklenburg H19
100 / 1.0
37 / 0.67
88 / 1.0
78 / 1.0
KY086299 Chicken associated smacovirus strain RS BR 2015 4
KY086300 Chicken associated smacovirus strain RS BR 2015 3
68 / 0.92
KY086301 Chicken associated smacovirus strain RS BR 2015 2
100 / 1.0
KP233179 Human smacovirus 1 isolate France 1 2009 3663
98 / 1.0
KT862228 Bovine faeces associated smacovirus 4 isolate GP3 45917 cow
KT862220 Sheep faeces associated smacovirus 1 isolate 47 Fec58729 Sheep
KM573771 Dromedary stool associated circular ssDNA virus isolate DcSCV c1345
98 / 1.0
KM573775 Dromedary stool associated circular ssDNA virus isolate DcSCV c1358
52 / -
54 / 0.99
KM573772 Dromedary stool associated circular ssDNA virus isolate DcSCV c1378
96 / 1.0
98 / 1.0
KT862218 Bovine faeces associated smacovirus 2 isolate 23 Fec30587 cow
KT862221 Sheep faeces associated smacovirus 2 isolate 47 Fec60415 Sheep
25 / -
KJ577816 Porcine stool associated circular virus 9 isolate FP1
40 / -
KJ577817 Porcine stool associated circular virus 8 isolate GP2
KY086298 Chicken associated smacovirus strain RS BR 2015 1
59 / 0.55
KC545227 Porcine stool associated circular virus 3 isolate 3L7
84 / -
100 / 1.0
KJ577818 Porcine stool associated circular virus 2 isolate TP3
29 / -
KT862219 Sheep faeces associated smacovirus 3 isolate 47 Fec58091 Sheep
KM573770 Dromedary stool associated circular ssDNA virus isolate DcSCV c1072
JX274036 Porcine stool associated circular virus isolate Cass
87 / 1.0
KJ577810 Porcine stool associated circular virus 1 isolate DP2
58 / 0.95
73 / 0.99
KJ577811 Porcine stool associated circular virus 1 isolate DP3
62 / 0.97
LC386197 Papio cynocephalus associated smacovirus PcSmV3-ZM09-74
100 / 1.0
LC386198 Papio kindae associated smacovirus PkSmV3-ZM09-76
53 / 0.83
28 / 0.83
100 / 1.0
LC386205 Papio cynocephalus associated smacovirus PcSmV3-ZM09-51
26 / -
KP233189 Black howler monkey smacovirus isolate SF1
KF880727 Turkey stool associated circular ssDNA virus strain TuSCV
98 / -
KX838318 Human feces associated smacovirus isolate BWA7684
KJ577814 Porcine stool associated circular virus 7 isolate EP3 C
26 / 0.77
100 / 1.0
KP233192 Gorilla smacovirus isolate SF4
73 / 1.0
KT862225 Porcine faeces associated smacovirus 1 isolate 49 Fec25 pig
KJ577819 Porcine stool associated circular virus 6 isolate XP1
42 / 0.71
91 / 1.0
KP860908 Rat stool associated circular ssDNA virus isolate KS 11 0582
KT600068 Human feces smacovirus 2 strain SmaCV2
100 / 1.0
100 / 1.0
LC386195 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-86
95 / 1.0
LC386196 Chlorocebus cynosuros associated smacovirus CcSmV1-ZM09-96
57 / 0.99
KP233194 Lemur smacovirus isolate SF5
75 / 1.0
GQ351272 Chimpanzee stool associated circular ssDNA virus isolate DP152
89 / 1.0
GQ351274 Chimpanzee stool associated circular ssDNA virus isolate GM476
KM598409 Odonata associated circular virus 21 isolate OdasCV 21 US 1679SC3 12
100 / 1.0
KT862222 Bovine faeces associated smacovirus 3 isolate 48 Fec5 cow
0.7
🟥 Current study
🟩 Bosviscomavirus
🟦 Drosmacovirus
🟨 Huchismacovirus
🟪 Porprismacovirus
🟩 Cosmacovirus
🟦 Dragsmacovirus

Figure 4. Phylogenetic tree constructed based on the Rep amino acid sequences of Zambian non-human primate CRESS DNA viruses and other known smacoviruses. Tree topology following the ML approach is shown and annotated with the support of the posterior probability from the Bayesian inference approach. Smacoviruses identified in the study are shown in red color.

interchangeable modules", i.e. shuffling of the *CP* ORF by recombination may conceivably impact on virus tropism of recombinants. Our results are in agreement with the notion of recombination patterns including a mechanistic predisposition to recombination in virion-strand replication origin and recombination breakpoints which significantly tend to occur in intergenic regions or at 5′ and 3′ termini of genes rather than within the genes of ssDNA viruses[23,27]. Recombination breakpoints are known to be disfavoured within coding regions, as observed in the CP. Therefore, genes in ssDNA viruses preferentially move as modules which contain >50% of the coding region and natural selection disfavours viruses harbouring recombinant proteins which leads to the observed nonrandom distribution of breakpoint observed[27]. The modular genetic exchange by recombination within non-coding regions have also been previously implicated in the emergence of new viral strains[28,29]. Whether these, or related phenomena, exist for recombinant smarcoviruses warrants further study.

PkSmV1-ZM09-64 showed a unisense genome organization of *CP* and *Rep* ORF similar to previously reported CRESS DNA virus, PigSCV (JQ023166) (Fig. 1)[8]. The precise reasons underlying this ORF organization by these CRESS DNA viruses remain unclear. It is possible that this genome organization may have arisen from errors during recombination event between ancestral viruses which led to a unidirectional ORF organization instead of the more common ambisense bidirectional genome organization evident in the majority of known smacoviruses.

In conclusion, our studies indicate the presence of previously unrecognized CRESS DNA viruses in the NHP virome and provide further evidence of the extent of the genetic diversity of DNA viruses in primates.

Materials and Methods

Ethical statement and sample collection.

All animal experiments were approved by the then Zambia Wildlife Authority (ZAWA), now the Department of National Parks and Wildlife, Ministry of Tourism and Arts and performed in accordance with the relevant guidelines and regulations (certificate no. 2604). Tissue and faecal samples were collected from NHPs (*Chlorocebus cynosuros*, n = 25; *Papio spp*., n = 25) in Mfuwe District in 2009 and used for different research projects as reported previously[30,31]. For NHP species typing, the mitochondrial *cytb* gene was amplified and sequenced from genomic DNA extracted from spleen tissues of the NHPs, as described elsewhere[31].

Metagenomic analysis using high-throughput sequencing.

Viral nucleic acids were extracted from the pooled faecal suspensions as described previously[32], and double-stranded cDNA was synthesized with the

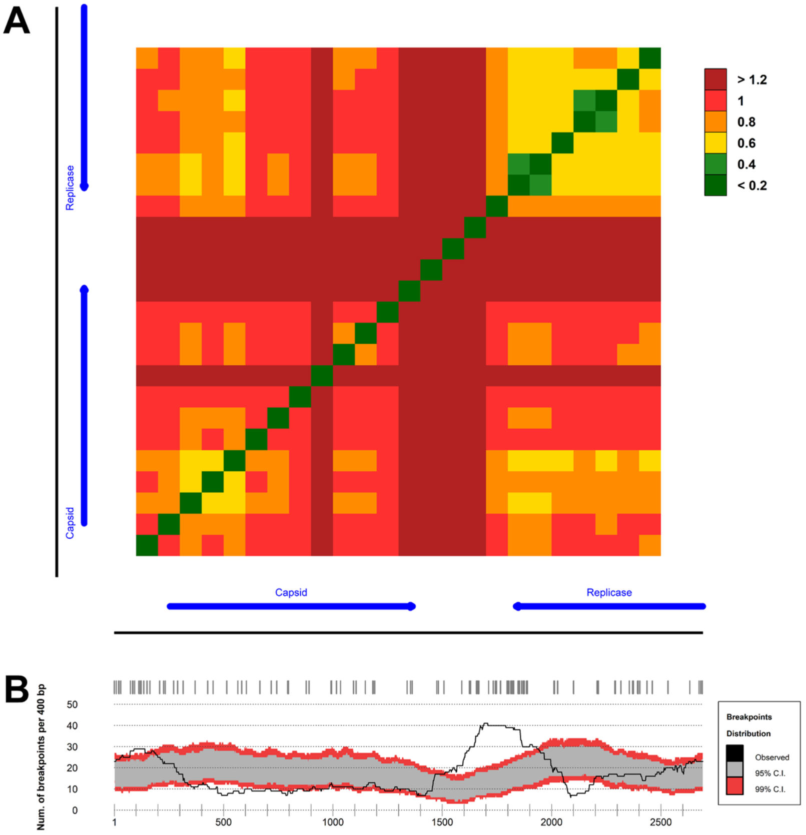

Figure 5. Recombination analysis of smacovirus species. (**A**) Phylogenetic compatibility matrix over the species of smacoviruses including ZM-SmVs. The respective regions in the smacoviruses alignment correspond to the normalized Robinson-Foulds distance between two neighbor-joining (NJ) trees. (**B**) Distribution of recombination breakpoints within the smacoviruses alignment. Breakpoints detected by recombination detection program (RDP) analysis are shown at top of the panel. The vertical axis is the number of recombination boundaries per window. The solid black line indicates the number of observed breakpoints. The grey and red areas indicate 95 and 99% confidence intervals of each breakpoint, respectively.

PrimeScript Double Strand cDNA Synthesis kit (Takara BIO, Shiga, Japan). Sequencing libraries were prepared with Nextera XT DNA Sample Preparation kit (Illumina, San Diego, CA) and sequenced on the Illumina MiSeq platform (Illumina). The obtained reads were compared against NCBI NT/NR database as described previously[33]. The sequence reads related to smacoviruses were *de novo* assembled to contigs using CLC Genomics Workbench software (CLC bio, Aarhus, Denmark).

PCR screening, whole genome sequencing, and genome annotation. Based on the nucleotide sequences of the generated contigs, six different pairs of primer sets were designed for PCR screening with GENETYX software (GENETYX, Tokyo, Japan) (Supplementary Table S1). DNA was extracted from faecal samples for each individual NHP with the High Pure Viral Nucleic Acid kit (Roche Diagnostics, Mannheim, Germany) and PCR screened for putative smacoviruses with Tks Gflex DNA Polymerase (TAKARA BIO). PCR products were sequenced and the sequences were used to design additional primers for the complete genome amplification of CRESS DNA viruses by inverse PCR. The amplicons were subsequently cloned into a pCR4-Blunt-TOPO vector (Invitrogen; Thermo Fischer Scientific, Waltham, MA) and sequenced by a primer walking strategy. The whole circular genome of each CRESS DNA virus was assembled with Phred and Phrap[34] with quality scores >30 in all assembled nucleotide positions and annotated using Geneious[35]. The pairwise identity among sequences was calculated with Sequence Demarcation Tool (SDT) v1.2[36].

Phylogenetic analysis. The complete genome nucleotide sequences and predicted amino acid protein sequences were aligned with MAFFT using the algorithm FFT-NS-i[37]. To infer the phylogenetic relation between sequenced and available samples maximum likelihood (ML) approaches with IQ-TREE v1.6.5[38] were used to determine the best substitution model, infer the topology and the branch support with a bootstrap of 1,000 repetitions. Additionally, Bayesian inference approaches with MrBayes v3.2.6[39] were used to search for the best substitution model and estimate the posterior probability of the inferred branches with chains of one million states. Three phylogenetic trees were inferred in this study for the whole genome nucleotide sequences (Fig. 2), and the amino acid sequences of CP (Fig. 3) and Rep (Fig. 4). The tree topology of the ML approach was used and annotated with the support of the posterior probability from the Bayesian approach.

Recombination analysis. The genome multiple sequence alignment was assessed for evidence of recombination events by the suite of methods in the recombination detection program (RDP) v.4.58[40,41]. Detected recombination events required statistical support $p < 0.01$ and the distribution of recombination breakpoints were analyzed with a sliding window of 400 nucleotides and one nucleotide step, with 1,000 permutations for estimating the statistical support of the breakpoint distribution. To assess the effects of the recombination events on the phylogenetic relationships among sequences, a compatibility matrix was built[25], where the compatibility of two windows with 300 nucleotides from a sliding window and 100 nucleotides per step is defined as the normalized Robinson-Foulds distance[42] between the corresponding neighbor-joining phylogenetic trees under Tamura-Nei substitution model. The compatibility reflects how similar are the inferred phylogenies for any two genome windows ranging between 0 (identical topologies) to 1 (completely dissimilar topologies).

Data Availability

The whole genomes of the identified viruses in this study were submitted to the GenBank/EMBL/DDBJ database under accession numbers of LC386195-LC386205.

References

1. Simmonds, P. *et al*. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol* **15**, 161–168, https://doi.org/10.1038/nrmicro.2016.177 (2017).
2. Zhao, L., Rosario, K., Breitbart, M. & Duffy, S. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses: Ubiquitous Viruses With Small Genomes and a Diverse Host Range. *Adv Virus Res* **103**, 71–133, https://doi.org/10.1016/bs.aivir.2018.10.001 (2019).
3. King, A. M. Q. *et al*. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Arch Virol* **163**, 2601–2631, https://doi.org/10.1007/s00705-018-3847-1 (2018).
4. Rosario, K., Duffy, S. & Breitbart, M. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* **157**, 1851–1871, https://doi.org/10.1007/s00705-012-1391-y (2012).
5. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267–276, https://doi.org/10.1038/nrg2323 (2008).
6. Varsani, A. & Krupovic, M. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch Virol* **163**, 2005–2015, https://doi.org/10.1007/s00705-018-3820-z (2018).
7. Steel, O. *et al*. Circular replication-associated protein encoding DNA viruses identified in the faecal matter of various animals in New Zealand. *Infect Genet Evol* **43**, 151–164, https://doi.org/10.1016/j.meegid.2016.05.008 (2016).
8. Cheung, A. K. *et al*. A divergent clade of circular single-stranded DNA viruses from pig feces. *Arch Virol* **158**, 2157–2162, https://doi.org/10.1007/s00705-013-1701-z (2013).
9. Cheung, A. K. *et al*. Identification of several clades of novel single-stranded circular DNA viruses with conserved stem-loop structures in pig feces. *Arch Virol* **160**, 353–358, https://doi.org/10.1007/s00705-014-2234-9 (2015).
10. Sachsenröder, J. *et al*. Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group A rotavirus. *J Gen Virol* **95**, 2734–2747, https://doi.org/10.1099/vir.0.070029-0 (2014).
11. Lima, D. A. *et al*. Faecal virome of healthy chickens reveals a large diversity of the eukaryote viral community, including novel circular ssDNA viruses. *J Gen Virol* **98**, 690–703, https://doi.org/10.1099/jgv.0.000711 (2017).
12. Woo, P. C. *et al*. Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology* **471–473**, 117–125, https://doi.org/10.1016/j.virol.2014.09.020 (2014).
13. Ng, T. F. *et al*. A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evol* **1**, vev017, https://doi.org/10.1093/ve/vev017 (2015).
14. Blinkova, O. *et al*. Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* **91**, 74–86, https://doi.org/10.1099/vir.0.015446-0 (2010).
15. Phan, T. G. *et al*. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98–104, https://doi.org/10.1016/j.virol.2015.03.011 (2015).
16. Dayaram, A. *et al*. Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect Genet Evol* **30**, 278–287, https://doi.org/10.1016/j.meegid.2014.12.037 (2015).
17. Rosario, K. *et al*. Virus discovery in all three major lineages of terrestrial arthropods highlights the diversity of single-stranded DNA viruses associated with invertebrates. *PeerJ* **6**, e5761, https://doi.org/10.7717/peerj.5761 (2018).
18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2 (1990).
19. Sachsenröder, J. *et al*. Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One* **7**, e34631, https://doi.org/10.1371/journal.pone.0034631 (2012).
20. Kim, H. K. *et al*. Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* **93**, 635–639, https://doi.org/10.1099/vir.0.037838-0 (2012).
21. Phan, T. G. *et al*. The fecal virome of South and Central American children with diarrhea includes small circular DNA viral genomes of unknown origin. *Arch Virol* **161**, 959–966, https://doi.org/10.1007/s00705-016-2756-4 (2016).
22. Díez-Villaseñor, C. & Rodriguez-Valera, F. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat Commun* **10**, 294, https://doi.org/10.1038/s41467-018-08167-w (2019).
23. Martin, D. P. *et al*. Recombination in eukaryotic single stranded DNA viruses. *Viruses* **3**, 1699–1738, https://doi.org/10.3390/v3091699 (2011).
24. Kazlauskas, D., Varsani, A. & Krupovic, M. Pervasive Chimerism in the Replication-Associated Proteins of Uncultured Single-Stranded DNA Viruses. *Viruses* **10**, 187, https://doi.org/10.3390/v10040187 (2018).
25. Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J Virol* **80**, 11827–11832, https://doi.org/10.1128/JVI.01100-06 (2006).
26. Carr, M. *et al*. Discovery of African bat polyomaviruses and infrequent recombination in the large T antigen in the Polyomaviridae. *J Gen Virol* **98**, 726–738, https://doi.org/10.1099/jgv.0.000737 (2017).
27. Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* **83**, 2697–2707, https://doi.org/10.1128/JVI.02152-08 (2009).
28. Gonzalez, G., Koyanagi, K. O., Aoki, K. & Watanabe, H. Interregional Coevolution Analysis Revealing Functional and Structural Interrelatedness between Different Genomic Regions in Human Mastadenovirus D. *J Virol* **89**, 6209–6217, https://doi.org/10.1128/JVI.00515-15 (2015).
29. Muslin, C., Joffret, M. L., Pelletier, I., Blondel, B. & Delpeyroux, F. Evolution and Emergence of Enteroviruses through Intra- and Inter-species Recombination: Plasticity and Phenotypic Impact of Modular Genetic Exchanges in the 5′ Untranslated Region. *PLoS Pathog* **11**, e1005266, https://doi.org/10.1371/journal.ppat.1005266 (2015).

30. Sasaki, M. *et al*. Distinct Lineages of Bufavirus in Wild Shrews and Nonhuman Primates. *Emerg Infect Dis* **21**, 1230–1233, https://doi.org/10.3201/eid2107.141969 (2015).
31. Carr, M. *et al*. Isolation of a simian immunodeficiency virus from a malbrouck (Chlorocebus cynosuros). *Arch Virol* **162**, 543–548, https://doi.org/10.1007/s00705-016-3129-8 (2017).
32. Sasaki, M. *et al*. Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *J Gen Virol* **96**, 440–452, https://doi.org/10.1099/vir.0.071209-0 (2015).
33. Gonzalez, G. *et al*. An optimistic protein assembly from sequence reads salvaged an uncharacterized segment of mouse picobirnavirus. *Sci Rep* **7**, 40447, https://doi.org/10.1038/srep40447 (2017).
34. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. *Error probabilities. Genome Research* **8**, 186–194, https://doi.org/10.1101/gr.8.3.186 (1998).
35. Kearse, M. *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649, https://doi.org/10.1093/bioinformatics/bts199 (2012).
36. Muhire, B. M., Varsani, A. & Martin, D. P. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* **9**, e108277, https://doi.org/10.1371/journal.pone.0108277 (2014).
37. Katoh, K. & Standley, D. M. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933–1942, https://doi.org/10.1093/bioinformatics/btw108 (2016).
38. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**, 268–274, https://doi.org/10.1093/molbev/msu300 (2015).
39. Ronquist, F. *et al*. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* **61**, 539–542, https://doi.org/10.1093/sysbio/sys029 (2012).
40. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1**, vev003, https://doi.org/10.1093/ve/vev003 (2015).
41. Martin, D. P., Murrell, B., Khoosal, A. & Muhire, B. Detecting and Analyzing Genetic Recombination Using RDP4. *Methods Mol Biol* **1525**, 433–460, https://doi.org/10.1007/978-1-4939-6622-6_17 (2017).
42. Steel, M. A. & Penny, D. Distributions of tree comparison metrics—some new results. *Syst Biol* **42**, 126–141 (1993).

Acknowledgements

We thank the Department of National Parks and Wildlife of the Ministry of Tourism and Arts (formerly ZAWA), Government of the Republic of Zambia, for assistance with sample collection in Zambia. This work was supported by the Japan Initiative for Global Research Network of Infectious Diseases (J-GRID) from Japan Agency for Medical Research and Development (AMED) (JP18fm0108008); Grants-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (16H06429, 16H06431, 16K21723); and Japan Society for the Promotion of Science (JSPS) KAKENHI (16H05805).

Author Contributions

M.S. and H.S. conceived the research, P.D.A., M.S., G.G. and W.P. conducted the experiments and analyzed the data. M.S., B.M.H., A.S.M., K.I., Y.O. and H.S. contributed samples/reagents/analysis tools. P.D.A., M.S., G.G., M.C., A.S.M. and H.S. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41358-z.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.