1 **The Smc5/6 complex belongs to a novel class of evolutionary-constrained restriction**
2 **factors and is antagonized by all mammalian hepatitis B virus X proteins**
3
4 Fabien Filleton[1-6]*, Fabien Abdul[7]*, Laetitia Gerossier[8], Alexia Paturel[8], Janet Hall[8], Michel
5 Strubin[7], Lucie Etienne[1-6]§
6
7 [1] CIRI, Centre International de Recherche en Infectiologie, Lyon, France ; [2] INSERM,
8 U1111 ; [3] Ecole Normale Supérieure ENS de Lyon ; [4] CNRS, UMR5308 ; [5] University of
9 Lyon, Lyon I, UMS3444/US8 BioSciences Gerland ; [6] Labex ANR ECOFECT of Université
10 de Lyon ; [7] Department of Microbiology and Molecular Medicine, University Medical Centre
11 (C.M.U.) / University of Geneva, Rue Michel-Servet 1, 1211 Geneva 4, Switzerland ; [8]
12 CRCL- UMR Inserm 1052 - CNRS 5286, 151 cours Albert Thomas, 69300 Lyon, France
13
14 * Contributed equally to this work
15 §: Correspondence: Lucie Etienne, CIRI, ENS-Lyon, 46 Allée d'Italie, 69364 Lyon, France.
16 E-mail: lucie.etienne@ens-lyon.fr
17
18 **Keywords:** Smc5/6 complex, evolution of SMC genes, Hepatitis B virus, HBx, restriction
19 factor, antagonist, positive selection, virus-host interaction, genetic conflict

Abstract

The Smc5/6 complex, in addition to its essential housekeeping functions, is a restriction factor against human hepatitis B virus (HBV) and is targeted for degradation by the regulatory HBx protein. Here, we examined its evolutionary history and the conservation of this HBx-mediated degradation by mammalian hepadnaviruses. Phylogenetic and positive selection analyses found that the six Smc5/6 subunits have been highly conserved in primates with only Smc6 showing signatures of potential virus-host "arms-race". Despite substantial sequence divergence, HBx proteins from hepatitis viruses naturally infecting bats, rodents, or primates efficiently degraded Smc5/6 in various species cells and rescued the replication of an HBx-deficient HBV in primary human hepatocytes. These findings point to an evolutionary-conserved antagonism of Smc5/6 by HBx. They further suggest that Smc5/6, like SERINCs, belongs a novel class of restriction factors that have dual antiviral and essential cellular functions and have therefore limited evolutionary opportunities to escape from viral antagonists.

36 **Introduction**

37 Infection with hepatitis B virus (HBV) is a leading cause of chronic hepatitis and liver
38 cancer in humans (World Health Organization). HBV is a member of the *Hepadnaviridae*
39 family of DNA viruses, which have co-evolved with their host species for millions of years
40 (Dill et al., 2016; Lauber et al., 2017; Suh et al., 2014). Today, hepadnaviruses are found to
41 naturally infect, in a species-specific manner, mammals, birds, as well as fish and amphibians.
42 In mammals, HBVs are present in rodents, bats, and several primates including human,
43 chimpanzee, gibbon, orangutan, and the New World wooly monkey. Mammalian
44 hepadnaviruses (orthohepadnaviruses) all contain a gene encoding a small regulatory protein,
45 HBx, that is thought to have arisen *de novo* in the orthohepadnavirus lineage (Suh et al.,
46 2014). HBx has long been known to play a central role in HBV replication and pathogenesis
47 (Benhenda, Cougot, Buendia, & Neuveut, 2009; Feitelson, Bonamassa, & Arzumanyan, 2014;
48 Lucifora et al., 2011) and has recently been shown to have a key role in promoting HBV
49 transcription by antagonizing the restriction function of the infected cell's Structural
50 Maintenance of Chromosome (SMC) Smc5/6 complex (Decorsiere et al., 2016; Murphy et al.,
51 2016). However, whether this property has been conserved among the X-containing
52 hepadnaviruses is unknown.

53 The Smc5/6 complex is, together with cohesin and condensin, one of the three SMC
54 complexes found in eukaryotes (Cobbe & Heck, 2004; Uhlmann, 2016). These complexes all
55 have essential housekeeping functions, playing fundamental roles in chromosome replication,
56 segregation and repair (reviewed in (Uhlmann, 2016). Condensin controls chromosome
57 condensation during mitosis and cohesin maintains cohesion between the newly replicated
58 sister chromatids. However, the role of the Smc5/6 complex is less well understood. It has
59 reported functions in DNA replication and repair, but its exact mode of action remains elusive

60 (Fernandez-Capetillo, 2016; Hwang et al., 2017; Kegel & Sjogren, 2010; Menolfi, Delamarre,
61 Lengronne, Pasero, & Branzei, 2015; Watanabe, 2005).

62 In addition to its essential cellular activities, a novel function of Smc5/6 complex as an
63 HBV restriction factor has been recently dissected: in the absence of HBx, the Smc5/6
64 complex binds to the HBV episomal DNA genome and inhibits viral transcription (Decorsiere
65 et al., 2016; Murphy et al., 2016; Niu et al., 2017). HBx antagonizes this effect by high-
66 jacking the host DDB1-containing E3 ubiquitin ligase complex to target the Smc5/6 complex
67 for ubiquitin-mediated degradation, thereby enabling productive HBV gene expression
68 (Decorsiere et al., 2016).

69 Most genes encoding for antiviral restriction factors have been engaged in an
70 "evolutionary arms-race" with the viruses they inhibit (Duggal & Emerman, 2012; Johnson,
71 2013). Indeed, during long-term co-evolution, pathogenic viruses and their hosts are
72 constantly under the selective pressure of the other for survival. As a result, host restriction
73 factors evolve rapidly and display signatures of positive (diversifying) selection. These
74 signatures can be identified by analyzing the codon sequences of orthologous genes from a
75 large number of related species. At virus-host interaction sites, one can witness adaptive
76 changes, including frequent amino acid changes (where a higher non-synonymous
77 substitution rate, dN, than the synonymous rate, dS, is indicative of positive selection), and
78 insertion/deletions (indels) or splicing variants as ways to modify the virus-host interface and
79 to escape from viral antagonists (Daugherty & Malik, 2012; Jia et al., 2009; Johnson, 2013;
80 Lim, Malik, & Emerman, 2010; Sauter et al., 2009; Sawyer, Wu, Emerman, & Malik, 2005).

81 To assess whether the Smc5/6 complex underwent such a selective pressure we
82 determined its evolutionary history in primates. As for the other SMC complexes, the core of
83 the Smc5/6 complex is formed by a heterodimer of two SMC proteins, Smc5 and Smc6
84 (Gligoris & Lowe, 2016), which associate with four additional subunits known as non-SMC

85 elements (Nsmce1-4) (Figure 1A). We found that all six subunits have been strictly conserved
86 with just Smc6 showing signatures of potential virus-host evolutionary "arms-race". To
87 investigate whether the differences in Smc6 impact on the ability of HBx to antagonize the
88 Smc5/6 complex, we performed evolutionary-guided functional assays. These revealed that
89 the HBx protein from six divergent orthohepadnaviruses could all efficiently degrade Smc5/6
90 in a species-independent manner and rescue replication of an HBV mutant lacking a
91 functional HBx protein in human cells. This suggests that Smc5/6 degradation is a strict
92 requirement for the establishment of mammalian HBV infection. From an evolutionary
93 perspective, the absence of species-specificity indicates that mammalian hepadnaviruses have
94 not shaped Smc6 adaptive evolution and that the Smc5/6 complex represents a novel class of
95 evolutionary-constrained restriction factors. It also raises the intriguing possibility that the
96 Smc5/6 complex may restrict other, yet unidentified pathogenic viruses.
97
98 **Results**
99 **Overall evolutionary conservation of the Smc5/6 complex in primates**
100 To trace the evolutionary history of the Smc5/6 complex, we compared the sequence
101 of its six subunits in primates (Figure 1A). As a comparison, we also analyzed the
102 evolutionary history of all the other primate SMC genes. These include the *Smc1* and *Smc3*
103 genes, which encode the core cohesin subunits, and the *Smc2* and *Smc4* genes encoding the
104 condensin core subunits. The sequences of these genes were retrieved from publically
105 available datasets (Table 1, Dataset S1). To perform more robust phylogenetic and selection
106 analyses, we obtained additional primate species sequences using RT-PCR approaches (Table
107 1, Figure 1B; see Methods). Overall, we included up to 20 simian primate species in our
108 positive selection analyses to span 40 million years of evolution (McBee, Rozmiarek,
109 Meyerson, Rowley, & Sawyer, 2015; Perelman et al., 2011). We found that the synteny of the

genes was conserved during primate evolution, although some subunits had duplicated pseudogenes in a few primate species (Figure S1). Amongst the core SMC proteins, the most conserved are the cohesin Smc1 and Smc3 subunits, which share 100% and 99.9% pairwise identity at the amino acid level in a simian primate alignment, respectively (Dataset S1). Smc6 was the least conserved SMC protein with 97.4% pairwise identity in simian primates. Using the Genetic Algorithm for Recombination Detection (GARD) on the complete set of genes (Kosakovsky Pond, Posada, Gravenor, Woelk, & Frost, 2006), evidence of recombination (GARD, $p<0.05$) was only found for *Nsmce3*, and therefore subsequent phylogenetic and selection analyses were performed on both the whole *Nsmce3* gene and the two *Nsmce3* gene fragments identified (from 1-246 bp and 247-912 bp). Phylogenetic analysis of the ten genes showed that the gene trees derived from nucleotide alignments were largely in accordance with the accepted species tree from Perelman and colleagues (Figure 1B, and Dataset S1) (Perelman et al., 2011).

To assess whether the Smc5/6 complex and the *Smc1-4* genes have experienced diversifying selection during primate evolution, we performed four types of positive selection analyses. Firstly, we used the BUSTED method that tests whether a gene has experienced positive selection on at least one site or one branch during evolution (Murrell et al., 2015). Of the six Smc5/6 complex subunits and the four *SMC* genes from the cohesin and condensin complexes, only one, *Smc6*, showed gene-wide evidence of episodic positive selection (BUSTED, $p<0.05$, Figure 1C). These findings were confirmed using the PARRIS method that also detects evidence of positive selection using a codon alignment (Scheffler, Martin, & Seoighe, 2006), although the level of significance was not reached for the *Smc6* gene ($p=0.19$; Figure 1C). Thirdly, we ran the Codeml program from the PAML package (Yang, 2007) on our codon alignments to compare two models: one that allows positive selection at certain sites (M8, alternative hypothesis) and one that disallows positive selection (M7, null

135 hypothesis). We then performed a likelihood ratio test (LRT) to examine which of the two
136 models fitted better our data. Overall, there was no convergence for seven of the X genes we
137 tested because they were too conserved (see Methods for details) and no evidence of
138 significant positive selection was found for the remaining genes (Figure 1C). Finally, we used
139 the Bio++ package from Guéguen *et al.* (Gueguen et al., 2013) to similarly test for evidence
140 of positive selection (M7 *vs.* M8 as implemented in Bio++), but with two main advantages
141 over PAML. First, the DNA substitution models use non-stationary matrices, which allow
142 nucleotide composition to change over time, and therefore improve the fitting to real data.
143 Second, the codon frequency better fits biological assumptions, as compared to PAML
144 (Gueguen et al., 2013). Using Bio++, we obtained higher values of likelihood and, for *Smc6*, a
145 model allowing positive selection was favored over a model that disallows positive selection
146 (Bio++ M7 *vs.* M8, p=0.05; Figure 1C). Overall, these studies show that the *SMC* genes for
147 cohesin and condensin, as well as the six Smc5/6 complex subunits, have been very highly
148 conserved during primate evolution. This is consistent with their essential cellular
149 housekeeping functions. The only exception is the *Smc6* gene for which weak signatures of
150 adaptation were found. Thus, and in contrast to most other known antiviral restriction factors,
151 the Smc5/6 complex does not appear to have evolved under strong positive selection in
152 primates.
153
154 **Evidence of site-specific adaptive selection in primate Smc6**
155 Our gene-wide evolutionary analyses showed that Smc6 may have been under positive
156 selection in primates. It is formally possible that most of the Smc6 protein has been highly
157 conserved due to its housekeeping function, whilst just a few sites have been engaged in a
158 virus-host interaction. Indeed, previous studies have shown that essential cellular proteins that
159 are usurped by viruses for their replication have evolved under a strong purifying selection

160 background with only the virus-host interaction sites showing rapid evolution (Demogines,
161 Abraham, Choe, Farzan, & Sawyer, 2013; Enard, Cai, Gwennap, & Petrov, 2016; Lou et al.,
162 2016; Ng et al., 2015; Yang, Wong, & Nielsen, 2005). To determine if this was the case for
163 *Smc6*, we characterized in more detail its evolutionary history during primate evolution.
164 Using the BranchSite-REL algorithm from HYPHY (Kosakovsky Pond et al., 2011), we
165 found that positive selection has occurred on a proportion of branches of the primate *Smc6*
166 phylogeny (episodic positive selection; BS-REL, $p<0.05$). To look more specifically for site-
167 specific positive selection, we used MEME and FUBAR, which are the most accepted
168 methods from HYPHY, as well as the posterior probabilities at codon sites in PAML and
169 Bio++ (M8 model) (Figure 2A). We confirmed that most codons in Smc6 have been
170 extremely conserved, with over 90% having a non-synonymous to synonymous nucleotide
171 (dN/dS) substitution ratio lower than 1 (Figure 2B). However, a few sites were identified as
172 having evolved under significant positive selection by one or several methods, consistent with
173 site-specific positive selection in the primate *Smc6* gene (Figure 2).

174 In addition to single codon substitutions, other forms of genetic changes may also be
175 adaptive in a virus-host "genetic conflict" (Daugherty & Malik, 2012; Etienne, 2015; Lou et
176 al., 2016). In particular, recombination, gene deletion or duplication, insertions and deletions
177 (indels) can also be advantageous for the host in the context of a virus-host "arms race".
178 These "genetic innovations" would be missed in typical methods screening for positive
179 selection. We found some indels in genes encoding for several Smc5/6 complex subunits
180 (Dataset S1). In particular, we found a 5-amino acid (aa) indel in the N-terminal region of
181 Smc5 (Figure 2C). The five New World monkey species, including the two sequenced in this
182 study, display a TASFT motif at this position. However, this 5-aa motif was not present in
183 any of the retrieved *Catarrhini* species sequences (n=13) (Figure 2C Hominoids and Old
184 World Monkeys; Dataset S1). To decipher if this 5-aa indel was specific to primates, we

185 extended our analysis to mammals. We found a significant plasticity within this region with
186 both indels and amino acid changes during mammalian evolution. For example, the
187 prosimians' proteins contain a 5-aa motif, but with amino acid differences in the TASFT
188 motif (TVSFT and AVAFT, respectively; Figure 2C). Another remarkable example of this N-
189 terminal plasticity was found in bats (*Chiroptera*). Most bat species contain a 5 amino acid
190 stretch (residues 36-40) but the sequence differs significantly (TVSFI, TVSFT, PDPFT,
191 TDTFT), while some bats carry an 8-10 amino acid deletion within this region (*Rhinolofus*
192 and *Hipposideros* species) (Figure S2). Therefore, although most of the Smc6 protein
193 sequences has been very conserved in primates and more generally in mammals, a few sites
194 have been under significant positive selection and show substantial genetic plasticity. These
195 signatures of genetic changes in an essential protein could be reminiscent of an evolutionary
196 "arms-race" with pathogenic viruses.
197

198 **HBx and the woodchuck WHx counterpart promote degradation of the Smc5/6 complex**
199 **in diverse mammalian cell types and species**

200 We next examined whether the identified interspecies variability of the Smc5/6
201 complex and the marks of adaptive selection in Smc6 have functional consequences on the
202 ability of HBx to promote the degradation of the Smc5/6 complex. Consistent with previous
203 studies (Decorsiere, et al. 2016; Murphy, et al. 2016), we found that HBx transduced in
204 various human cell lines, including HepG2 (hepatocyte carcinoma), 293T (kidney epithelial),
205 and HeLa (epithelial adenocarcinoma), triggers a similar decrease in endogenous Smc6 and
206 Nsmce4A protein levels (Figure S3). Similar results were obtained in U2OS (human
207 osteosarcoma) and A549 (human lung adenocarcinoma) cell lines (data not shown). This
208 indicates that this HBx activity is not cell-type specific.

209 We next tested the capacity of human HBx and the woodchuck WHx counterpart to
210 promote degradation of the Smc5/6 complex in cells derived from various primate species,
211 including human, tantalus African green monkey (*Chlorocebus tantalus*), vervet African
212 green monkey (*Chlorocebus pygerythrus*), and the New World owl monkey (*Aotus*
213 *trivirgatus*), as well as in ferret (*Mustela putorius furo*, carnivore) and mouse (*Mus musculus*,
214 rodent) cells (Figure 3). We amplified and sequenced the *Smc6* gene from these cells which
215 confirmed the breadth of our host species panel (Figure S4). Importantly, our host panel
216 sampled the variants at all the positive selection sites in Smc6, as well as other genetic
217 variability marks (Figure 3A). Cells from these six different species were transduced with a
218 lentivector encoding for GFP alone, GFP-HBx or GFP-WHx. Five days later, we found that
219 HBx and WHx expression had specifically triggered the degradation of the Smc5/6 complex
220 in all species tested (Figure 3B). This was true even for host species like owl monkey, which
221 harbor a 5-amino acid insertion in the N-terminal region of Smc6 (Figure 3A-B, Figure 2C)
222 and for Old World monkeys that, in contrast to New World monkeys and Hominoids, are not
223 natural hosts for hepadnaviruses (Figure 3B). This further shows that inter-species variability,
224 including amino-acid differences at sites under adaptive evolution, in Smc6 do not
225 significantly impact on HBx-mediated degradation of the complex (Figure 3A-B). Taken
226 together these results show that the antagonism of endogenous Smc5/6 by the viral HBx and
227 WHx is independent of the cell type and the mammalian host species.
228

229 **Mammalian hepadnavirus HBx proteins show a conserved ability to counteract the**
230 **restriction activity of the human Smc5/6 complex**

231 In a complementary experiment, we functionally compared the X proteins from
232 diverse mammalian hepadnaviruses for their ability to antagonize the Smc5/6 complex in
233 human cells. To span the entire orthohepadnavirus evolutionary history, we examined, in

234 addition to the human HBx and woodchuck WHx, a newly cloned X protein from a hepatitis
235 B virus that infects the New World wooly monkey and the newly cloned X proteins encoded
236 by three viruses infecting distantly related bat species; *Hipposideros cf. ruber* (roundleaf bat),
237 *Rhinolophus alcyone* (horseshoe bat), and *Uroderma bilobatum* (tent-making bat) (Figure
238 4A). As shown in Figure 4B these X proteins have highly divergent amino acid sequences
239 with some regions sharing essentially no homology. Despite this high sequence divergence,
240 all X proteins showed comparable ability to trigger degradation of the human Smc6 and
241 Nsmce4A proteins, when transduced in either HepG2 or 293T cells (Figures 4C and S5).
242 Previous work has shown that the Smc5/6 complex binds episomal DNA templates to block
243 transcription and that inactivation of the complex leads to an increased episomal gene
244 expression (Decorsière et al., 2016). Accordingly, a similar increase in expression of a
245 transiently transfected episomal luciferase reporter construct was observed in all
246 orthohepadnavirus X protein-expressing cells (Figure 4D). Thus, the ability to degrade and to
247 counteract the restriction activity of the Smc5/6 complex is conserved among mammalian
248 hepadnavirus X proteins.
249

250 **Divergent mammalian X proteins efficiently rescue replication of an HBx-deficient**
251 **hepatitis B virus in primary human hepatocytes**

252 We finally tested whether the X proteins from non-human orthohepadnaviruses would
253 substitute for HBx in an HBV replication assay. Primary human hepatocytes (PHH) were
254 transduced with lentiviruses expressing GFP alone or fused to one of the six
255 orthohepadnavirus X proteins, and four days later infected with the wild-type HBV or an
256 HBx-deficient HBV mutant (HBVΔX) (Figure 5A). As shown previously, human HBx
257 provided in trans fully rescued the replication defect of HBVΔX as measured by HBe and
258 HBs antigen secretion (Decorsière et al 2016). Strikingly, the X proteins from wooly monkey

259 (WMHBx), woodchuck (WHx), and three bat (TBHBx, HBHBx, and RBHBx) viruses were
260 all capable of restoring HBVΔX replication to comparable levels as shown by HBe and HBs
261 antigen secretion (Figure 5A). This occurred in the absence of changes in viral cccDNA levels
262 (Figure 5B) and was accompanied by a concomitant decrease in Smc6 protein levels (Figure
263 5C). These results provide functional evidence that orthohepadnavirus X proteins have a
264 conserved capacity to antagonize the antiviral function of Smc5/6 and that this occurs in a
265 species-independent fashion.
266
267 **Discussion**
268 In this study, we traced the evolutionary history of the genes encoding the six
269 components of the Smc5/6 complex, as well as the genes for the Smc1-4 core subunits of
270 cohesin and condensin. We show that Smc1-4 and the Smc5/6 complex have been highly
271 conserved in primates. The only exception is the *Smc6* gene, which shows some signatures of
272 adaptive selection during mammalian evolution. These include several sites under positive
273 selection and insertion/deletion events during mammalian evolution, which could be
274 reminiscent of a virus-host evolutionary "arms-race".
275 To examine if HBV, which is currently the only virus reported to be restricted by the
276 Smc5/6 complex, has contributed to shape the evolution of the complex, we tested several
277 orthohepadnavirus X proteins for their ability to antagonize the Smc5/6 complex across
278 species. We found that all the X proteins we tested, including those encoded by distantly
279 related bat HBVs, were equally efficient in antagonizing Smc5/6 in human and other
280 mammalian cells. This conserved property suggests that the interaction between the Smc5/6
281 complex and HBx is independent of the sites under genetic conflicts that we identified in this
282 study. Thus, HBx does not seem to have driven the evolution of Smc5/6 in primates and,
283 more broadly, in mammals.

284 Overall, we did not identify strong signatures of positive selection in the Smc5/6
285 complex as typically found in other known restriction factors (Daugherty & Malik, 2012;
286 Duggal & Emerman, 2012; Etienne, 2015; Kluge et al., 2015). This is surprising given that
287 the antagonistic relationship between HBx and Smc5/6 has likely been played out over tens of
288 millions of years (Suh et al., 2014). There are at least three possible explanations for this
289 apparent inconsistency. Firstly, Smc5/6 has an evolutionarily highly conserved architecture
290 and it performs fundamental functions in cellular genome maintenance, and thus has likely
291 evolved under strong purifying selection. This is in contrast to most other described
292 restriction factors, such as APOBEC3G and TRIM5, that are dedicated to their antiviral
293 intrinsic immune functions (Kluge et al., 2015; Malim & Bieniasz, 2012). However, our
294 findings show similarities both at the evolutionary and functional levels with the Serine
295 incorporator proteins SERINC3 and SERINC5. These cellular proteins were recently found to
296 act as restriction factors against lentiviruses (Heigele et al., 2016; Murrell, Vollbrecht,
297 Guatelli, & Wertheim, 2016; Rosa et al., 2015; Usami, Wu, & Gottlinger, 2015). Despite their
298 antiviral function and their antagonism by the lentiviral Nef protein, SERINC3 and SERINC5
299 also show little evidence for positive selection (Murrell, et al. 2016). Like Smc5/6, the
300 SERINC proteins have important cellular functions and are not part of the interferon signaling
301 pathway. The Smc5/6 complex and the SERINC proteins may therefore fall into a separate
302 category of restriction factors that have dual antiviral and essential cellular functions, and
303 therefore have limited evolutionary opportunities. In an extreme scenario, host proteins might
304 not be able to "escape" from the viral antagonist and it is likely that viruses have adapted to
305 target such constrained host proteins (i.e. "viral strategy") (Enard et al., 2016; Jager et al.,
306 2011). The thorough evolutionary analysis of such constrained restriction factors would
307 benefit from novel bioinformatic methodologies specifically designed to identify adaptive
308 evolution that operates on a background of strong purifying selection and that may further

309 involve genetic innovations other than site-specific positive selection (e.g. indels,
310 recombination).

311 A second possible explanation for the lack of strong positive selection in the Smc5/6
312 complex is that orthohepadnaviruses may not have been strong drivers of mammalian genome
313 adaptation. Indeed, acute HBV infection is mostly asymptomatic, and when evolving into
314 chronic infection the symptoms and associated morbidity appear late in life, after the
315 reproductive age (WHO IARC Monograph). Arguing against this interpretation are the
316 findings of Enard and colleagues, who provided evidence that most cellular genes reported in
317 the literature to encode HBV-interacting proteins show a strong excess of adaptation (Enard et
318 al., 2016). This, however, remains to be functionally demonstrated because, to our
319 knowledge, no study to date has reported evidence for a direct evolutionary "arms-race"
320 between an HBV protein and a cellular factor.

321 A third possibility is that HBx interacts with the Smc5/6 complex indirectly, through a
322 yet unknown intermediate cellular protein. Although we cannot formally exclude this
323 possibility, none of our studies so far suggests that this is the case. In particular, no common
324 cellular protein in addition to the DDB1 subunit of the E3 ligase and the six Smc5/6 subunits
325 was recovered from HBx-expressing cells in pull down experiments with either HBx or
326 Smc5/6, as would be expected for an adaptor protein bridging HBx to the complex (FA and
327 MS, unpublished). We can further exclude any "evolutionary conflict" between DDB1 and
328 HBV, because DDB1 is under strong purifying selection (data not shown) and our assays in
329 heterologous species cells allow us to robustly exclude any species-specificity between HBx
330 and a hypothetical endogenous intermediate factor. Nevertheless, the exact HBx-Smc5/6
331 interface remains unknown. Evolutionary and mutagenesis analyses of HBx did not allow us
332 to identify the viral determinants of the interaction with Smc5/6 (data not shown). The

333 analysis of structural/docking models ((Letko, Booiman, Kootstra, Simon, & Ooms, 2015;
334 Richards et al., 2015) for example) might contribute to solve this virus-host interface.
335 It is really remarkable that all the orthohepadnavirus X proteins that we tested
336 antagonize the Smc5/6 complex and share the ability to fully substitute for human HBx in an
337 HBV infection assay using primary human hepatocytes, especially given their low sequence
338 identity (<38% identity for some pairs). These findings suggest that the Smc5/6 antiviral
339 restriction activity is conserved among mammals and provide additional evidence that
340 antagonizing the Smc5/6 complex is a major function of HBx during HBV infection.
341 (Chouteau et al., 2001; Drexler et al., 2013; Lempp et al., 2017). Our findings further imply
342 that, in contrast to what has been documented for other restriction factors (Sawyer & Elde,
343 2012), Smc5/6 does not act as a species-barrier to the potential zoonotic transmission of bat or
344 primate HBVs to humans (Chouteau et al., 2001; Drexler et al., 2013; Lempp et al., 2017;
345 Yan et al., 2012).
346 HBVs are however not restricted to mammals as they are also found in birds, fishes,
347 and amphibians (Dill et al., 2016; Geoghegan, Duchene, & Holmes, 2017; Hahn et al., 2015).
348 Because these latter viruses appear to lack an *X* gene (Dill et al., 2016; Hahn et al., 2015; Suh
349 et al., 2014), it would be interesting to explore if they are also restricted by Smc5/6, and if so,
350 what strategy do these divergent hepadnaviruses use to circumvent the antiviral restriction.
351 Finally, because mammalian *Smc6* has some evidence of adaptive evolution
352 independently of HBV pressure, it raises the possibility that the *Smc6* gene has been engaged
353 in an evolutionary "arms-race" with other pathogenic viruses. For example, the evolutionary
354 adaptive changes identified in the lentiviral restriction factor MxB have not been driven by
355 lentiviruses but likely by other pathogens (P. S. Mitchell, Young, Emerman, & Malik, 2015;
356 Murrell et al., 2016). In addition, restriction factors with a broad antiviral spectrum, such as
357 MxA or PKR, have been evolutionarily driven by several pathogens (Elde, Child, Geballe, &

358 Malik, 2009; Patrick S. Mitchell, Emerman, & Malik, 2013). It will be interesting to
359 determine whether other DNA viruses are restricted by the Smc5/6 complex and, if so,
360 whether they have contributed to shape the evolution of the Smc5/6 complex.
361
362

363 **Methods**

364

365 **Cell lines and culture**

366 Cells were maintained in DMEM supplemented with 10% fetal calf serum. Human
367 cell lines used in the study were: human embryonic kidney 293T cells, the human hepatoma
368 cell lines HepG2 (ATCC HB-8065), and HeLa cells. Old World monkey cells used were:
369 tantalus African green monkey AGM (*Chlorocebus tantalus*) COS-7 cells (a gift from Branka
370 Horvat, CIRI Lyon) and vervet AGM (*Chlorocebus pygerythrus*) Vero cells. New World
371 monkey cells were: Owl monkey (*Aotus trivirgatus*) OMK cells (CelluloNet Lyon) and
372 Coton-headed tamarin (*Saguinus oedipus*) B95a cells (a gift from Branka Horvat). We also
373 used ferret (MPF, *Mustela putorius furo*) and mouse cells (NIH3T3, a gift from Theophile
374 Olmann's lab, CIRI Lyon). The species identity of all the cell lines used in this study was
375 confirmed by amplification and sequencing of the Cytochrome b and/or beta-actin (data not
376 shown).

377

378 **Expression Plasmids**

379 The lentivirus vectors pWPT expressing GFP, GFP-HBx, and GFP-WHx have been
380 previously described (Leupin, Bontron, Schaeffer, & Strubin, 2005; Lin-Marq, Bontron,
381 Leupin, & Strubin, 2001; van Breugel et al., 2012). The X coding regions (synthesized by
382 Genewiz) from hepadnaviruses infecting the New World wooly monkey (*Lagothrix*)
383 (WMHBx), three distant bat species including the roundleaf bat (*Hipposideros cf. ruber*), the
384 horseshoe bat (*Rhinolophus Alcyone*), and the tent-making bat (*Uroderma bilobatum*)
385 (RBHBx, HBHBx, TBHBx, respectively) (Drexler et al., 2013) were expressed from the same
386 vector. The X insert was ligated to the pWPT-GFP backbone between the PstI and NotI sites
387 following the T4 DNA (NEB-M0202) ligation protocol. The plasmid was then transformed

388 following the High Efficiency Transformation Protocol using NEB 10-beta Competent E. coli
389 (NEB-C3019). All the X-expressing plasmids were checked by Sanger sequencing using
390 eGFP-F primer, 5'-CAT GGT CCT GCT GGA GTT CGT G-3', and pWPT-R, 5'-CAT GGT
391 CCT GCT GGA GTT CGT G-3'.
392
393 **Transfection and transduction**
394 For recombinant lentivirus production, plasmids were transfected in 293T cells by the
395 Calcium Phosphate-method. Briefly, $4.5x10^6$ cells were plated in a 10cm-dish and transfected
396 12h later with 15 µg of the X protein-expressing lentiviral vectors, 10 µg of packaging
397 plasmid (psPAX2, gift from Didier Trono (Addgene plasmid # 12260)), and 5 µg of envelope
398 (pMD2G, gift from Didier Trono (Addgene plasmid # 12259)). Media was changed 6h post-
399 transfection. After 48h, the viral supernatants were collected and stored at -80°C. The viral
400 supernatants were titrated by transducing 293T or HeLa cells and measuring the GFP
401 expression by FACS 5 days later. The different mammalian cell lines were transduced by
402 plating 0.1-0.5 $x10^6$ cells in 12-well plates and adding 4-6h later appropriate volumes of viral
403 supernatants to obtain at least 75% of GFP positive cells. Because of the strong post-entry
404 block against the lentiviral vector in OMK cells due to the owl monkey TRIMcyp (Sayah,
405 Sokolskaja, Berthoux, & Luban, 2004), cyclosporine A was added (final concentration
406 2.5µM) to increase transduction efficiencies in these cells. Five days post-transduction, cells
407 were collected for FACS and Western blot analysis. For AGM Cos-7 cells, transduction
408 efficiencies were low and GFP-positive cells were isolated using FACS AsriaII sorting.
409
410 **Luciferase reporter assay**
411 Transfection of plasmid DNA in HepG2 cells was performed using X-tremeGENE HP DNA
412 Transfection Reagent (Roche) following the manufacturer's instructions. For luciferase

413 reporter gene assay cells were typically seeded at a density of about $6x10^5$ cells per 30mm
414 diameter well ($1x10^5$ cells/cm2) and reverse-transfected with 30 ng of reporter plasmid DNA
415 pCMV-Gluc (New England Biolabs) and 2 µg of empty EBS-PL vector. The next day cells
416 were transduced with lentiviral vectors encoding GFP or the diverse GFP-tagged viral X
417 proteins. Five to seven days later GLuc activity was measured by adding 5 µL of the cell
418 supernatant sample to 50 µL room temperature assay buffer (100 mM NaCl, 35 mM EDTA,
419 0.1% tween 20, 300 mM Sodium ascorbate, 0,8 µM coelenterazine in PBS 1X), and
420 immediately measuring luminescence in a luminometer (Glomax, Promega).
421

422 **Western blotting**

423 Western blot analyses were performed as previously described (Decorsiere et al.,
424 2016). Cells were disrupted in 2% SDS and briefly sonicated. The protein concentration was
425 estimated and normalized using the BCA Protein Assay (Novagen). The membranes were
426 probed with 1:5,000 mouse monoclonal anti-GFP antibody (Roche 11814460001) to detect
427 the GFP-tagged X proteins, 1:1,000 mouse monoclonal antibodies against Smc6 (Abgent
428 AT3956a), 1:1,000 rabbit polyclonal anti-Nse4 (Abgent AP9909A), 1:10,000 mouse
429 monoclonal anti-GAPDH (Sigma-Aldrich G8795) antibodies. Horseradish peroxidase-
430 conjugated sheep anti-rabbit or anti-mouse IgG (Amersham Biosciences, at 1:5,000) were
431 used as secondary antibodies. Detection was performed by ECL (Pierce).
432

433 **PHH isolation, HBV infection, ELISA**

434 PHHs were isolated from human liver tissue as previously described (Gripon, Diot, &
435 Guguen-Guillouzo, 1993) from three donors. They were infected with PCR normalized HBV
436 or HBVΔX at an MOI of 500 viral genome equivalents per cell (Lucifora et al., 2011). An
437 ELISA (Autobio Diagnostics) was used to determine the amount of HBe and HBs antigens in

438 culture media from infected or mock infected PHHs. The efficiency of infection was
439 controlled by quantitative PCR analyses specific for the viral cccDNA using beta-globin
440 levels to normalize cell numbers. DNA from infected and transduced cells was extracted
441 using the Epicentre kit. For the cccDNA analysis, T5 exonuclease (TSE44111K Epicentre)
442 was added and the DNA further incubated at 37°C for 30 min and 30 min at 70°C. Specific
443 TaqMan probes and primers were then used to assess the cccDNA content as described in
444 (Allweiss et al., 2017).
445 Transduction levels were estimated by the GFP expression in transduced cells by FACS
446 analyses (BD FACS Calibur).
447
448 ***De novo* sequencing of Smc5/6 genes**
449 Total RNA was extracted from 10^7 cells following the manufacturer instructions
450 (NucleoSpin RNA Blood, Macheray-Nagel 740200). Reverse transcription was performed
451 using the SuperScript III Reverse Transcriptase (ThermoFisher 18080) with random hexamers
452 and oligo(dT). Single-round PCR of overlapping fragments were generated using the Q5 high
453 fidelity DNA polymerase (NEB M0491) following the manufacturer instructions. Sequences
454 of the *Smc6* gene from owl monkey (OMK), coton-headed monkey (B95a), tantalus and
455 vervet AGM (Cos7 and Vero), human (293T and HepG2), and ferret (MPF) cells were
456 obtained. Primers used for the amplification and the sequencing are given in Table S1.
457
458 **Host phylogenetic analyses**
459 Sequences of primate or mammalian orthologous genes were retrieved from publically
460 available databases using UCSC Blat and NCBI Blastn with the human sequence as the query.
461 After including sequences obtained from *de novo* sequencing in this study, the sequences of
462 the orthologues were codon-aligned using Muscle (Edgar, 2004) with minor adjustments

463 (alignments are in Dataset S1). The name of each sequence has been uniformed: the first three
464 letters of the host genus name followed by the first three letters of the species name (e.g.
465 papAnu for *Papio Anubis*, or macMul for *Macaca mulata*). When sequences were retrieved
466 using Blat on the primate full-genome assembly, the name of the genome assembly was used
467 (e.g. panTro4 for *Pan troglodytes*, or hg38 for human). The synteny of each locus of interest
468 was analyzed in UCSC using Blat and the Genome Browser. When necessary, the genomic
469 sequences were further retrieved and aligned to a reference gene to determine the
470 pseudogenes and the gene orders (Figure S1).
471 We used GARD from HYPHY to perform the recombination analyses with a cut-off at
472 $p<0.05$ (Kosakovsky Pond et al., 2006; Pond, Frost, & Muse, 2005). PhyML was used for the
473 phylogenetic reconstructions with a HKY+G+I model and aLRT or 1,000 bootstrap replicates
474 for branch support (Guindon et al., 2010).
475
476 **Positive selection analyses**
477 Maximum-likelihood tests to assay for positive selection were performed using three
478 platforms: HYPHY from Kosakovsky Pond and colleagues, PAML from Yang and
479 colleagues, and Bio++ from Dutheil, Guéguen and colleagues (Gueguen et al., 2013; Pond et
480 al., 2005; Yang, 2007). In HYPHY, we used PARRIS to detect if a subset of sites in the
481 alignment have evolved under positive selection (Scheffler et al., 2006). We further used the
482 more recent BUSTED method, which detects gene-wide evidence of positive selection within
483 a codon-alignment (Murrell et al., 2015). In PAML, we used Codeml with the corresponding
484 gene tree inferred with PhyML as input. Parameters were checked using M0 model. The gene-
485 coding sequence alignments were fit to models that disallow (M7) or allow (M8) positive
486 selection. For several genes, we did not get convergence certainly because genes were too
487 conserved and the sum of dS across the tree was too low. We therefore exclude the Codeml

488 and Bio++ analyses of seven genes, for which the p or the q parameters of the beta
489 distribution was extreme (<0.05 or >99 (Yang, 2007)) (Figure 1C, NA). The likelihood of
490 models was compared using a chi-squared test to derive p values.

491 For the *Smc6* gene, for which we found evidence of positive selection in some of the
492 previously described methods, we further analyzed (i) which lineage(s) during primate
493 evolution have been subjected to positive selection (i.e. when did the gene experience rapid
494 evolution?) and (ii) which sites have been under positive selection (i.e. where has the gene
495 evolved more rapidly than expected?). For branch-specific analyses, we used BS-REL from
496 HYPHY, which identifies if certain lineages have undergone positive selection (Kosakovsky
497 Pond et al., 2011). To detect episodic site-specific positively selected sites, we used MEME
498 from HYPHY (Murrell et al., 2012). We also ran FUBAR in Datamonkey, which uses
499 Bayesian inference to detect positive and negative selection at individual sites (Murrell et al.,
500 2013). In Bio++ and Codeml, we used the Bayesian Posterior Probability and the BEB
501 analysis from the respective M8 model to identify codons with dN/dS>1 (sites with posterior
502 probability > 0.90 are presented here, Figure 2A).

503

504 **Virus phylogenetics**

505 The nucleotide sequences of the *X* gene from orthohepadnaviruses were retrieved
506 using Blastn with the human HBx as the query. Only one sequence per orthohepadnaviral
507 lineage was retained for this analysis. We performed the amino acid alignment with Muscle
508 (total length, 161aa) and we used PhyML for the phylogenetic reconstructions with a
509 JTT+G+I model and 1,000 bootstrap replicates for branch support (Guindon et al., 2010).

510

511 **Ethics statement**

512 PHHs were prepared from surgical liver resections provided by Pr Michel Rivoire's,
513 Pr. Jean-Yves Mabrut's and Dr. Guillaume Passot's departments. Approval from the local and
514 national ethics committees (French Ministry of Research and Education numbers AC-2013-
515 1871, DC-2013-1870, DC-2008-235) and informed consent from patients were obtained.
516
517 **Accession numbers**
518 All the new *Smc6* sequences are available at GenBank under accession numbers MF624755 to
519 MF624761.
520

Acknowledgements

We would like to very sincerely thank Léa Picard and Laurent Guéguen for the Bio++ analyses and input on evolutionary analyses, Véronique Barateau for technical support with the flow cytometry analyses, Olivier Hantz and Isabelle Chemin for scientific discussions, and Loic Peyrot, Anaelle Dubois, and Francoise Berby for technical assistance. We thank Maud Michelet, Jennifer Molle, Loic Peyrot, Anaelle Dubois, Océane Floriot, Laura Dimer, Marie-Laure Plissonnier and Julie Lucifora (CRCL) for the isolation of primary human hepatocytes, and Pr Michel Rivoire's, Pr. Jean-Yves Mabrut's and Dr. Guillaume Passot's staff for providing liver resections. We thank Michael Emerman and Andrea Cimarelli for their comments on earlier version of the manuscript. We thank all the contributors of publically available genome sequences. This work was supported by an amfAR Mathilde Krim Phase II Fellowship (#109140-58-RKHF to LE), an FRM "*Projet Innovant*" grant (#ING20160435028 to LE), a FINOVI "recently settled scientist" grant (to LE), an ANR LabEx ECOFECT grant (GrASP to LE), an ANRS grant (#ECTZ19143 to LE), a fellowship from La Ligue Contre le Cancer (to LG), a grant from the Swiss National Science Foundation (310030-149626 to MS and FA) and by the Canton of Geneva.

Author contributions

Conceptualization: FA, LG, MS, LE

Formal analysis: FF, FA, LG, MS, LE

Funding acquisition: JH, MS, LE

Investigation: FF, FA, LG, AP, JH, MS, LE

Methodology: FA, LG, LE

Project administration: LE

Resources: FF, FA, LG, JH, MS, LE

Supervision: FA, JH, MS, LE

Writing – original draft: FF, LE

Writing – review and editing: FF, FA, LG, JH, MS, LE

References

Allweiss, L., Volz, T., Giersch, K., Kah, J., Raffa, G., Petersen, J., . . . Dandri, M. (2017). Proliferation of primary human hepatocytes and prevention of hepatitis B virus reinfection efficiently deplete nuclear cccDNA in vivo. *Gut*. doi:10.1136/gutjnl-2016-312162

Benhenda, S., Cougot, D., Buendia, M. A., & Neuveut, C. (2009). Hepatitis B virus X protein molecular functions and its role in virus life cycle and pathogenesis. *Adv Cancer Res, 103*, 75-109. doi:10.1016/S0065-230X(09)03004-8

Chouteau, P., Le Seyec, J., Cannie, I., Nassal, M., Guguen-Guillouzo, C., & Gripon, P. (2001). A short N-proximal region in the large envelope protein harbors a determinant that contributes to the species specificity of human hepatitis B virus. *J Virol, 75*(23), 11565-11572. doi:10.1128/JVI.75.23.11565-11572.2001

Cobbe, N., & Heck, M. M. (2004). The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol, 21*(2), 332-347. doi:10.1093/molbev/msh023

Daugherty, M. D., & Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet, 46*, 677-700. doi:10.1146/annurev-genet-110711-155522

Decorsiere, A., Mueller, H., van Breugel, P. C., Abdul, F., Gerossier, L., Beran, R. K., . . . Strubin, M. (2016). Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature, 531*(7594), 386-389. doi:10.1038/nature17170

Demogines, A., Abraham, J., Choe, H., Farzan, M., & Sawyer, S. L. (2013). Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol, 11*(5), e1001571. doi:10.1371/journal.pbio.1001571

Dill, J. A., Camus, A. C., Leary, J. H., Di Giallonardo, F., Holmes, E. C., & Ng, T. F. (2016). Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. *J Virol, 90*(17), 7920-7933. doi:10.1128/JVI.00832-16

Drexler, J. F., Geipel, A., Konig, A., Corman, V. M., van Riel, D., Leijten, L. M., . . . Drosten, C. (2013). Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc Natl Acad Sci U S A, 110*(40), 16151-16156. doi:10.1073/pnas.1308049110

Duggal, N. K., & Emerman, M. (2012). Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat Rev Immunol, 12*(10), 687-695. doi:10.1038/nri3295

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Elde, N. C., Child, S. J., Geballe, A. P., & Malik, H. S. (2009). *Protein kinase R reveals an evolutionary model for defeating viral mimicry.* (Vol. 457): Nature Publishing Group.

Enard, D., Cai, L., Gwennap, C., & Petrov, D. A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *Elife, 5*(e12469), 1-25. doi:10.7554/eLife.12469

Etienne, L. (2015). Virus-Host Evolution and Positive Selection *Virus-Host Evolution and Positive Selection in Encyclopedia of AIDS* (pp. pp1-13).

Feitelson, M. A., Bonamassa, B., & Arzumanyan, A. (2014). The roles of hepatitis B virus-encoded X protein in virus replication and the pathogenesis of chronic liver disease. *Expert Opin Ther Targets, 18*(3), 293-306. doi:10.1517/14728222.2014.867947

Fernandez-Capetillo, O. (2016). The (elusive) role of the SMC5/6 complex. *Cell Cycle, 15*(6), 775-776. doi:10.1080/15384101.2015.1137713

Geoghegan, J. L., Duchene, S., & Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog, 13*(2), e1006215. doi:10.1371/journal.ppat.1006215

Gligoris, T., & Lowe, J. (2016). Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends Cell Biol, 26*(9), 680-693. doi:10.1016/j.tcb.2016.04.002

Gripon, P., Diot, C., & Guguen-Guillouzo, C. (1993). Reproducible high level infection of cultured adult human hepatocytes by hepatitis B virus: effect of polyethylene glycol on adsorption and penetration. *Virology, 192*(2), 534-540. doi:10.1006/viro.1993.1069

Gueguen, L., Gaillard, S., Boussau, B., Gouy, M., Groussin, M., Rochette, N. C., . . . Dutheil, J. Y. (2013). Bio++: efficient extensible libraries and tools for computational molecular evolution. *Mol Biol Evol, 30*(8), 1745-1750. doi:10.1093/molbev/mst097

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Hahn, C. M., Iwanowicz, L. R., Cornman, R. S., Conway, C. M., Winton, J. R., & Blazer, V. S. (2015). Characterization of a Novel Hepadnavirus in the White Sucker (Catostomus commersonii) from the Great Lakes Region of the United States. *J Virol, 89*(23), 11801-11811. doi:10.1128/JVI.01278-15

Heigele, A., Kmiec, D., Regensburger, K., Langer, S., Peiffer, L., Sturzel, C. M., . . . Kirchhoff, F. (2016). The Potency of Nef-Mediated SERINC5 Antagonism Correlates with the Prevalence of Primate Lentiviruses in the Wild. *Cell Host Microbe, 20*(3), 381-391. doi:10.1016/j.chom.2016.08.004

Hwang, G., Sun, F., O'Brien, M., Eppig, J. J., Handel, M. A., & Jordan, P. W. (2017). SMC5/6 is required for the formation of segregation-competent bivalent chromosomes during meiosis I in mouse oocytes. *Development, 144*(9), 1648-1660. doi:10.1242/dev.145607

Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., . . . Krogan, N. J. (2011). Global landscape of HIV-human protein complexes. *Nature, 481*(7381), 365-370. doi:10.1038/nature10719

Jia, B., Serra-Moreno, R., Neidermyer, W., Rahmberg, A., Mackey, J., Fofana, I. B., . . . Evans, D. T. (2009). Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. *PLoS Pathog, 5*(5), e1000429. doi:10.1371/journal.ppat.1000429

Johnson, W. E. (2013). Rapid adversarial co-evolution of viruses and cellular restriction factors. *Curr Top Microbiol Immunol, 371*, 123-151. doi:10.1007/978-3-642-37765-5_5

Kegel, A., & Sjogren, C. (2010). The Smc5/6 complex: more than repair? *Cold Spring Harb Symp Quant Biol, 75*, 179-187. doi:10.1101/sqb.2010.75.047

Kluge, S. F., Sauter, D., & Kirchhoff, F. (2015). SnapShot: antiviral restriction factors. *Cell, 163*(3), 774-774 e771. doi:10.1016/j.cell.2015.10.019

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D., Delport, W., & Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol, 28*(11), 3033-3043. doi:10.1093/molbev/msr125

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol, 23*(10), 1891-1901. doi:10.1093/molbev/msl051

Lauber, C., Seitz, S., Mattei, S., Suh, A., Beck, J., Herstein, J., . . . Bartenschlager, R. (2017). Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of

Non-enveloped Fish Viruses. *Cell Host Microbe, 22*(3), 387-399 e386. doi:10.1016/j.chom.2017.07.019

Lempp, F. A., Wiedtke, E., Qu, B., Roques, P., Chemin, I., Vondran, F. W. R., . . . Urban, S. (2017). Sodium taurocholate cotransporting polypeptide is the limiting host factor of hepatitis B virus infection in macaque and pig hepatocytes. *Hepatology, 66*(3), 703-716. doi:10.1002/hep.29112

Letko, M., Booiman, T., Kootstra, N., Simon, V., & Ooms, M. (2015). Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface. *Cell Rep, 13*(9), 1789-1799. doi:10.1016/j.celrep.2015.10.068

Leupin, O., Bontron, S., Schaeffer, C., & Strubin, M. (2005). Hepatitis B virus X protein stimulates viral genome replication via a DDB1-dependent pathway distinct from that leading to cell death. *J Virol, 79*(7), 4238-4245. doi:10.1128/JVI.79.7.4238-4245.2005

Li, T., Robert, E. I., van Breugel, P. C., Strubin, M., & Zheng, N. (2010). A promiscuous alpha-helical motif anchors viral hijackers and substrate receptors to the CUL4-DDB1 ubiquitin ligase machinery. *Nat Struct Mol Biol, 17*(1), 105-111. doi:10.1038/nsmb.1719

Lim, E. S., Malik, H. S., & Emerman, M. (2010). Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. *J Virol, 84*(14), 7124-7134. doi:10.1128/JVI.00468-10

Lin-Marq, N., Bontron, S., Leupin, O., & Strubin, M. (2001). Hepatitis B virus X protein interferes with cell viability through interaction with the p127-kDa UV-damaged DNA-binding protein. *Virology, 287*(2), 266-274. doi:10.1006/viro.2001.1036

Lou, D. I., Kim, E. T., Meyerson, N. R., Pancholi, N. J., Mohni, K. N., Enard, D., . . . Sawyer, S. L. (2016). An Intrinsically Disordered Region of the DNA Repair Protein Nbs1 Is a Species-Specific Barrier to Herpes Simplex Virus 1 in Primates. *Cell Host Microbe, 20*(2), 178-188. doi:10.1016/j.chom.2016.07.003

Lucifora, J., Arzberger, S., Durantel, D., Belloni, L., Strubin, M., Levrero, M., . . . Protzer, U. (2011). Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. *J Hepatol, 55*(5), 996-1003. doi:10.1016/j.jhep.2011.02.015

Malim, M. H., & Bieniasz, P. D. (2012). HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb Perspect Med, 2*(5), a006940. doi:10.1101/cshperspect.a006940

McBee, R. M., Rozmiarek, S. A., Meyerson, N. R., Rowley, P. A., & Sawyer, S. L. (2015). The effect of species representation on the detection of positive selection in primate gene data sets. *Mol Biol Evol, 32*(4), 1091-1096. doi:10.1093/molbev/msu399

Menolfi, D., Delamarre, A., Lengronne, A., Pasero, P., & Branzei, D. (2015). Essential Roles of the Smc5/6 Complex in Replication through Natural Pausing Sites and Endogenous DNA Damage Tolerance. *Mol Cell, 60*(6), 835-846. doi:10.1016/j.molcel.2015.10.023

Mitchell, P. S., Emerman, M., & Malik, H. S. (2013). An evolutionary perspective on the broad antiviral specificity of MxA. *Current opinion in microbiology, 16*(4), 493-499. doi:10.1016/j.mib.2013.04.005

Mitchell, P. S., Young, J. M., Emerman, M., & Malik, H. S. (2015). Evolutionary Analyses Suggest a Function of MxB Immunity Proteins Beyond Lentivirus Restriction. *PLoS Pathog, 11*(12), e1005304. doi:10.1371/journal.ppat.1005304

Murphy, C. M., Xu, Y., Li, F., Nio, K., Reszka-Blanco, N., Li, X., . . . Su, L. (2016). Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep, 16*(11), 2846-2854. doi:10.1016/j.celrep.2016.08.026

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol, 30*(5), 1196-1205. doi:10.1093/molbev/mst030

Murrell, B., Vollbrecht, T., Guatelli, J., & Wertheim, J. O. (2016). The Evolutionary Histories of Antiretroviral Proteins SERINC3 and SERINC5 Do Not Support an Evolutionary Arms Race in Primates. *J Virol, 90*(18), 8085-8089. doi:10.1128/JVI.00972-16

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., . . . Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Mol Biol Evol, 32*(5), 1365-1371. doi:10.1093/molbev/msv035

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet, 8*(7), e1002764. doi:10.1371/journal.pgen.1002764

Ng, M., Ndungo, E., Kaczmarek, M. E., Herbert, A. S., Binger, T., Kuehne, A. I., . . . Chandran, K. (2015). Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. *Elife, 4*. doi:10.7554/eLife.11785

Niu, C., Livingston, C. M., Li, L., Beran, R. K., Daffis, S., Ramakrishnan, D., . . . Fletcher, S. P. (2017). The Smc5/6 Complex Restricts HBV when Localized to ND10 without Inducing an Innate Immune Response and Is Counteracted by the HBV X Protein Shortly after Infection. *PLoS One, 12*(1), e0169648. doi:10.1371/journal.pone.0169648

Perelman, P., Johnson, W. E., Roos, C., Seuanez, H. N., Horvath, J. E., Moreira, M. A., . . . Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genet, 7*(3), e1001342. doi:10.1371/journal.pgen.1001342

Pond, S. L., Frost, S. D., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics, 21*(5), 676-679. doi:10.1093/bioinformatics/bti079

Richards, C., Albin, J. S., Demir, O., Shaban, N. M., Luengas, E. M., Land, A. M., . . . Harris, R. S. (2015). The Binding Interface between Human APOBEC3F and HIV-1 Vif Elucidated by Genetic and Computational Approaches. *Cell Rep, 13*(9), 1781-1788. doi:10.1016/j.celrep.2015.10.067

Rosa, A., Chande, A., Ziglio, S., De Sanctis, V., Bertorelli, R., Goh, S. L., . . . Pizzato, M. (2015). HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature, 526*(7572), 212-217. doi:10.1038/nature15399

Sauter, D., Schindler, M., Specht, A., Landford, W. N., Munch, J., Kim, K. A., . . . Kirchhoff, F. (2009). Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe, 6*(5), 409-421. doi:10.1016/j.chom.2009.10.004

Sawyer, S. L., & Elde, N. C. (2012). A cross-species view on viruses. *Curr Opin Virol, 2*(5), 561-568. doi:10.1016/j.coviro.2012.07.003

Sawyer, S. L., Wu, L. I., Emerman, M., & Malik, H. S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A, 102*(8), 2832-2837. doi:10.1073/pnas.0409853102

Sayah, D. M., Sokolskaja, E., Berthoux, L., & Luban, J. (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature, 430*(6999), 569-573. doi:10.1038/nature02777

Scheffler, K., Martin, D. P., & Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics, 22*(20), 2493-2499. doi:10.1093/bioinformatics/btl427

Suh, A., Weber, C. C., Kehlmaier, C., Braun, E. L., Green, R. E., Fritz, U., . . . Ellegren, H. (2014). Early mesozoic coexistence of amniotes and hepadnaviridae. *PLoS Genet, 10*(12), e1004559. doi:10.1371/journal.pgen.1004559

Uhlmann, F. (2016). SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Biol, 17*(7), 399-412. doi:10.1038/nrm.2016.30

Usami, Y., Wu, Y., & Gottlinger, H. G. (2015). SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature, 526*(7572), 218-223. doi:10.1038/nature15400

van Breugel, P. C., Robert, E. I., Mueller, H., Decorsiere, A., Zoulim, F., Hantz, O., & Strubin, M. (2012). Hepatitis B virus X protein stimulates gene expression selectively from extrachromosomal DNA templates. *Hepatology, 56*(6), 2116-2124. doi:10.1002/hep.25928

Watanabe, Y. (2005). The importance of being Smc5/6. *Nat Cell Biol, 7*(4), 329-331. doi:10.1038/ncb0405-329

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., . . . Li, W. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife, 2012*(1), 1--28. doi:10.7554/eLife.00049

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol, 24*(8), 1586-1591. doi:10.1093/molbev/msm088

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol, 22*(4), 1107-1118. doi:10.1093/molbev/msi097

Figure legends

Figure 1. Smc6 is the least conserved subunit of the Smc5/6 complex in primates

A, Architecture of the Smc5/6 complex. The complex is made of two core subunits (Smc5 and Smc6) and four non-SMC elements (Nscme1-4).

B, Phylogenetic analysis of primate *Smc6* genes. Sequences were aligned with MUSCLE and phylogeny was performed with PhyML and an HKY+I+G model with aLRT as statistical support (**, aLRT>0.8). Newly sequenced genes (arrow) are indicated. The newly sequenced *Smc6* from *Chlorocebus pygerythrus* (chlPyg_Vero cells) is not represented because the nucleotide sequence is identical to the retrieved *Chlorocebus sabaeus* (chlSab) sequence of *Smc6*. Alignments and phylogenies of the ten SMC analyzed genes are available in Supplementary Dataset S1.

C, Positive selection analysis of the indicated genes during primate evolution. Shown are the p-values obtained using four different methods (BUSTED, PARRIS, PAML Codeml, and Bio++; see Methods). The p-values of the maximum-likelihood tests indicate whether the model that allows positive selection better fits the data (*: statistically significant). NA, results are not available because convergence was not obtained for these genes and/or analyses (see Methods).

Figure 2. Evidence of episodic site-specific positive selection in Smc6 during primate evolution

A, Specific sites in Smc6 are under positive selection. Codon-alignments were analyzed using four different positive selection tests: MEME, which detects site-specific episodic positive selection; FUBAR, similar in a Bayesian framework; and Bio++ and PAML Codeml (M8), which detect site-specific positive selection (see Methods). The table shows the codon-sites

showing significant positive selection (i.e. that passed the widely accepted p-value or posterior probability (PP) cut-offs for each method). The statistical thresholds used in each test are shown in the table. Codons identified as being under positive selection in at least two of the four tests are indicated by an asterisk.

B, Graphic depicting the proportion of sites in Smc6 at a given dN/dS ratio, calculated with BUSTED. A very few number of Smc6 sites are under positive selection. Ratio dN/dS<1, negative selection; dN/dS=1, neutrality; and dN/dS>1, positive selection.

C, Marks of "genetic conflicts" in Smc6. Sites under positive selection in primates, as well as the plasticity of the N-terminal region of Smc6 in mammals are shown. Amino acid alignment was performed with MUSCLE, and residue's color coding is from RasMol. A and B correspond to the globular domain that contain a Walker A and Walker B motif, respectively. Dashes indicate gaps. One non-primate mammal species was arbitrarily chosen for the schematic representation. The mammal sequence illustrates only one possibility of the natural inter-species sequence variations that have been important in this region (Dataset S3). For example, the extensive plasticity of the N-terminal region during bat evolution is further illustrated in Figure S2.

Figure 3. HBx and WHx can degrade the Smc5/6 complex in cells from diverse mammalian species

A, Amino acid differences at the sites of "genetic conflict" in Smc6 between the host species tested in panel B. Note that all statistically significant marks of potential evolutionary "arms-race" identified in Figure 2 are represented. Asterisks denote the sites that were found under positive selection by at least two methods.

B, The human (HBx) and woodchuck (WHx) HBV X proteins promote degradation of the Smc5/6 complex in primate, rodent, and carnivore cells (n=6 species). Cells were either mock

transduced or transduced with a lentivector encoding GFP, GFP-HBx or GFP-WHx. After 4-5 days, cell lysates were prepared and run on an SDS gel and probed for Smc6 and Nsmce4A, two essential subunits of the Smc5/6 complex, using primary antibodies against the human protein. Note that rodent Smc6 could not be detected. GAPDH serves as a loading control.

Figure 4. Highly divergent mammalian HBV X proteins show a conserved property of antagonizing Smc5/6 restriction in human cells

A, Phylogenetic analysis of the X proteins from hepadnaviruses that naturally infect mammals. The viral X proteins tested in our *in vitro* functional assays (C and D) are highlighted by an asterisk. Phylogenetic analysis of orthohepadnaviral X proteins was performed using a 161-amino acid alignment obtained with MUSCLE (Dataset S2) and the tree was built with PhyML and a JTT+I+G model with 1,000 bootstrap replicates. Bootstrap values (> 800/1000) are indicated at the nodes. The tree was rooted for representation purposes according to Drexler et al. 2013 (but the outgroup of orthohepadnavirus is still under debate (Dill et al., 2016)). The scale bar indicates the number of amino acid substitutions per site. We analyzed the X proteins from GSHBV, ground squirrel; ASHBV, arctic squirrel; woodchuck WHV; three bat viruses (RBHBV, HBHBV, and TBHBV naturally infecting *Hipposideros cf. ruber* (roundleaf bat), *Rhinolophus alcyone* (horseshoe bat), and *Uroderma bilobatum* (tent-making bat), respectively); wooly monkey WMHBV; human HBV and other indicated hominoids.

B, Amino acid alignment of the viral X proteins used in B and C. The black-to-white gradient depicts high-to-low sequence identity. The open reading frames (ORFs) overlapping with HBx are shown, as well as the DDB1-binding region in the human HBx protein (Li, Robert, van Breugel, Strubin, & Zheng, 2010). Of note, we suspect that all orthohepadnavirus X

proteins are able to recruit the DDB1 complex although the domain by which they do so appears non-homologous at the sequence level.

C, Degradation of the human Smc5/6 complex by mammalian X proteins. Human hepatoma HepG2 cells were either mock transduced (Mock) or transduced with a lentivector expressing GFP alone or fused to the indicated X proteins. Western blot analysis of the endogenous Smc6 and Nsmce4A was performed (see Methods). GAPDH serves as a loading control. Supplementary Data are found in associated Figure S5.

D, Effect of mammalian X proteins on transiently transfected reporter gene activity. HepG2 cells were transfected with a luciferase reporter construct and the next day transduced with lentiviral vectors expressing the indicated proteins as above. At day 5-7, the luciferase activity was measured and the fold-increase of RLU (relative light units) versus the GFP control condition (set at 1) is shown. The mean of two independent experiments is shown, along with the standard deviation.

Figure 5. The X proteins from six orthohepadnaviruses can all fully rescue the replication defect of an HBx-deficient HBV in primary human hepatocytes (PHHs)

A, PHHs were mock-transduced or transduced with GFP or the indicated X proteins and infected with wild-type HBV or an HBx-deficient HBV (HBVΔX). HBe and HBs antigen secretion was quantified 7 days later by ELISA. Antigen concentrations are relative to wild-type HBV, which was set to 100. Data are mean ± SEM of independent experiments performed with three different PHH donors.

B, The HBV cccDNA levels were measured at day 7 post-infection by real-time PCR. Values are expressed relative to β-globin mRNA levels to normalize to cell number. The results are the mean +/- SEM for the levels seen in PHHs from two donors.

C, Smc6 degradation in PHHs expressing X proteins from different HBV lineages. Protein extracts where prepared from the cells above and Smc6 protein levels were analyzed by Western blot. Actin serves as a loading control. Supplementary Data are found in associated Figure S6.

Table 1. Primate species and sequences that we included in the evolutionary analyses of the SMC complexes. The asterisks indicate the sequences newly obtained in this study. The total number of primate species (sp) and individuals (ID) included in each gene analysis is indicated at the bottom. The abbreviated species names are found in parentheses (See Methods for details).

Supplementary Information and Data

Supplementary Figures

Figure S1. Synteny conservation of Smc5/6 complex genes during primate evolution. The six essential genes of the complex were analyzed in UCSC with the Genome Browser and Blat. For each gene, at least four primate species were fully analyzed: two great apes (human and gibbon), at least one Old World monkey (baboon and/or macaque), and at least one New World monkey (squirrel monkey and/or marmoset). The genes of interest are shown with their orientation (white arrow). Nscme1-4 are abbreviated Nse1-4. The genes in the flanking regions are shown in grey boxes. The reference of the chromosome or assembly is given on the right. For three subunits (Nscme2, Nsmce3, and Nscme4a), we found evidence of potential pseudogenes with premature stop codons (annotated with an asterisk) in non-human primate species.

Figure S2. Plasticity of the N-terminal region of Smc6 during the bat's evolution. Truncated amino acid alignment (region 20-47) of Smc6 sequences from bats (names of bat species are abbreviated: three first letters of the genus followed by the three first letters of the species). On the left, cladogram of the bat Smc6. Amino acid alignment was performed with Muscle, and residue's color coding is from RasMol (in Geneious). Dashes indicate gaps.

Figure S3. HBx from human HBV degrades the human Smc5/6 complex in a cell-type independent manner. Protein expression of endogenous Smc6 and Nscme4A from three human cell lines (HepG2, 293T, HeLa cells) that were previously transduced with a lentivector expressing GFP only (GFP) or the HBx protein from human HBV with the GFP reporter (HBx). GAPDH serves as a loading control.

Figure S4. Phylogeny of the mammalian *Smc6* highlighting the species that were functionally tested in Figure 3. Our panel of host species cell sample Smc6 diversity. Nucleotide sequences were aligned with MUSCLE and phylogeny was built with PhyML and a HKY+I+G model with aLRT as statistical support (Dataset S5, Table S2). The tree was rooted according to the accepted mammalian species tree. Host species that were functionally tested are highlighted with the name of the corresponding cell line in red.

Figure S5. Associated to Figure 4. **The X proteins from six divergent mammalian hepadnaviruses degrade the Smc5/6 complex from (A) human (293T cells), (B) New World owl monkey (OMK cells), and (C) mouse (NIH3T3 cells).**
As in Figure 4, human 293T cells (A), owl monkey OMK cells (B), or mouse NIH3T3 cells were transduced with a lentivector expressing only GFP (control), or the GFP-fused X protein

from diverse hepadnaviruses: human HBV, woodchuck WHV, wooly monkey WMHBV, and viruses from three bat species (RBHBV, HBHBV, and TBHBV naturally infecting *Hipposideros cf. ruber* (roundleaf bat), *Rhinolophus alcyone* (horseshoe bat), and *Uroderma bilobatum* (tent-making bat), respectively), or a mock control. Western blot analysis of the endogenous Smc6 and Nsmce4A was performed (see Methods). Note that mouse Smc6 could not be detected. GAPDH serves as a loading control.

Figure S6. Associated to Figure 5. The X proteins from six orthohepadnaviral lineages can all rescue the replication of an HBx-deficient HBV in primary human hepatocytes (PHHs).

A, Smc6 degradation was seen in PHHs expressing the X protein from different lineages. Protein extracts where prepared from infected cells on day 7 and Smc6 levels analyzed by western blot. Data shown is a representative gel from the PHHs from one donor (different from Figure 5C).

B, Transduction levels were estimated by the GFP expression (from HBx-GFP or GFP alone) in transduced cells by FACS analyses (BD FACS Calibur).

Supplementary Tables

Table S1. Primers used to amplify and sequence endogenous Smc6 from different mammalian cell lines.

Table S2. Species used for the phylogeny of the mammalian *Smc6* (Figure S4). In red: the genes that were newly sequenced

Supplementary Dataset

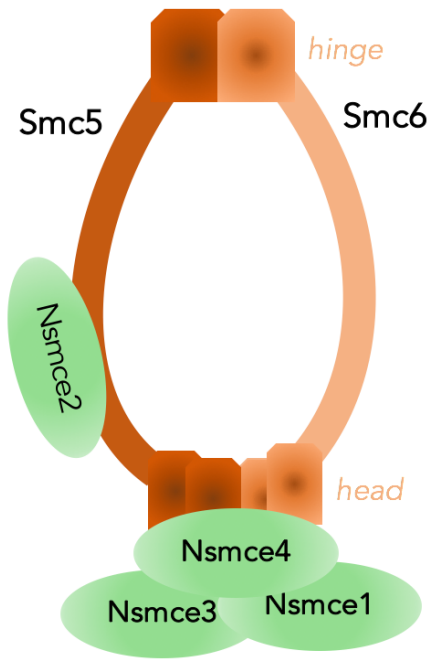
Dataset S1. Host gene alignments used in the study (fasta files) and phylogenetic analyses (newick format)

Dataset S2. Orthohepadnaviral HBx amino acid alignment (interleaved phylip format)

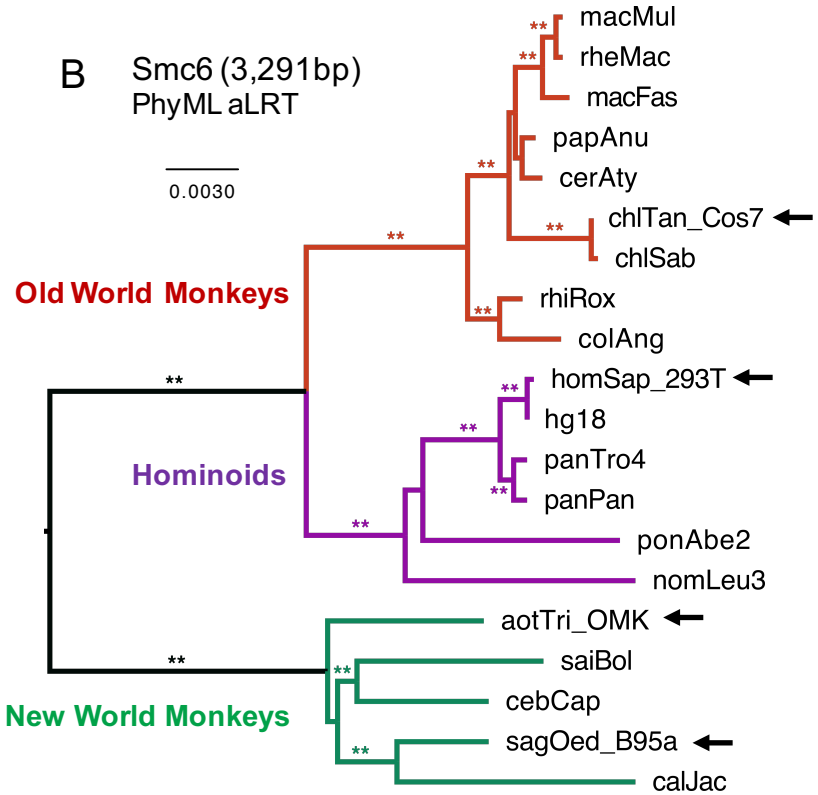
Dataset S3. Phylogenetic analyses of Smc5/6 in mammals (fasta and newick format)

Figure 1 – Filleton et al.

A



B



C

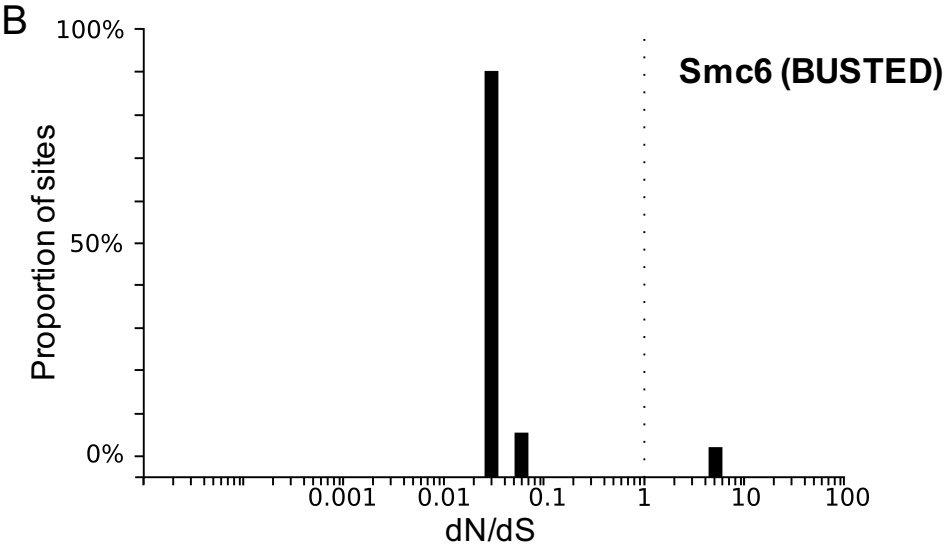
	Smc5/6 complex						Cohesin (Smc genes)		Condensin (Smc genes)	
	Nsmce1	Nsmce2	Nsmce3	Nsmce4a	Smc5	Smc6	Smc1	Smc3	Smc2	Smc4
BUSTED (p value)	1	0.98	1	1	1	0.02*	1	1	0.81	1
PARRIS (p value)	0.48	1	1	0.81	1	0.19	1	1	1	1
PAML Codeml M7 vs M8 (p value)	1	NA	NA	NA	NA	0.48	NA	NA	NA	1
Biopp M7 vs M8 (p value)	1	NA	NA	NA	NA	0.05*	NA	NA	NA	0.94

Figure 2 – Filleton et al.

A

Method		Codons under positive selection						
MEME	(p<0.1)		279	426	489	628		930
FUBAR	(PP>0.8 for β>α)	173		426			697	930
BPP M8	(PP>0.9)						697	930
PAML M8	(PP>0.9)							930
Intersect	Total	173	279	426*	489	628	697*	930*

B



C

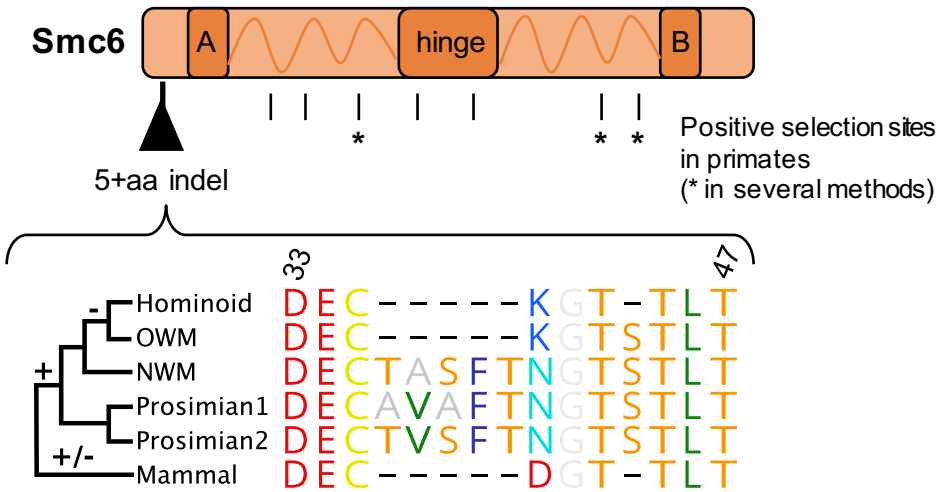
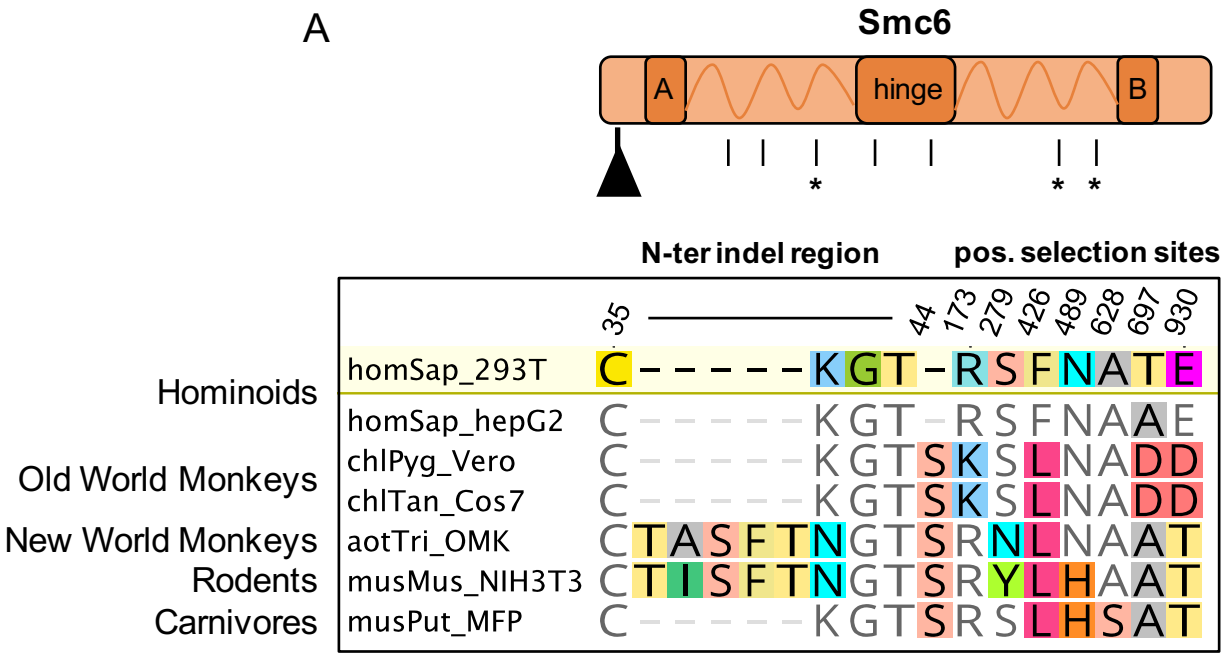


Figure 3 – Filleton et al.

A



B

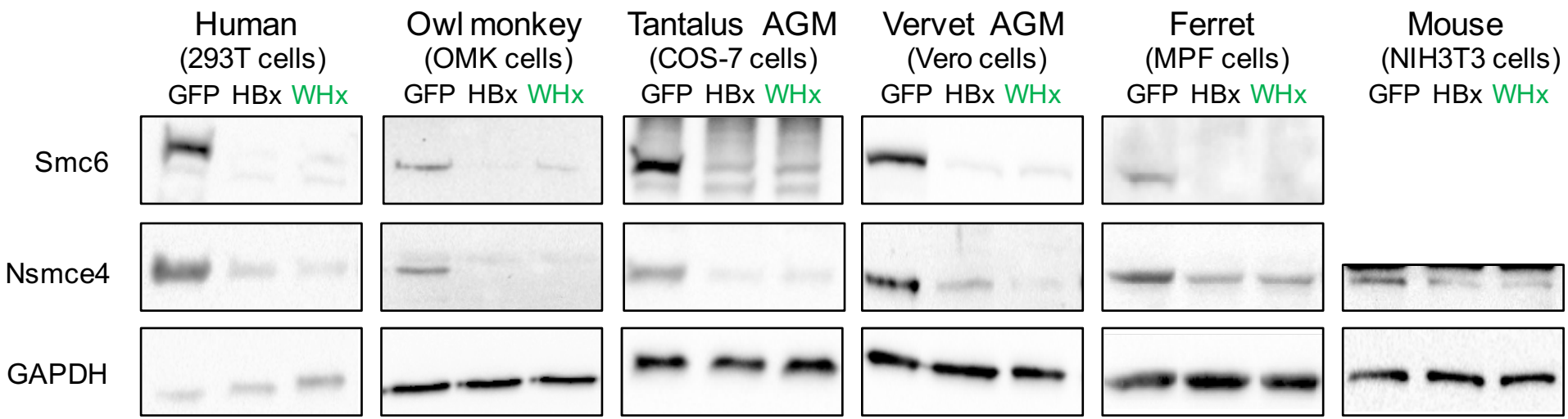
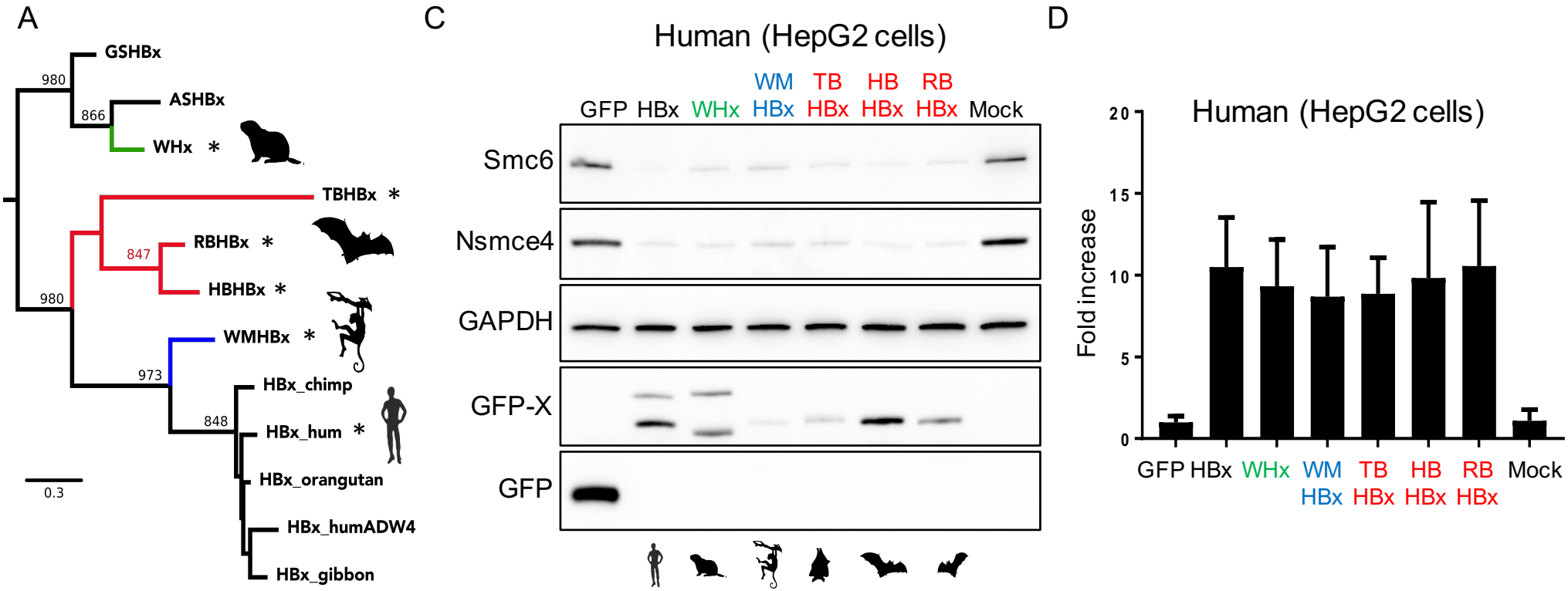


Figure 4 – Filleton et al.



B

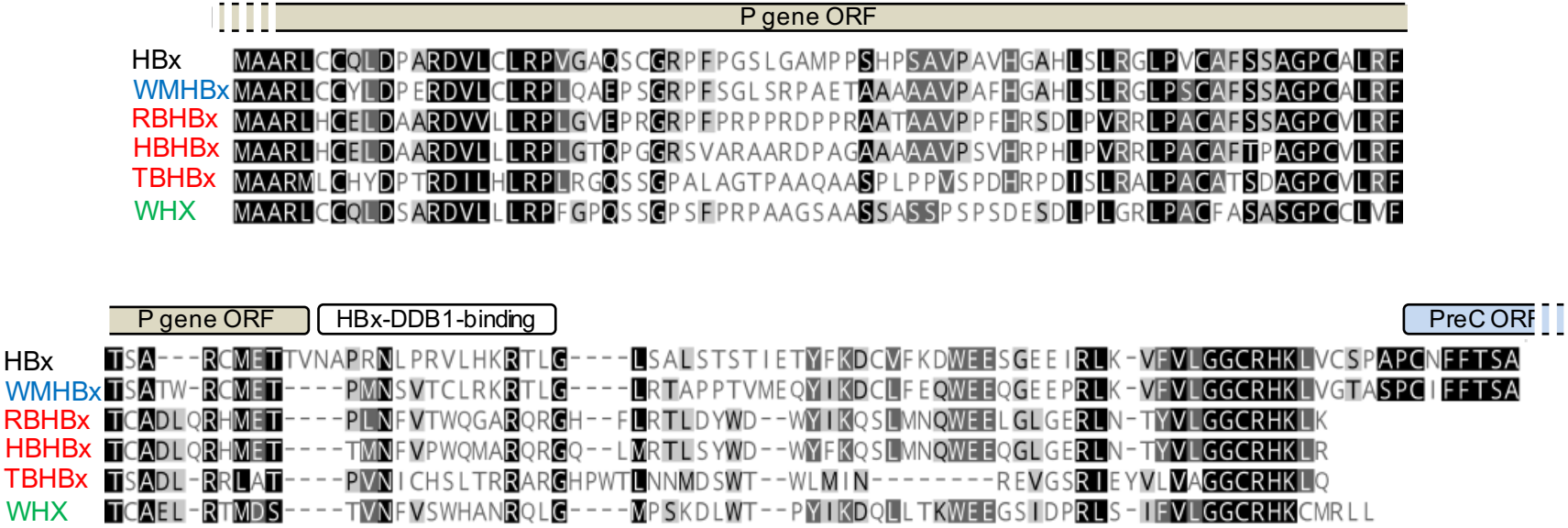
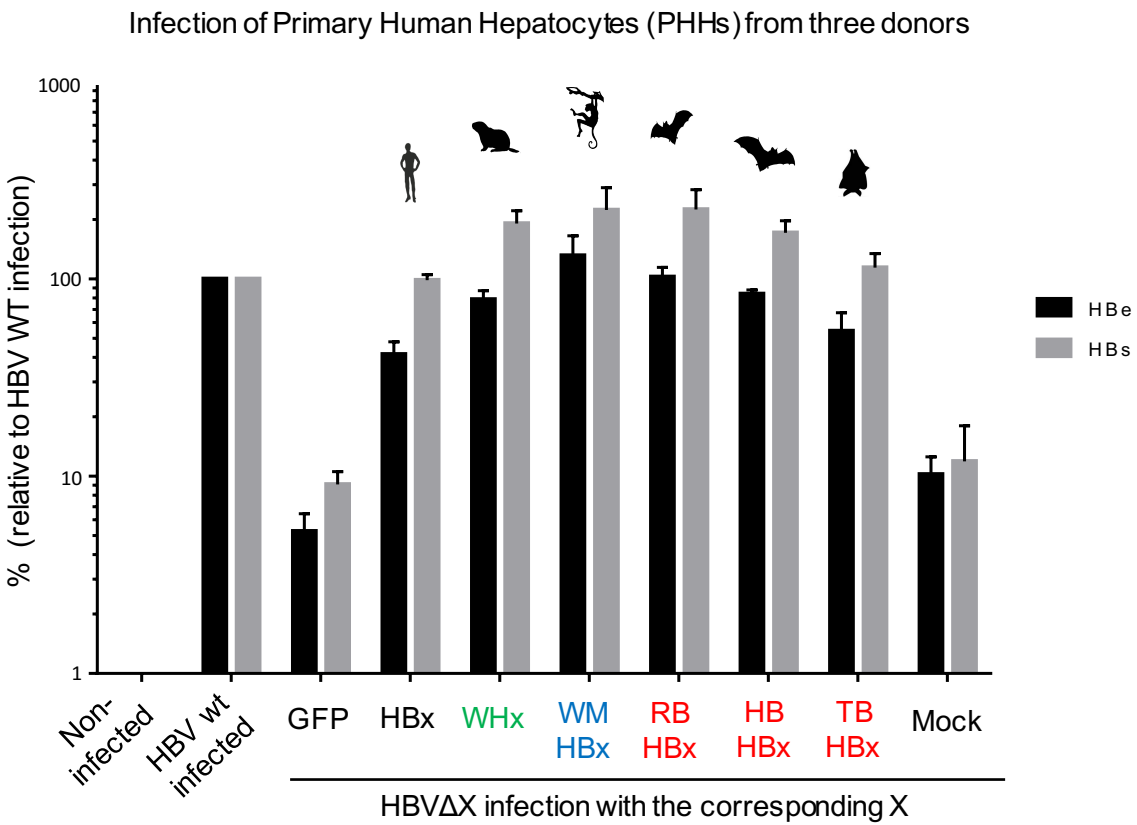
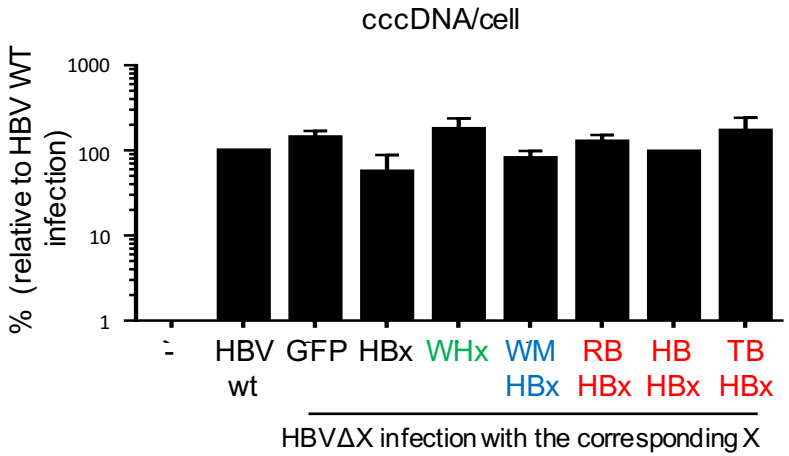


Figure 5 – Filleton et al.

A



B



C

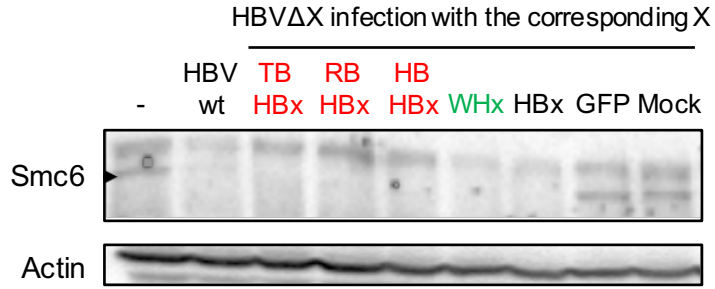


Table 1 – Filleton et al.

		Smc5/6 complex						Cohesin (Smc genes)		Condensin (Smc genes)	
		Nsmce1	Nsmce2	Nsmce3	Nsmce4a	Smc5	Smc6	Smc1	Smc3	Smc2	Smc4
Hominoids	Human	1	1	1	1	1	2*	1	1	1	1
	Chimp (panTro)	1	1	1	1	2	1	1	2	1	1
	Bonobo (panPan)	1	1	1		1	1	1	1	1	1
	Gorilla (gorGor)	1	1	1		1	1		2	1	1
	Orangutan (ponAbe)	1	1	1	1		1	1	1		
	Gibbon (nomLeu)	1	1	1				1		1	1
Old World Monkeys (OWMs)	Olive baboon (papAnu)	1	1	1	1	1	1		1	1	1
	Sooty mangabey (cerAty)	1	1	1		1	1	1		1	1
	Drill (manLeu)	1	1	1		1		1	1		1
	Sabaeus AGM (chlSab)	1	1	1	1	1	1	1	1	1	1
	Tantalus AGM (chlTan)						1*				
	Rhesus macaque (macMul)	2	2	2	1	1	2	1	1	2	1
	Pig-tailed macaque (macNem)	1		1	1	1		1		1	1
	Crab-eating macaque (macFas)	1	1	1	1	1	1	1	1	1	1
	Colobus (colAng)	1	1	1	1	1	1	1	1	1	1
	Golden snub-nosed monkey (rhiRox)	1	1	1	1	1	1	1	1	1	1
	Black snub-nosed monkey (rhiBie)	1	1	1	1			1	1		1
New World Monkey (NWMs)	Marmoset (calJac)		1	1	1	1	1	1	1	1	1
	Coton-headed tamarin (sagOed)						1*				
	Three-stripped owl monkey (aotTri)						1*				
	Nancy Ma's owl monkey (aotNan)	1	1		1	1		1	1	1	1
	Squirrel monkey (saiBol)	1	1	1		1	1	1		1	
	Capuchin (cebCap)	1	1	1	1	1	1	1	1		1
		20 ID / 19 sp	20 sp	20 ID / 19 sp	14 sp	18 ID / 17 sp	20 ID / 18 sp	18 sp	18 ID / 16 sp	17 ID / 16 sp	18 sp

Supplementary Data

Figure S1 – Filleton et al.

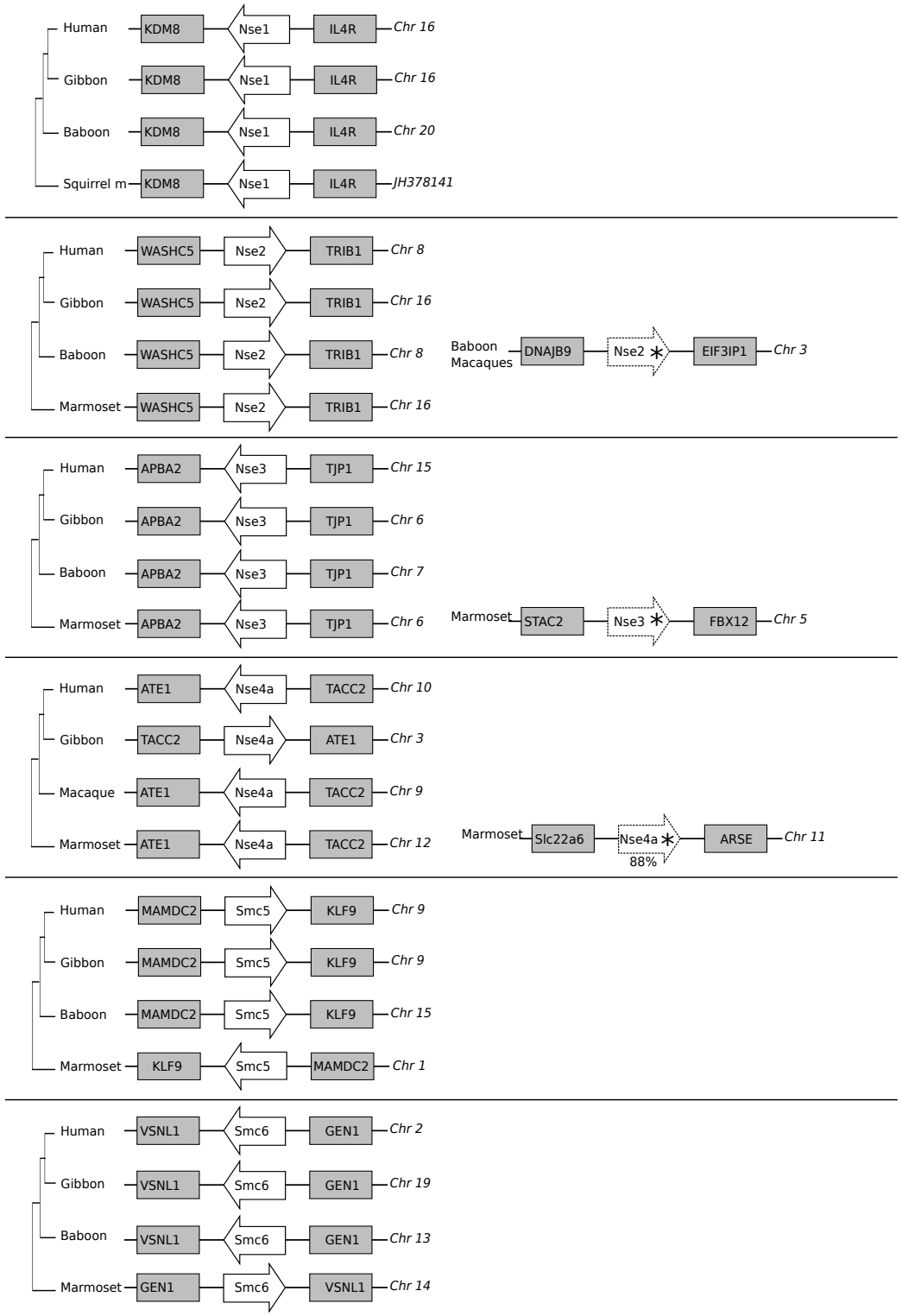


Figure S2 – Filleton et al.

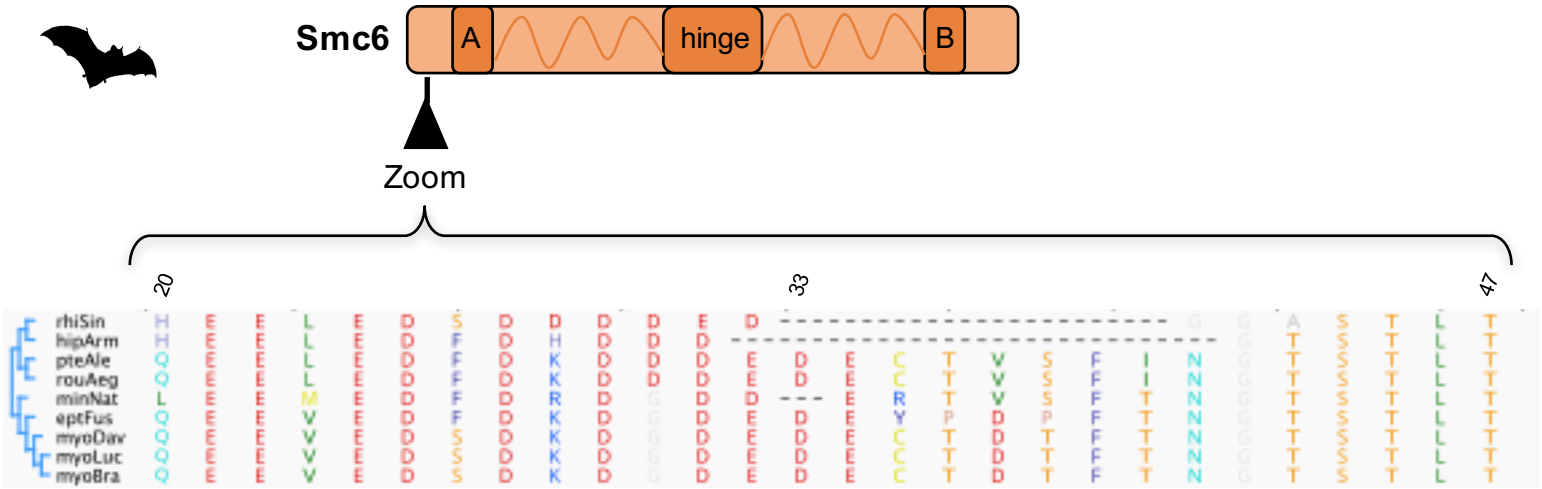


Figure S3 – Filleton et al.

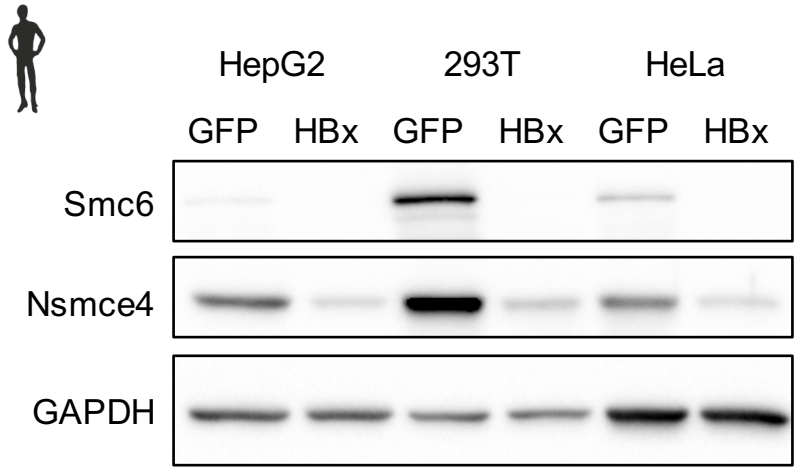


Figure S4 – Filleton et al.

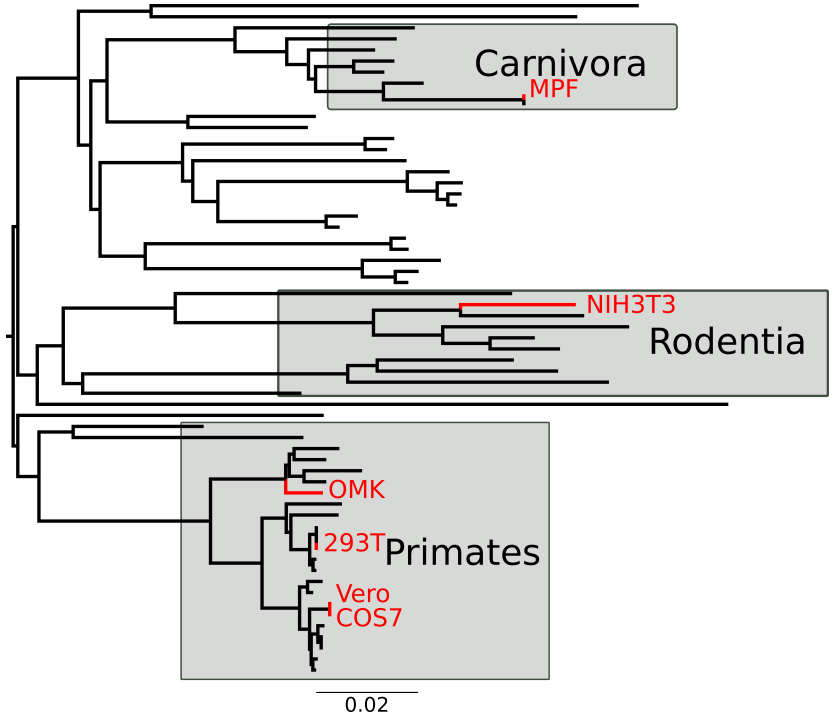


Figure S5 – Filleton et al.

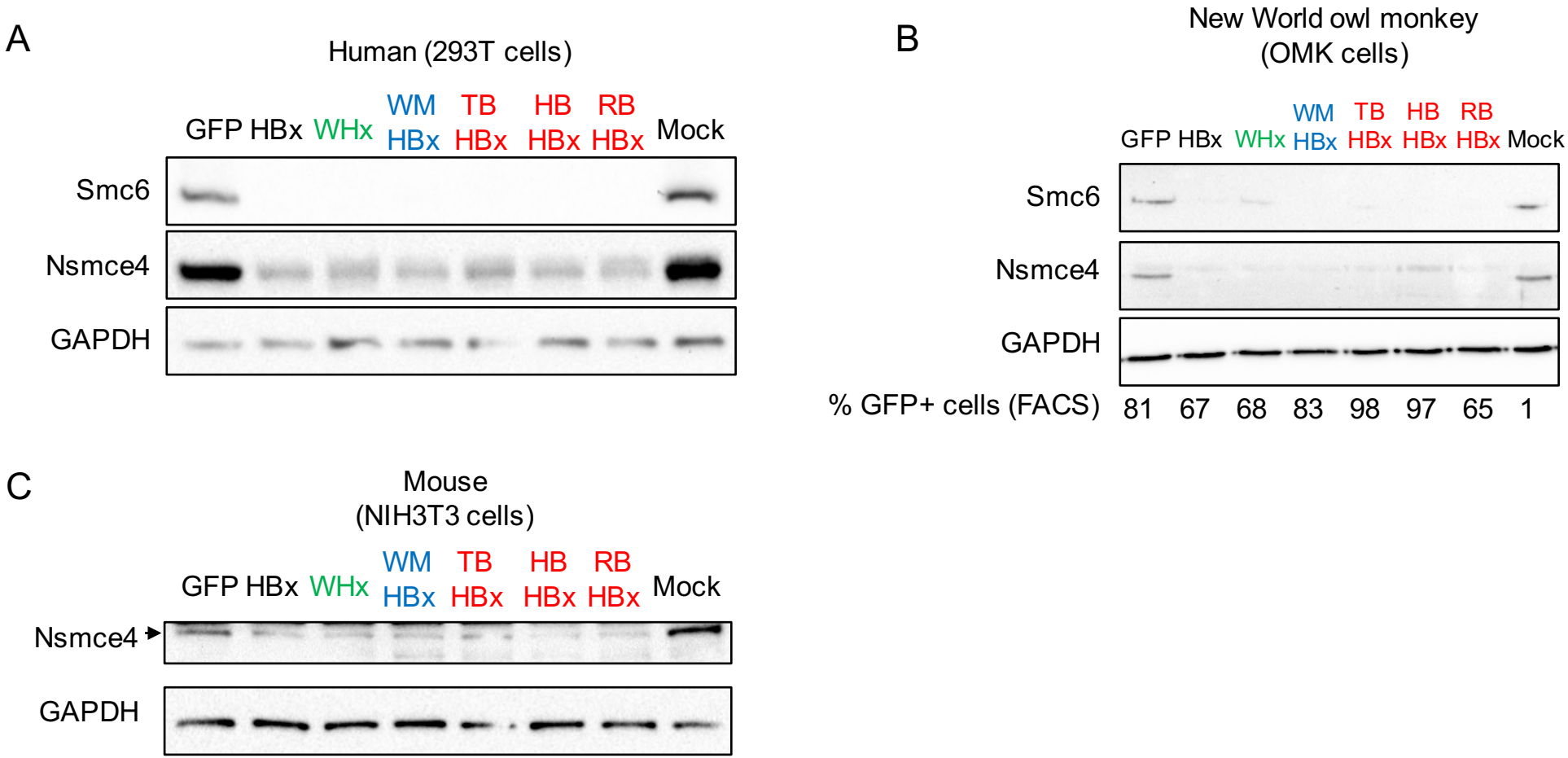
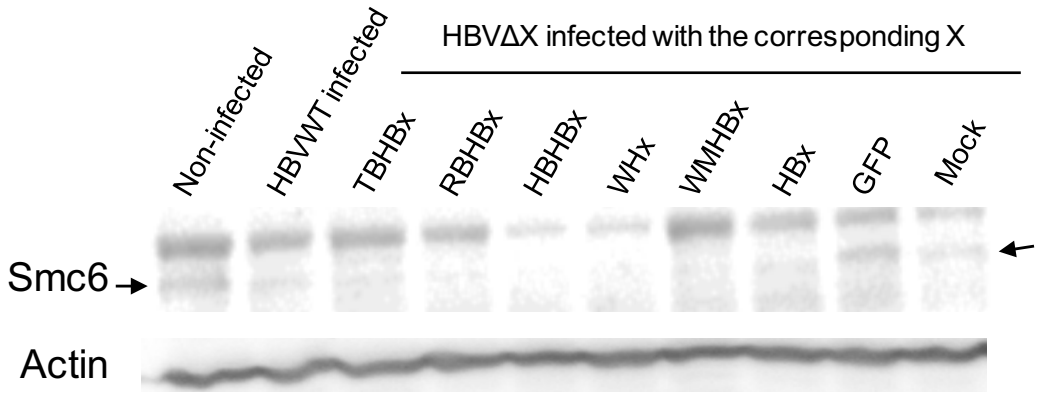
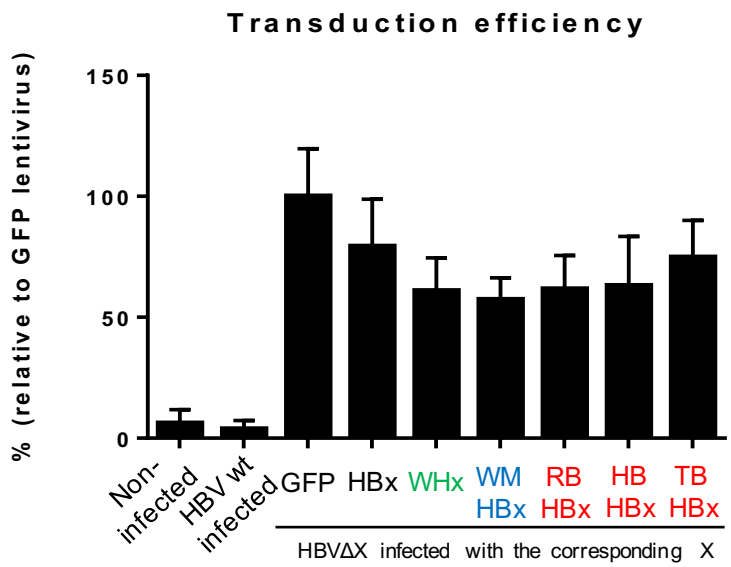


Figure S6 – Filleton et al.

A



B