Estimating hepatitis B virus cccDNA persistence in chronic infection

Katrina A. Lythgoe[1,2], Sheila F. Lumley[3,4], Lorenzo Pellis[5], Jane A. McKeating[6+] and Philippa C. Matthews[3,4,7+]

[1] Big Data Institute, University of Oxford, Old Road Campus, Oxford, OX3 7LF.
[2] Department of Zoology, University of Oxford, Medawar Building, South Parks Road, Oxford OX1 3SY.
[3] Nuffield Department of Medicine, University of Oxford, Medawar Building, South Parks Road, Oxford OX1 3SY.
[4] Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU.
[5] Department of Mathematics, University of Manchester, Manchester, UK
[6] Nuffield Department of Medicine Research Building, University of Oxford, Oxford, OX3 7LF.
[7] NIHR Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU.

[+] Denotes equal contribution

Corresponding author: Katrina Lythgoe, Email: katrina.lythgoe@bdi.ox.ac.uk

Running header: cccDNA persistence during HBV infection

Key words: hepatitis B virus, cccDNA, persistence, evolution, dynamics, reservoir, modelling

Author contributions:
KAL: Conceptualisation, Formal analysis, Funding acquisition, Methodology, Writing – original draft, Writing – Review & Editing
SFL: Formal analysis, Investigation, Writing – original draft
LP: Methodology, Writing – Review & Editing
JAM: Conceptualisation, Writing – Review & Editing
PCM: Conceptualisation, Writing – Review & Editing

Abstract

Hepatitis B virus (HBV) infection is a major global health problem with over 240 million infected individuals at risk of developing progressive liver disease and hepatocellular carcinoma. HBV is an enveloped DNA virus that establishes its genome as an episomal, covalently closed circular DNA (cccDNA) in the nucleus of infected hepatocytes. Currently available standard-of-care treatments for chronic hepatitis B (CHB) include nucleos(t)ide analogues (NA) that suppress HBV replication but do not target the cccDNA and hence rarely cure infection. There is considerable interest in determining the lifespan of cccDNA molecules to design and evaluate new curative treatments. We took a novel approach to this problem by developing a new mathematical framework to model changes in evolutionary rates during infection which, combined with previously determined within-host evolutionary rates of HBV, we used to determine the lifespan of cccDNA. We estimate that during HBe-antigen positive ($HBeAg^{POS}$) infection the cccDNA lifespan is 61 (36-236) days, whereas during the $HBeAg^{NEG}$ phase of infection it is only 26 (16-81) days. We found that cccDNA replicative capacity declined by an order of magnitude between $HBeAg^{POS}$ and $HBeAg^{NEG}$ phases of infection. Our estimated lifespan of cccDNA is too short to explain the long durations of chronic infection observed in patients on NA treatment, suggesting that either a sub-population of long-lived hepatocytes harbouring cccDNA molecules persists during therapy, or that NA therapy does not suppress all viral replication. These results provide a greater understanding of the biology of the cccDNA reservoir and can aid the development of new curative therapeutic strategies for treating CHB.

Author Summary

Nearly one million people die each year due to hepatitis B virus (HBV) related diseases. Although antiviral treatments for HBV exist, cure is rare and treatment is typically life-long, reflecting the persistence of episomal copies of the viral DNA (cccDNA) in the liver. Our knowledge of the cccDNA reservoir in chronic hepatitis B (CHB) is limited. HBV has a high mutation rate and the key determinants of cccDNA dynamics can be inferred by examining the rate of viral evolution. Combining a mathematical model and known rates of HBV evolution we estimate the cccDNA lifespan during different phases of CHB. Our results provide important insights into the dynamics of the HBV reservoir that will inform the design of future treatment interventions.

75 **INTRODUCTION**

76 Hepatitis B Virus (HBV) is a global health problem, with more than 240 million
77 chronically infected individuals at risk of developing liver fibrosis, cirrhosis, and
78 hepatocellular carcinoma [1]. HBV is the prototypic member of the *hepadnaviruses*, a
79 family of small, enveloped hepatotropic viruses with a partially double-stranded
80 relaxed circular DNA (rcDNA) genome that replicates via reverse transcription of
81 pregenomic RNA (pgRNA). Following the infection of hepatocytes, the rcDNA genome
82 is imported to the nucleus and converted to covalently closed circular DNA (cccDNA),
83 that provides the transcriptional template for pregenomic and subgenomic RNAs.
84 Newly synthesized pgRNA is assembled into nucleocapsids that undergo reverse
85 transcription to generate rcDNA, which is subsequently enveloped and released as
86 infectious virions. Alternatively, capsids can be redirected to the nucleus to replenish
87 and maintain the episomal pool of cccDNA and this intracellular amplification pathway,
88 together with the long half-life of cccDNA, contributes to viral persistence [2,3].

89

90 Chronic hepatitis B (CHB) is usually treated with nucleos(t)ide analogues (NAs) that
91 inhibit the reverse transcription of pgRNA to rcDNA. However, these therapies do not
92 directly target the cccDNA reservoir [4] and viremia rebounds once treatment is
93 stopped, even when peripheral levels of viral DNA have remained undetectable for
94 months or years [5]. There is a growing impetus to identify curative therapies for HBV
95 [6]. Despite its central role in the HBV life cycle, our understanding of the viral and
96 host factors that regulate cccDNA abundance and half-life is limited[7]. cccDNA half-
97 life can be defined as the time for the number of copies in the liver to reduce by half
98 and will depend on a number of factors, including cccDNA 'lifespan' (the time an

99 individual cccDNA molecule persists) and the genesis of new cccDNA via extra-cellular
100 virus infection or intra-cellular amplification [8,9] (Fig.1a). A recent *in vitro* study
101 reported a 40 day half-life of HBV cccDNA in infected HepG2-NTCP cells [2], with an
102 estimated lifespan of 58 days. However, the cccDNA lifespan in the human liver is
103 unknown.
104
105 The lifespan of cccDNA molecules most likely changes over the course of CHB,
106 influenced by host and viral factors [10], including the rate of hepatocyte proliferation
107 [11,12]. The natural history of CHB is often classified by four clinical or virological
108 phases of infection and/or hepatitis (according to European Association for the Study
109 of the Liver 2017[13]). However, it is increasingly recognised that this description fails
110 to capture the diversity of viral replication in the liver and misleadingly suggests a
111 linear progression over time from one phase to the next. To avoid difficulties
112 associated with this rigid phenotypic structure of disease, and to focus on viral
113 dynamics rather than disease pathology, we instead consider two states: $HBeAg^{POS}$
114 that is associated with high peripheral HBV DNA levels (viral load - VL), and $HBeAg^{NEG}$,
115 which is associated with lower VLs [14]. Early in infection cccDNA is transcriptionally
116 active and translation of pre-core/pgRNA results in detectable levels of hepatitis B e
117 antigen (HBeAg) in the periphery, associated with high VL [15]. In later stages of
118 infection after HBeAg seroconversion there is a loss of HBeAg that associates with
119 more efficient immune targeting of infected cells [16], leading to a reduction in VL and
120 a shortening of cccDNA lifespan. The higher hepatocyte death rates during $HBeAg^{NEG}$
121 CHB infection will induce hepatocyte proliferation [15]. The extent to which cccDNA is
122 lost during hepatocyte mitosis is uncertain [9], but unless all cccDNA episomes survive

123 mitosis, the increased proliferation rate of infected cells will shorten the average
124 lifespan of cccDNA [12,17,18].
125
126 Double-stranded DNA viruses typically have low mutation rates, but since rcDNA is
127 generated via error prone reverse transcription in the *hepadnaviridae*, they have
128 higher mutation rates than other DNA viruses (Fig 1A) [19]. The estimated mutation
129 rate for avian hepadnavirus is $2x10^{-5}$ substitutions per site per infected cell (s/s/c) [20],
130 and an upper limit of $8.7x10^{-5}$ s/s/c has been estimated for HBV [21], similar to
131 estimates for RNA retroviruses [22]. The evolutionary rate measures how quickly
132 mutations become fixed in a population over a period of time [23]. Strikingly, the
133 evolutionary rate of HBV is much lower than for RNA viruses with similar mutation
134 rates [24–26]. Different mechanisms could explain this observation, including the
135 biological constraint of multiple overlapping reading frames in the HBV genome [24],
136 limited viral population size in the liver, or long cccDNA lifespan [24,25]. Both
137 evolutionary constraint and population size should only influence the rate of evolution
138 of variants that experience selection [27]. For neutral or near neutral mutations, the
139 long cccDNA lifespan provides the simplest explanation for the low evolutionary rate
140 of HBV [25].
141
142 We propose the within-host evolutionary rate of HBV can be used to estimate cccDNA
143 lifespan and the replicative capacity (a combined measure of extra-cellular viral
144 production and intra-cellular amplification) of the virus. We developed a novel
145 mathematical model that uses published HBV mutation and evolutionary rates
146 [20,24,26], to infer the lifespan and replicative capacity of cccDNA during different

147 phases of CHB. Although we use our model to consider the neutral rate of evolution,
148 it is essentially an ecological model, in that we do not assume any viral phenotypic
149 evolution. To the best of our knowledge, these are the first estimates of cccDNA
150 lifespan in treatment naïve subjects and provide important insights into the HBV
151 reservoir that will be valuable for the design and evaluation of future treatment
152 interventions.
153

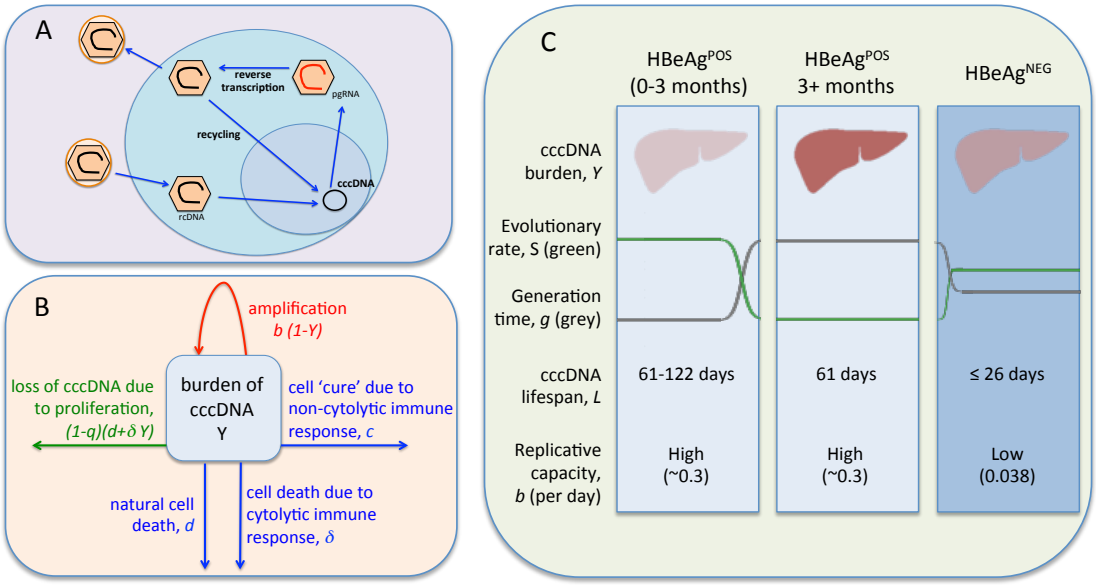


Fig 1: **Summary of the HBV life cycle, mathematical model, and estimated cccDNA lifespan. A:** Simplified HBV replication cycle. A virus particle containing relaxed circular DNA (rcDNA) enters a hepatocyte (blue circle) and is uncoated. The rcDNA is transported to the nucleus (purple circle) and repaired to generate cccDNA. This cccDNA is the transcriptional template for all viral RNAs, including pre-genomic (pgRNA), which is transported to the cytoplasm, encapsidated, and converted into rcDNA by error-prone reverse transcription. The encapsidated rcDNA can be transported back into the nucleus to form more cccDNA (intra-cellular amplification), or enveloped and released as virions that can infect hepatocytes (extra-cellular amplification). **B**: Structure of the mathematical model. This is a single compartment model representing the burden of cccDNA in the liver, Y, over the course of infection. The cccDNA burden can increase due to amplification (intra- and extra-cellular), where *b* is a measure of the within-host replicative capacity of cccDNA. cccDNA can be cleared from the liver due to natural cell death, at rate *d* per day, cytolytic immune responses at rate δ per day, and non-cytolytic immune responses at rate *c* per day. Proliferation can also result in loss of cccDNA at rate $(1-q)(d+\delta Y)$, where *q* is probability that an individual cccDNA survives mitosis. **C**: Representation of the model dynamics and key results, where the numbers give the most likely values inferred by fitting the mutation and evolutionary rates to the model. The darker the colours on the figure the higher the cccDNA burden (reds) and the stronger the immune response (blues).

176 **RESULTS**

177 We developed a mathematical model describing the number of cccDNA molecules in
178 the liver that is independent of infected cell frequency, and accounts for intra- and
179 extra-cellular cccDNA amplification and loss of cccDNA during hepatocyte mitosis (Fig
180 1B and Methods, Eqs 1 and 2). Using this model we derived expressions for the viral
181 generation time, defined as the typical time for one cccDNA molecule to generate
182 another cccDNA molecule at time *t* since infection, *g(t)* (Eq 5), and the neutral rate of
183 evolution at time *S(t)* (Eq 6). At equilibrium, we show that the lifespan of cccDNA, L_E,
184 is equal to the virus generation time, g_E, which is given by the neutral mutation rate
185 divided by the neutral rate of evolution, μ/S_E (Eq 8). The notation used throughout is
186 given in Table 1.

187

188 **Table 1: Notation used to describe the model**

$N(t)$, N_E	Number of cccDNA molecules at time t since infection and at equilibrium
K	Maximum possible number of cccDNA molecules
$Y(t)$, Y_E	cccDNA burden in the liver at time t since infection and at equilibrium, expressed as a proportion of the maximum possible number of cccDNA, $Y(t)=N(t)/K$, $Y_E= N_E/K$
$\rho(t)$	Per capita proliferation rate of cccDNA at time t since infection
$g(t)$, g_E, $g_{Y<<1}$	Generation time of cccDNA at time t since infection, at equilibrium, and when few cells are infected
$S(t)$, S_E	Neutral rate of evolution at time t since infection and at equilibrium

L_E , $L_{Y<<1}$	Life span of cccDNA molecules at equilibrium and when few cells are infected
β	Per capita replicative capacity, defined as the per capita growth rate of cccDNA when few cells are infected and in the absence of infected cell death or loss of cccDNA due to non-cytolytic immune responses.
R_0	The basic reproductive number of cccDNA (the number of cccDNA molecules a single cccDNA produces in its lifetime in an otherwise uninfected population of hepatocytes)
$b=\beta K$	Replicative capacity of cccDNA (a rescaled measure of the per capita replicative capacity)
d	Per capita natural death rate of hepatocytes
δ	Per capita additional death rate of infected hepatocytes due to cytolytic immune responses
c	Per capita loss rate of cccDNA due to non-cytolytic immune responses
q	Probability that a cccDNA molecule survives mitosis
μ	Mutation rate of cccDNA (substitutions per site per reproduction)

[+] Throughout we use days as the time unit and all rates are per day.

Lifespan of cccDNA

From the published estimates for the mutation [20] and evolutionary rates of HBV [24,26], we inferred the probability distributions for cccDNA lifespan during $HBeAg^{POS}$ and $HBeAg^{NEG}$ phases of CHB using Eq 8. The predicted lifespan of cccDNA during $HBeAg^{POS}$ infection, and when VLs are high and stable, is 61 days (5th-95th percentiles 36-236 days; Fig 2, orange line). In contrast, during $HBeAg^{NEG}$ infection the lifespan of cccDNA is estimated at only 26 days (16-81 days; Fig 2, blue line).

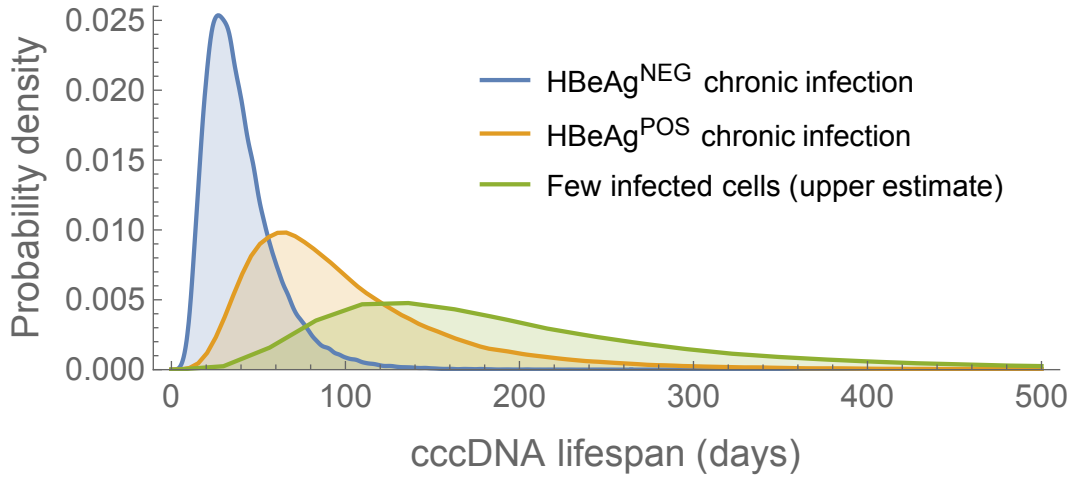
197

198 The shorter lifespan of cccDNA during HBeAg^{NEG} compared to HBeAg^{POS} infection can
199 be explained by higher rates of cccDNA clearance (Eq 9). This may reflect changes in
200 the immune environment due to HBe-antigen seroconversion that is associated with
201 increased cytolytic and non-cytolytic immune responses (δ and c respectively).
202 Increased host immune responses during HBeAg^{NEG} infection could push the basic
203 reproduction number, R_0, of cccDNA below one (Eq 3) due to the higher clearance
204 rates of cccDNA molecules, and also due to reduced replicative capacity, b, of cccDNA.
205 If R_0<1, the number of cccDNA will not reach a stable level but will continually decline.
206 In this non-equilibrium situation the lifespan of cccDNA may be less than our inferred
207 26 days since the viral generation time will be greater than the lifespan of cccDNA (Fig
208 3 and methods).

209

210 Our model suggests cccDNA lifespan can be up to two times longer when few cells are
211 infected compared to when most cells are infected cells (see methods; compare Eqs 9
212 and 10). When few cells are infected there is less cell death due to cytolytic immune
213 responses, a lower rate of hepatocyte proliferation to maintain the number of
214 hepatocytes, and consequently reduced loss of cccDNA via mitosis of infected cells.
215 This is of more than theoretical interest, because when estimating how long it will take
216 to deplete the cccDNA reservoir on treatment, it is the lifespan of cccDNA when
217 relatively few cells are infected that is important since treatment is known to reduce
218 the cccDNA load. The maximum expected cccDNA lifespan, corresponding to HBeAg^{POS}
219 infection, few infected cells, and no cccDNA surviving mitosis, is 123 days (71-472
220 days; Fig 2, green line). Reports for duck hepatitis B virus (DHBV) show a high

221 proportion of cccDNA survives mitosis [18]. In contrast, for HBV recent experimental
222 [9,17] and modelling [12] results suggest that relatively few cccDNA molecules survive
223 mitosis, making this longer lifespan a reasonable expectation.
224



225
226

227 **Fig 2. Probability distributions for cccDNA lifespan during different stages of HBV**
228 **infection.** The distributions for cccDNA in stable HBeAg^{POS} and HBeAg^{NEG} chronic
229 infection are based on the neutral mutation rate and rate of neutral evolution (orange
230 and blue lines, respectively). If the cccDNA burden during HBeAg^{NEG} infection is not
231 stable, but gradually falling (i.e. the basic reproduction number, R_0, is less than one)
232 the lifespan will be slightly less than inferred here. The upper estimate reflects the
233 maximum likely cccDNA lifespan when few cells are infected, based on the neutral
234 rate of evolution during HBeAg^{POS} infection and assuming no cccDNA survives mitosis
235 (q=0; green line).
236

237 ***Dynamics of the mathematical model***

238 To demonstrate the behaviour of our model we present examples of the dynamics
239 when no cccDNA survives mitosis (q=0, Fig 3; see S1 Fig for model dynamics when q=1).
240 We used parameters that are compatible with our estimated cccDNA generation times

241 (61 days during HBeAg^{POS} infection and 26 days during HBeAg^{NEG} infection). Since
242 hepatocytes are long-lived we defined the natural death rate as d=0.002 per
243 hepatocyte per day throughout and, for simplicity, we set c=0 under the assumption
244 that cytolytic responses have greater antiviral activity than non-cytolytic responses.
245 We assume a neutral mutation rate μ=2x10^{-5} s/s/c [20]. The model dynamics when
246 q=1 are similar to the case where $q=0$, apart from the lifespan of cccDNA in the early
247 stages of infection is predicted to be higher if q=1 (see below). A graphical
248 representation of the results is given in Fig 1C, and a summary of the parameters in
249 Table 2.
250
251 **Table 2: Parameters used to describe the model dynamics[1,2]**

	Model 1 No *cccDNA survives mitosis*		Model 2 *All cccDNA survives mitosis*	
	HBeAg^{POS}	HBeAg^{NEG}	HBeAg^{POS}	HBeAg^{NEG}
q	0	0	1	1
g	g_E=61	g_E=26 $^3g_{Y<<1}$=26 $^3g_{Y<<1}$=26	g_E=61	g_E=26 $^3g_{Y<<1}$=26 $^3g_{Y<<1}$=26
R_0	30.0	13.6 1.0 0.7	37.5	7.9 1.0 0.7
b	0.3	0.3 0.038 0.038	0.3	0.3 0.038 0.038

d	0.002	0.002	0.002	0.002
δ	0.006	0.018 0.034 0.050	0.014	0.036 0.036 0.052
c	0	0	0	0
μ	$2x10^{-5}$	$2x10^{-5}$	$2x10^{-5}$	$2x10^{-5}$

[1]All rates are given per day, and generation times are listed as days

[2]Where three values are given these refer to the alternative parameters used for the different trajectories presented in Fig 3 and S1 Fig (top parameter, black line; middle parameter, blue line; bottom parameter, orange line).

[3]Since R0<=1, a non-zero equilibrium is not reached.

HBeAg^{POS} infection

The replicative capacity of cccDNA, *b,* was chosen to be 0.3 per day so that the peak number of cccDNA molecules in the liver is reached at approximately 3 months since infection, in line with reported observations [28]. The per capita death rate of infected cells due to cytolytic immune responses, δ, was determined assuming a cccDNA generation time at equilibrium of 61 days, and solving Eq 9 for δ (giving δ=0.006 per day if q=0; the associated R_0 is 30).

Under these assumptions, during the first few months of infection the cccDNA burden (number of cccDNA divided by the maximum number of cccDNA) increases rapidly, leading to a short viral generation time predicted by the model of 3.3 days (Eq 11, Fig 3, S1 Fig). A recent study estimated an eclipse period of approximately 3 days for a

270 newly infected cell to produce viral particles [2], so our estimated viral generation time
271 seems reasonable. This short generation time of cccDNA during early infection
272 contrasts with the long cccDNA lifespan (dots in Fig 3, S1 Fig; note the longer lifespan
273 predicted if *q=0* compared to *q*=1). The neutral rate of evolution is also predicted to
274 be high during this early stage of infection due to the short generation time.
275
276 As infection progresses, the viral generation time increases due to fewer susceptible
277 target cells (Eq 5), in line with results in epidemiology [29], and this in turn reduces
278 the evolutionary rate (Eq 6). This dependency of evolutionary rate on epidemiological
279 dynamics has been noted in a previous simulation study on within-host viral infection
280 [30], but is generally an underappreciated factor influencing evolutionary rates. At
281 equilibrium, the estimated viral generation time and cccDNA lifespan are the same,
282 and it is this equivalency that enables us to determine these parameters from the
283 neutral rate of evolution, independent of the parameters of the model (see Methods).
284 Due to the long lifespan of infected hepatocytes, a high cccDNA burden is reached in
285 the model. This is in line with observations that most hepatocytes are infected at peak
286 infection [31].
287
288 *HBeAg^{NEG} infection*
289 We assumed the transition from HBeAg^{POS} to HBeAg^{NEG} occurs after an arbitrary
290 amount of time after HBeAg^{POS} equilibrium is reached and associates with a reduced
291 cccDNA generation time from 61 to 26 days. If this reduced generation time is not
292 accompanied by a decrease in replicative capacity, only a modest fall in the cccDNA
293 burden is predicted (Fig 3, blue line; δ=0.018 per day and *b*=0.3 per day). However,

294 this is inconsistent with *in vivo* infections, where the number of cccDNA molecules and
295 extracellular HBV DNA levels (VL) typically decline by several orders of magnitude after
296 transition to $HBeAg^{NEG}$ infection [32,33]. If the cccDNA burden is low (Y<<1), then
297 replicative capacity, b, is estimated by the reciprocal of the generation time ($b=1/g$;
298 Eq 5). For a generation time of 26 days, this gives b=0.038 per day, leading to an
299 estimated 10-fold reduction in the ability of cccDNA to reproduce during $HBeAg^{NEG}$
300 compared to $HBeAg^{POS}$ infection. In Fig 3, the orange line shows the model dynamics
301 given this decline in b, and when R_0=1 during $HBeAg^{NEG}$ infection (i.e. $\delta = 0.034$ per
302 day and b=0.038 per day). In this case, the cccDNA burden falls at a relatively modest
303 rate. Perhaps more likely is that $R_0 < 1$ and the number of cccDNA molecules continues
304 to decline. The green line shows the dynamics if R_0=0.7 (δ = 0.050 per day). However,
305 even with this modest increase in δ, the number of cccDNA is predicted to fall rapidly.
306
307 The difficulty in explaining low but steady VL using standard within-host virus models,
308 and the sensitivity of VL to model parameters when R_0 is close to one, have been
309 acknowledged previously, particularly in relation to HIV-1 infections [34–36]. Possible
310 explanations for the low numbers of cccDNA during $HBeAg^{NEG}$ infection and low rates
311 of spontaneous cure include the existence of a small number of hepatocytes that are
312 susceptible to infection, resulting in low numbers of cccDNA molecules even if R_0 is
313 high [34], or the existence of a metapopulation-type partitioned structure in the liver,
314 which enables the cccDNA to persist when R_0 is low [36].
315

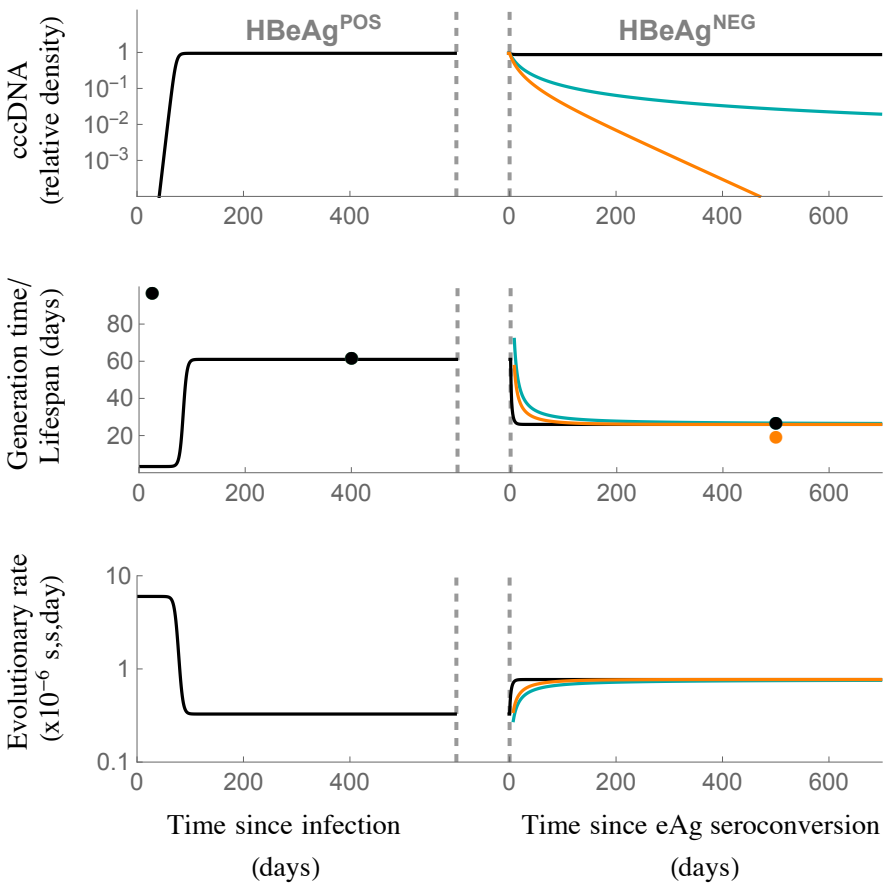


Fig 3. Population and evolutionary dynamics of cccDNA for the within-host model. We assumed no cccDNA survives mitosis (q=0; see S1 Fig for the case where q=1). The model was parameterised assuming the generation time, g, during chronic HBeAg^{POS} infection is 61 days, and during chronic HBeAg^{NEG} infection is 26 days, in line with our predictions for cccDNA generation time *in vivo*. The top panel shows cccDNA burden, where 1 represents the maximum. The middle panel shows the viral generation time (lines) and cccDNA lifespan during key stages of infection (dots, derived from Eqs 9 and 10). The bottom row shows the evolutionary rates. Black line: replicative capacity during HBeAg^{NEG} infection remains the same as during HBeAg-postive infection (b_{eAg+}= b_{eAg-}=0.3 per day). Blue line: replicative capacity falls to b_{eAg-}=0.038 per day during HBeAg^{NEG} infection, and R_0=1. Orange line: replicative capacity falls to b_{eAg-}=0.038 per day and R_0=0.7. See Table 1 for all other parameters.

330 ***Estimated time to eradicate cccDNA on treatment***

331 When few cells are infected, the inferred cccDNA lifespan is 123 days during HBeAg^{POS}
332 infection if q=0. Even with this longer estimate for cccDNA lifespan, if there are 10^{12}
333 cccDNA molecules at the start of treatment (see methods), we would expect the
334 reservoir to be depleted after less than ten years of treatment (Eq 13, Fig 4A).
335 Moreover, if treatment is initiated during HBeAg^{NEG} CHB the time to eradicate cccDNA
336 is predicted to be even faster (only 1.5 years) with a lifespan of 26 days, and a lower
337 number of cccDNA molecules ($2x10^9$) in the liver at the start of treatment. However,
338 these predictions are in stark contrast to what is observed in the clinic, where a high
339 proportion of individuals remain infected after many years of continuous treatment
340 [37] and there is no appreciable difference in treatment mediated cure in HBeAg^{NEG}
341 or HBeAg^{POS} patients [8,38]. The discrepancy may arise due to our estimated cccDNA
342 lifespan being too short. An estimated lifespan of 236 days during HBeAg-postive CHB
343 still lies within our 95% confidence interval, and would give a time to eradication, and
344 hence sterilizing cure between 18 and 36 years (Eq 13). However, this does not explain
345 the long time to eradicate cccDNA during HBeAg^{NEG} infection. Alternative explanations
346 include ongoing (albeit reduced) cccDNA amplification during NA treatment (b>0)
347 [8,39], or the presence of a long-lived subset of infected hepatocytes [17,40].

348

349 To evaluate these two scenarios, we modelled cccDNA dynamics in CHB patients on
350 treatment assuming different levels of viral replication (Fig 4B) or a subset of long-
351 lived cells (Fig 4C, S2 Fig). The dynamics of cccDNA are sensitive to the amount of
352 replication, making it unlikely that ongoing amplification alone explains the failure of
353 treatments to eliminate cccDNA. Apart from a narrow range of replicative capacities,

either a high and steady cccDNA burden, or relatively rapid cccDNA elimination, is predicted on treatment. The existence of a long-lived population of infected hepatocytes is more robust to differences in model parameters, with a gradual increase in the time to eradicate cccDNA as the death rate of long-lived cells is increased, making a long-lived population a more parsimonious explanation for the slow decline in the HBV reservoir. However, since the decay dynamics of the reservoir on treatment can be complex, and differ between individuals [39], a combination of factors most likely explains the clinical observations.

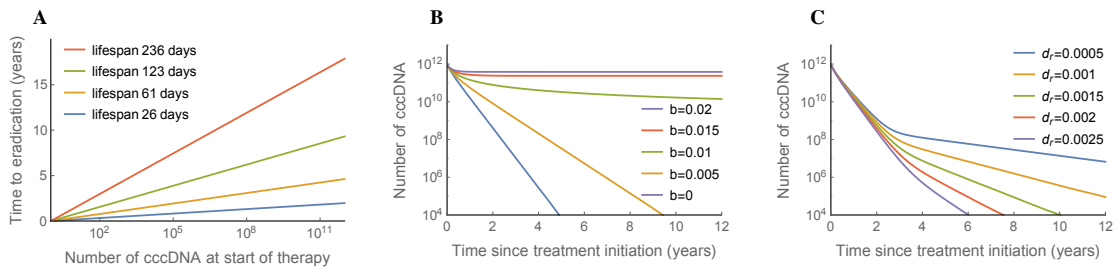


Fig 4. Effect of NA treatment predicted by the model. A: Predicted time for cccDNA to be eradicated in the absence of any cccDNA reproduction (b=0). B: cccDNA dynamics whilst on treatment, assuming some residual reproduction. For all runs d=0.002 per day, δ=0.006 per day, c=0, q=0. C: cccDNA dynamics on treatment, assuming no residual reproduction (b=0) but 0.1% cccDNA is long-lived, for different death rates of long-lived cells, d_r per day. For normal cells d=0.002 per day, δ=0.006 per day,c=0, q=0, and for long-lived cells δr=0, c_r=0, q=0. The maximum number of cccDNA was assumed to be 10^{12}, and all model runs were started at equilibrium in the absence of treatment (b=0.3 per day).

377 **DISCUSSION**

378 We provide a new model to estimate the HBV cccDNA lifespan based on reported
379 mutation and within-host evolutionary rates [20,24,26]. The lifespan of cccDNA is an
380 important component of the half-life of the cccDNA reservoir, which describes how
381 the population of cccDNA molecules in an individual declines over time. We predict an
382 average cccDNA lifespan of 61 days during HBeAg^{POS} CHB compared to only 26 days in
383 the HBeAg^{NEG} phase of infection. Although estimates for the mutation and
384 evolutionary rates for HBV are associated with high levels of uncertainty, our
385 predicted lifespan is in agreement with *in vitro* studies showing a 40 day half-life of
386 HBV cccDNA [2] and an estimated half-life of 33-57 days in woodchucks and ducks *in*
387 *vivo* [41,42]. As far as we are aware, this is the first time cccDNA lifespan has been
388 estimated during untreated infection. The lower lifespan during HBeAg^{NEG} infection is
389 consistent with a study in which VL data during therapy was fitted to a mathematical
390 model, concluding that the turnover of infected cells is higher if therapy is initiated
391 during HBeAg^{NEG} infection [16], although our predictions for cccDNA persistence are
392 longer [16] .

393

394 The shorter cccDNA lifespan during HBeAg^{NEG} CHB may reflect host immune responses,
395 with our model suggesting a doubling of the clearance rate compared to HBeAg^{POS}
396 infection. However, this increased clearance rate is predicted to have a modest effect
397 on the total number of cccDNA molecules. As well as inferring the lifespan of cccDNA,
398 we inferred cccDNA replicative capacity (a combined measure of intra and extra-
399 cellular amplification). Our results predict an approximate ten-fold reduction in
400 replicative capacity between HBeAg^{POS} and HBeAg^{NEG} phases of infection, which may

401 reflect host innate or adaptive immune responses suppressing viral production and/or
402 intra-cellular amplification. This can explain the lower cccDNA levels reported in
403 HBeAg^{NEG} CHB [32,33], and is consistent with observations that the replicative capacity
404 of cccDNA in the HBeAg^{NEG} phase of infection is reduced compared to HBeAg^{POS}
405 infection [32]. This may reflect immune control at the level of the viral epigenome, but
406 without cell death [43].
407
408 In our modelling approach, we have assumed that viral dynamics are driven by host-
409 factors, in the absence of viral phenotypic evolution. The HBeAg^{NEG} phase of infection
410 is often associated with an emergence of precore mutations that may affect HBV
411 replication and stability of cccDNA [44,45] . Viral phenotypic evolution is likely to have
412 an important role in viral dynamics, and we speculate that these viral phenotypic
413 changes could increase the rate of intra-cellular amplification, but at the expense of
414 killing the host cell, and therefore could represent a form of intra-host short-sighted
415 evolution [25].
416
417 Our estimates for cccDNA lifespan have implications for curative treatment strategies.
418 If NA therapy inhibits all cccDNA amplification, we would predict HBV to be cured after
419 1 to 10 years of continuous treatment. However, this is not observed in the clinic, with
420 only 1% of individuals clearing HBsAg each year [37]. Possible explanations for this
421 discrepancy are that NAs do not inhibit all intra- and extra-cellular amplification [8,39],
422 or the existence of long-lived infected cells [17,40]. Our model is consistent with the
423 presence of long-lived infected cells providing the most parsimonious explanation for
424 sustained infection on treatment. There is growing evidence that there is negligible

425 intra-cellular cccDNA amplification in human HBV infection [9], and since NA
426 treatment will inhibit the genesis of viral particles this will prevent extra-cellular
427 amplification. Furthermore, the dynamics of cccDNA clearance is sensitive to the
428 assumed amplification rates, and therefore if amplification alone explains the
429 dynamics we would expect to see a proportion of individuals clearing infection within
430 1-2 years of starting treatment. The presence of long-lived HBV infected cells has
431 parallels with the HIV reservoir, where long-lived latent-infected $CD4^+$ T cells prevent
432 cure [46]. Distinguishing between residual amplification and long-lived infected cells
433 will help define the expected impact of treatment strategies that prevent cccDNA
434 replication, compared to those directly targetting cccDNA. As HBV evolution will only
435 occur if there is cccDNA amplification, it may be possible to distinguish between these
436 two mechanisms by measuring the rate of cccDNA evolution whilst on treatment.
437
438 Our estimates of cccDNA persistence and amplification provide insights into
439 mechanisms underlying CHB and will inform our understanding of how spontaneous
440 or therapeutic clearance may be achieved. Given different infection profiles among
441 individuals, and limited datasets available for our model, the confidence intervals of
442 our estimations are wide. Our analysis exemplifies the power of modelling as a tool to
443 inform therapeutic interventions and highlights the need for genomic studies to
444 determine HBV evolutionary rates in CHB.
445
446
447
448

449 **METHODS**

450 To derive estimates of HBV cccDNA lifespan using the neutral mutation rate and the
451 rate of evolution we developed a deterministic mathematical model describing the
452 dynamics of cccDNA during the course of treatment naïve CHB. We used this model
453 to derive expressions for viral generation time and neutral rate of evolution, both of
454 which are predicted to change during the course of infection. Finally, we derived
455 expressions for the lifespan of cccDNA during (i) stable CHB and (ii) when the
456 proportion of infected cells is low, as would be expected in early stages of infection or
457 in the first few months of NA treatment.

458

459 ***A within-host model of HBV dynamics***

460 HBV cccDNA can replicate via intra-cellular and extra-cellular routes (Fig 1A), with a
461 reported copy number between 1-50 molecules within a single hepatocyte nucleus
462 [2,17,31,47–49] (the higher estimates tend to be for DHBV and lower estimates for
463 human HBV [9]). Since cccDNA can be lost during mitosis, we modelled the number of
464 cccDNA copies in the liver, rather than the number of infected cells. To do this, we
465 implicitly assume that viral production is proportional to the number of cccDNA
466 molecules. This is a reasonable assumption since VL has been reported to associate
467 with increasing cccDNA copy numbers [32,50].

468

469 We describe the number of copies of cccDNA in the liver at time *t* since infection, *N(t)*
470 as*:*

471

472 $\frac{dN(t)}{dt} = \beta N(t)[K - N(t)] - (d + \delta + c)N(t) - \rho(t)(1 - q)N(t)$ (Eq1a)

where the first term describes the increase in cccDNA due to intra- and extra-cellular amplification. We assume that the rate of increase is density dependent, with a maximum per capita growth rate β per day and a maximum possible number of cccDNA, K. We assume K is constant since proliferation ensures the number of hepatocytes in the liver remains stable during infection [15], and because evidence suggests there is a maximum number of copies of cccDNA that can persist within each hepatocyte, which is virally controlled [50–52]. In reality, the value of K is likely to fall during infection, for example due to sections of the liver becoming cirrhotic and/or if hepatocytes expressing HBsAg from integrated viral DNA have reduced susceptibility of infection. However, since ours is a density dependent model, this will have only a small effect on model dynamics if the change in K is much slower than viral dynamics, as is expected. Moreover, our estimates for cccDNA lifespan (below) are based on equilibrium values and are therefore independent of the value of K.

The second term describes the rate at which cccDNA is lost due to the natural death of hepatocytes and the host immune response, under the assumption that cccDNA is randomly distributed among infected hepatocytes. We assume that hepatocytes, and therefore cccDNA, have a natural per capita death rate d per day. Infected hepatocytes (and hence cccDNA) have an additional per capita death rate δ per day due to cytolytic immune responses, and cccDNA is lost at per capita rate c per day due to non-cytolytic immune responses.

496 The final term describes the loss of cccDNA due to cell proliferation. Uninfected and
497 infected hepatocytes are assumed to proliferate at the per capita rate $\rho(t)$ per day,
498 and hence cccDNA will be exposed to proliferation at rate, $\rho(t)$, with a probability *q*
499 that a cccDNA molecule will survive mitosis. Since the maximum possible number of
500 cccDNA, *K,* is assumed to be constant, proliferation and cell death are balanced,
501 hence:
502
503 $$dK + \delta N(t) = \rho(t)K \quad \text{(Eq1b)}$$
504
505 Here, the left-hand side represents the loss in carrying capacity due to the natural
506 death rate of all hepatocytes (including infected hepatocypes), and the additional
507 death rate incurred by infected hepatocytes as a direct result of infection. We are
508 making the assumption that cccDNA are randomly distributed among infected
509 hepatocytes, and that the natural death rate of hepatocytes is independent of
510 whether they are infected or not. Thus, for example, a 5% drop in the number
511 hepatoctyes results in a 5% drop in cccDNA carrying capacity. The right-hand side of
512 the equation represents consequential gain in cccDNA carrying capacity due to
513 proliferation that is needed for K to remain constant.
514
515 A complete expression for the dynamics of *N(t)* can be found by solving Eq 1b for $\rho(t)$
516 and substituting into Eq 1a. To simplify further, we consider the cccDNA burden in the
517 liver, $Y(t) = N(t)/K$, rather than the total number of cccDNA molecules, giving us:
518

519 $\frac{dY(t)}{dt} = b\,[1 - Y(t)]Y(t) - (d + \delta + c)Y(t) - (1 - q)[d + \delta Y(t)]Y(t)$

520 (Eq2)

521 where $b = \beta K$ is a rescaled measure of cccDNA replicative capacity. From this

522 equation we can calculate the basic reproductive rate of cccDNA, R_0, which is defined

523 as the number of new cccDNA molecules a single cccDNA molecule will produce in a

524 susceptible population of hepatocytes:

525

526 $R_0 = \frac{b}{d(2-q)+(\delta+c)}$ (Eq3)

527

528 If R_0<1, then the infection cannot be sustained in the long term. At equilibrium, the

529 cccDNA burden is given by:

530

531 $Y_E = Max\left[0, \frac{b-d(2-q)-(\delta+c)}{b+\delta(1-q)}\right]$ (Eq4)

532

533 which is equivalent to the cccDNA burden during stable chronic infection. Our model

534 considers the number of cccDNA molecules independent of their distribution within

535 cells. This is similar to the "single copy" modelling assumption used in [18], in which

536 only one cccDNA molecule can persist in a cell, and which was shown to produce

537 almost identical dynamics to one in which multiple copies of cccDNA are explicitly

538 modelled within infected cells [18].

539

540

541

542 ***An expression for the neutral rate of HBV evolution***

543 In a large well-mixed viral population, and in the absence of selection, the rate of
544 evolution at time *t* is given by $S(t) = \mu/g\ (t)$, where μ is the (neutral) mutation rate,
545 measured per site per viral generation, and $g(t)$ is the generation time [53]. For our
546 within-host model of HBV infection, g is equivalent to the typical amount of time it
547 takes for one cccDNA molecule to replicate another molecule. This is similar to the
548 meaning of the generation time in demography and epidemiology [29,54,55]. From Eq
549 2, we can calculate the 'backwards' generation time as:

550

551 $$g(t) = 1/[b(1 - Y(t))] \quad \text{(Eq5)}$$

552

553 At time *t* since initial infection, the neutral substitution rate is therefore given by:

554

555 $$S(t) = \frac{\mu}{g(t)} = \mu b[1 - Y(t)] \quad \text{(Eq6)}$$

556

557 Since intra- and extra-cellular amplification involve an error-prone reverse
558 transcription step (in both cases cccDNA is the template for producing pgRNA, which
559 is then reversed transcribed to form rcDNA), we have assumed they have similar
560 mutation rates. Substituting Y_E into Eq 6, we can find an expression for the neutral
561 rate of evolution rate at equilibrium:

562

563 $$S_E = \frac{\mu b[c+(2-q)(d+\delta)]}{b+\delta(1-q)} \text{ for } Y_E > 0 \quad \text{(Eq7)}$$

564

565 ***Lifespan of cccDNA during steady state infection***

566 In our model, at equilibrium the generation time of HBV will be equal to the typical
567 cccDNA lifespan, L_E. At equilibrium the number of cccDNA molecules remains
568 constant, and therefore the rate at which cccDNA is produced is equal to the rate at
569 which cccDNA is lost due to infected cell death, non-cytolytic clearance of cccDNA, and
570 proliferation of infected cells. Since the reciprocal of the production rate is equal to
571 the generation time, and the reciprocal of the rate cccDNA is lost is the typical lifespan
572 of cccDNA, at equilibrium, viral generation time and cccDNA lifespan are identical
573 ($g_E = L_E$). This relationship holds because of our assumption of constant death
574 rate and hence exponentially distributed lifetimes of cccDNA [29]; see [29,54] for
575 how this changes for different distributions.

576

577 Using the equivalence of g_E and L_E, the lifespan of cccDNA at equilibrium can be
578 determined from the mutation and neutral evolution rates by rearranging the first
579 part of Eq 6:

580

581 $$L_E = \frac{\mu}{s_E} \qquad \text{(Eq8)}$$

582

583 Substituting the expression for S_E from Eq 7 into Eq 8, we can write an expression for
584 the lifespan of cccDNA at equilibrium based on the model parameters:

585

586 $$L_E = \frac{b+\delta(1-q)}{b[c+(d+\delta)(2-q)]} \quad \text{for } Y_E > 0 \qquad \text{(Eq9)}$$

587

588 ***The lifespan of cccDNA when few cells are infected***

589 If infection increases the death rate of hepatocytes, then the level of proliferation (to
590 replace eliminated cells) will be larger the more cells are infected. Consequently, the
591 lifespan of cccDNA when few cells are infected (e.g. during early phases of infection
592 or during spontaneous clearance of infection, or after prolonged successful
593 suppressive treatment) may differ from the lifespan during $HBeAg^{POS}$ or $HBeAg^{NEG}$
594 steady state infection. By setting *Y<<1* in equation 2, we can derive an expression for
595 cccDNA lifespan when the copy number or burden is low:

596

597 $$L_{Y\ll 1} \approx \frac{1}{c+d(2-q)+\delta} \quad \text{(Eq10)}$$

598

599 Comparing the expressions for L_E and $L_{Y\ll 1}$, we can see that if all cccDNA survives
600 mitosis (*q=1*) or infection has a minimal effect on the death rate of infected cells (δ=0),
601 then cccDNA lifespan remains unchanged during infection (as long as *d* and *c* don't
602 change). However, if these conditions are not met, then the lifespan of cccDNA when
603 few cells are infected, $L_{Y\ll 1}$, can be up to double the lifespan during chronic stable
604 infection, L_E, for identical model parameters (e.g. when *q=c=d=0*, and *b>>* δ).

605

606 As we noted above, the cccDNA lifespan is only equivalent to the generation time at
607 equilibrium. Using equation 5, when few cells are infected, the generation time is
608 given by:

609

610 $$g_{Y\ll 1} \approx \frac{1}{b} \quad \text{(Eq11)}$$

611

612 This has also been observed in the epidemiological literature [29]. Combining

613 equations 3, 10 and 11 we see that:

614

615 $$R_0 = \frac{L_{Y\ll 1}}{g_{Y\ll 1}} \quad \text{(Eq12)}$$

616

617 If $R_0 > 1$ and few cells are infected (i.e. the number of cccDNA is increasing) the life

618 expectancy of cccDNA will be greater than the viral generation time, whereas if $R_0 <$

619 1 the life expectancy will be less than the viral generation time. This might be the case

620 if, for example, increased immune responses associated with HBeAg^{NEG} infection push

621 R_0 below one.

622

623 ***Estimating the generation time and lifespan of cccDNA from within-host***

624 ***evolutionary rates.***

625 During stable chronic infection, the lifespan of cccDNA, L_E, equals the viral generation

626 time, g_E, with $g_E = \mu/S_E$ (Eq 6). Although the mutation rate of HBV has not been

627 determined, for avian hepadnavirus it has been estimated at $2\text{x}10^{-5}$ s/s/c (in the range

628 $0.8\text{x}10^{-5}$ to $4.5\text{x}10^{-5}$; [20]). Since we are interested in the neutral rate of evolution, we

629 assume that a third of all mutations in non-overlapping reading frames are

630 synonymous, and that synonymous mutations are neutral or nearly neutral [56], giving

631 a neutral mutation rate of around $0.67\text{x}10^{-5}$ s/s/c ($0.3\text{x}10^{-5}$ to $1.5\text{x}10^{-5}$) in non-

632 overlapping reading frames. To incorporate the uncertainty associated with this

633 estimate, we assumed the probability of the true mutation rate is log-normally

634 distributed with mean $10^{-5.2}$ and standard deviation $10^{0.2}$.

635

636 Using longitudinal HBV sequence data, rates of evolution for non-overlapping regions
637 of the genome were generated using a relaxed clock method, inferring 16.1x10^{-8} (8.1
638 x10^{-8}, 25.5 x10^{-8}) substitutions per site per day (s/s/day) for $HBeAg^{POS}$ and 38.9 x10^{-8}
639 (27.2 x10^{-8}, 51.5 x10^{-8}) for $HBeAg^{NEG}$ chronic infection (the numbers in brackets give
640 the 5% and 95% highest posterior density (HPD) intervals; see Table 5 in [24]). In a
641 separate study, using data from [57], the synonymous rate of evolution in non-
642 overlapping genomic regions was estimated as half of the overall rate of evolution [26].
643 Assuming synonymous mutations are neutral, and that the ratio of synonymous to
644 nonsynonymous evolutionary rates is constant during infection, we therefore take the
645 neutral within-host rates of evolution during the $HBeAg^{POS}$ and $HBeAg^{NEG}$ phases of
646 infection to be half the rates of evolution reported in [24] for non-overlapping reading
647 frames. This gives a neutral rate of evolution of 8.0x10^{-8} (4.0x10^{-8}, 12.7x10^{-8}) s/s/day
648 during the $HBeAg^{POS}$ phase, and 19.5 x10^{-8} (13.6 x10^{-8} , 25.8 x10^{-8}) s/n/day during the
649 $HBeAg^{NEG}$ phase. We assumed the probability distributions of these rates are normally
650 distributed, with the standard deviation calculated using the difference between the
651 estimated rate and the 5% HPD.

652

653 We randomly sampled from each of the probability distribution functions (PDFs) for
654 the mutation rate and substitution rates and used these values to calculate the
655 generation time of cccDNA during $HBeAg^{POS}$ and $HBeAg^{NEG}$ CHB. This was repeated
656 100,000 times, from which the probability distributions for cccDNA generation time
657 during $HBeAg^{POS}$ and $HBeAg^{NEG}$ chronic infection were estimated using the built in
658 SmoothKernelDistribution function in *Mathematica* [58]. Assuming the number of

659 cccDNA rapidly reaches equilibrium during HBeAg^{POS} and HBeAg^{NEG} infection, the virus
660 generation time will provide an approximation of the cccDNA lifespan during stable
661 chronic infection (Fig 3).
662
663 ***Time to cccDNA eradication on treatment***
664 Apart from when treatment is first initiated, the number of infected cells on treatment
665 will be relatively low. Assuming eradication in our model is achieved when fewer than
666 one cccDNA molecule persists, and there is no cccDNA replication whilst on treatment,
667 the time to eradication can be approximated by:
668
669 $$t_{erad} \approx L_{Y \ll 1}\, Ln[N_{init}] \qquad \text{(Eq13)}$$
670
671 where N_{init} is the number of cccDNA when therapy is initiated and *Ln* is the natural
672 logarithm. To determine reasonable values for N_{init}, we multiplied the number of
673 hepatocytes in a human liver by the number of cccDNA per hepatocyte during
674 untreated infection. There are about 1.4x10^{8} hepatocytes per gram of human liver [59],
675 and an adult human liver is around 1.5kg, giving approximately 2x10^{11} hepatocytes in
676 total. In a recent study, an average of 6.3 copies of cccDNA per hepatocyte were found
677 during chronic HBeAg^{POS} infection, and 0.01 per hepatocyte during HBeAg^{NEG} infection
678 [32], which gives a total of approximately 1x10^{12} copies of cccDNA during HBeAg^{POS}
679 infection and 2x10^{9} copies of cccDNA during HBeAg^{NEG} infection.
680
681
682

683 ***Model assuming a subset of long-lived hepatocytes***

684 To model cccDNA dynamics in the presence of long-lived hepatocytes, we assume that
685 a constant proportion, α, of hepatocytes are long-lived, and the rest 'standard-lived',
686 with death of each cell type compensated for by proliferation of that same type. For
687 simplicity, we also assume there is no intra-cellular amplification, and with infection
688 occurring at the same per capita rate in standard- and long-lived uninfected cells.
689 Assuming the cccDNA carrying capacity of standard-lived cells is $(1-\alpha)K$, and of
690 long-lived cells is αK, the numbers of cccDNA in standard-lived hepatocytes, *N*, and in
691 long-lived hepatocytes, *L*, are given by:

692

693 $$\frac{dN(t)}{dt} = \beta[N(t)+L(t)]\left[\frac{(1-\alpha)K-N(t)}{(1-\alpha)K}\right] - (d+\delta+c)N(t) - \rho_N(t)(1-q)N(t)$$

694 (Eq14a)

695 $$\frac{dL(t)}{dt} = \beta[N(t)+L(t)]\left[\frac{\alpha K-L(t)}{\alpha K}\right] - (d_L+\delta_L+c_L)L(t) - \rho_L(t)(1-q)L(t) \quad \text{(Eq14b)}$$

696

697 where d_L, δ_L and c_L represent the per capita natural death rates, cytoloytic death rate
698 and clearance rates of cccDNA in long-lived cells, and ρ_N and ρ_L the per capita
699 proliferation rates of standard- and long-lived cells. Since deaths of standard (or long-
700 lived) cells are balanced by proliferation of standard (or long-lived) cells, we have the
701 relationships:

702

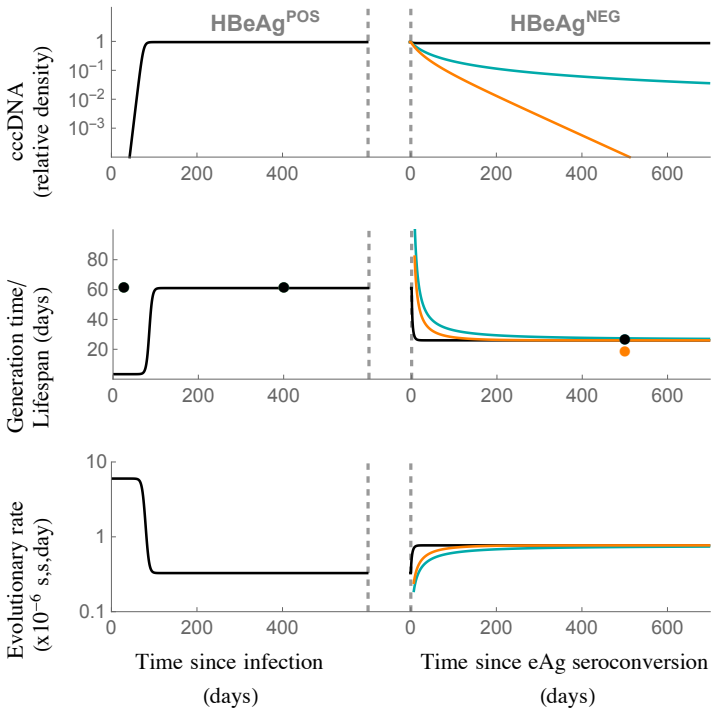
703 $$\rho_N(t)(1-\alpha)K = d(1-\alpha)K + \delta N(t) \quad \text{(Eq14c)}$$

704 $$\rho_L(t)\alpha K = d_L\alpha K + \delta_L L(t) \quad \text{(Eq14d)}$$

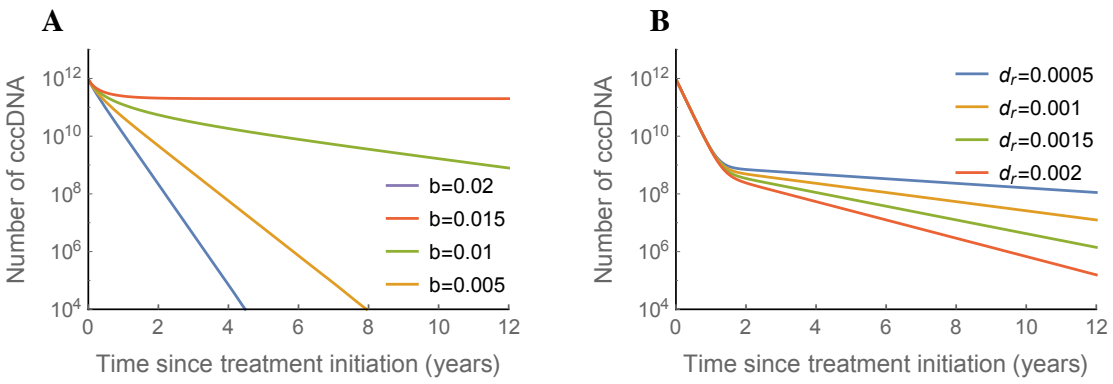
705

706 As before, we can substitute expressions for $\rho_N(t)$ and $\rho_L(t)$ found from solving Eqs
707 14c,d, into Eqs 14a,b, convert numbers into proportions, and simplify, to give
708 expressions for the burden of cccDNA in standard-lived cells, *Y(t),* and in long-lived
709 cells, *Z(t):*
710
711 $$\frac{dY(t)}{dt} = b\left[1 - \frac{Y(t)}{1-\alpha}\right][Y(t) + Z(t)] - (d + \delta + c)Y(t) - (1 - q)\left[d + \frac{\delta Y(t)}{1-\alpha}\right]Y(t)$$
712 (Eq15a)
713 $$\frac{dZ(t)}{dt} = b\left[1 - \frac{Z(t)}{\alpha}\right][Y(t) + Z(t)] - (d_L + \delta_L + c_L)Z(t) - (1 - q)\left[d + \frac{\delta_L Z(t)}{\alpha}\right]Z(t)$$
714 (Eq15b)
715
716
717 **ACKNOWLEDGEMENTS**
718 Thank you to Fabian Zoulim and Helen Fryer for their helpful comments on earlier
719 versions of this work. KAL and LP are supported by The Wellcome Trust and the
720 Royal Society grant numbers, 107652/Z/15/Z and 202562/Z/16/Z
721 JAM is supported by EU 2020 Research and Innovation Programme Consortia HEP-CAR
722 under grant agreement No.667273, MRC MR/R022011/1 and The Wellcome Trust IA
723 200838/Z/16/Z. PM is supported by The Wellcome Trust, grant number
724 110110/Z/15/Z.
725

SUPPORTING FIGURES



S1 Fig. Population and evolutionary dynamics of cccDNA for the within-host model assuming all cccDNA survives mitosis (*q*=1). The model was parameterised assuming the generation time, *g*, during chronic HBeAg^{POS} infection is 61 days, and during chronic HBeAg^{NEG} infection is 26 days, in line with our predictions for cccDNA generation time *in vivo*. The top panel shows the cccDNA burden, where 1 represents the maximum possible burden. The middle panel shows the viral generation time (lines) and cccDNA lifespan during key stages of infection (dots, derived from Eqs 9 and 10). The bottom row shows the evolutionary rates. Black line: replicative capacity during HBeAg^{NEG} infection remains the same as during HBeAg-postive infection (b_{eAg+}= b_{eAg-}=0.3 per day). Blue line: replicative capacity falls to b_{eAg-}=0.038 per day during HBeAg^{NEG} infection, and R_0=1. Orange line: replicative capacity falls to b_{eAg-}=0.038 per day and R_0=0.7. See Table 1 for all other parameters.



S2 Fig. Effect of NA treatment predicted by the model assuming all cccDNA survives mitosis (*q*=1). A: cccDNA dynamics whilst on treatment, assuming some residual reproduction. For all cells d=0.002 per day, δ=0.014 per day, c=0, q=1. B: cccDNA dynamics on treatment, assuming no residual reproduction (b=0) but 0.1% cccDNA is long-lived, for different death rates of long-lived cells, d_r per day. For normal cells d=0.002, δ=0.014, c=0, q=1, and for long-lived cells δr=0, c_r=0, q=0. The maximum number of cccDNA was assumed to be 10^{12}, and all model runs were started at equilibrium in the absence of treatment (b=0.3).

751 **REFERENCES**

752 1. WHO. Global hepatitis report. 2017.

753 2. Ko C, Chakraborty A, Chou W, Mckeating JA, Ryu W, Protzer U. Hepatitis B
754 virus genome recycling and de novo secondary infection events maintain
755 stable cccDNA levels. J Hepatol. European Association for the Study of the
756 Liver; 2018;69: 1231–1241. doi:10.1016/j.jhep.2018.08.012

757 3. Urban S, Schulze A, Dandri M, Petersen J. The replication cycle of hepatitis B
758 virus. J Hepatol. European Association for the Study of the Liver; 2010;52:
759 282–284. doi:10.1016/j.jhep.2009.10.031

760 4. Lucifora J, Protzer U. Attacking hepatitis B virus cccDNA – The holy grail to
761 hepatitis B cure. J Hepatol. European Association for the Study of the Liver;
762 2016;64: S41–S48. doi:10.1016/j.jhep.2016.02.009

763 5. Nassal M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure
764 of chronic hepatitis B. Gut. 2015;64: 1972–1984. doi:10.1136/gutjnl-2015-
765 309809

766 6. Revill PA, Chisari F V, Block JM, Dandri M, Gehring AJ, Guo H, et al. A global
767 scientific strategy to cure hepatitis B. Lancet Gastroenterol Hepatol. 2019;4:
768 545–548. doi:10.1016/S2468-1253(19)30119-0

769 7. Laporte V, Charlesworth B. Effective population size and population
770 subdivision in demographically structured populations. Genetics. 2002;162:
771 501–19.

772 8. Boyd A, Lacombe K, Lavocat F, Maylin S, Miailhes P, Lascoux-Combe C, et al.
773 Decay of ccc-DNA marks persistence of intrahepatic viral DNA synthesis under
774 tenofovir in HIV-HBV co-infected patients. J Hepatol. European Association for

775 the Study of the Liver; 2016;65: 683–691. doi:10.1016/j.jhep.2016.05.014
776 9. Tu T, Urban S. Virus entry and its inhibition to prevent and treat hepatitis B
777 and hepatitis D virus infections. Curr Opin Virol. Elsevier B.V.; 2018;30: 68–79.
778 doi:10.1016/j.coviro.2018.04.004
779 10. Locarnini S, Zoulim F. Molecular genetics of HBV infection. Antivir Ther.
780 2010;15: 3–14. doi:10.3851/IMP1619
781 11. Murray JM, Goyal A. In silico single cell dynamics of hepatitis B virus infection
782 and clearance. J Theor Biol. Elsevier; 2015;366: 91–102.
783 doi:10.1016/j.jtbi.2014.11.020
784 12. Goyal A, Ribeiro RM, Perelson AS. The Role of Infected Cell Proliferation in the
785 Clearance of Acute HBV Infection in Humans. Viruses. 2017; 350.
786 doi:10.3390/v9110350
787 13. EASL. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B
788 virus infection. J Hepatol. 2017;67: 370–398. doi:10.5152/tjg.2017.20817
789 14. Downs LO, Vawda S, Bester PA, Mcnaughton AL, Lythgoe KA, Wang T, et al.
790 Bimodal distribution and set point HBV DNA viral loads in chronic infection :
791 retrospective analysis of cohorts from the UK and South Africa [version 1 ;
792 peer review : awaiting peer review]. 2020; 1–6.
793 15. Mason WS, Gill US, Litwin S, Zhou Y, Peri S, Pop O, et al. HBV DNA Integration
794 and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered
795 Immune Tolerant. Gastroenterology. Elsevier, Inc; 2016;151: 986-998.e4.
796 doi:10.1053/j.gastro.2016.07.012
797 16. Ribeiro RM, Germanidis G, Powers KA, Pellegrin B, Nikolaidis P, Perelson AS, et
798 al. Hepatitis B Virus Kinetics under Antiviral Therapy Sheds Light on

799 Differences in Hepatitis B e Antigen Positive and Negative Infections. J Infect
800 Dis. 2010;202: 1309–1318. doi:10.1086/656528
801 17. Allweiss L, Volz T, Giersch K, Kah J, Raffa G, Petersen J, et al. Proliferation of
802 primary human hepatocytes and prevention of hepatitis B virus reinfection
803 efficiently deplete nuclear cccDNA in vivo. Hepatology. 2018;67: 542–552.
804 doi:10.1136/gutjnl-2016-312162
805 18. Reaiche-Miller GY, Thorpe M, Low HC, Qiao Q, Scougall CA, Mason WS, et al.
806 Duck hepatitis B virus covalently closed circular DNA appears to survive
807 hepatocyte mitosis in the growing liver. Virology. 2013;446: 357–364.
808 doi:10.1016/j.virol.2013.08.014
809 19. McNaughton AL, Arienzo VD, Ansari MA, Lumley SF, Littlejohn M, Revill P, et
810 al. Insights From deep sequencing of the HBV genome - Unique, tiny, and
811 misunderstood. Gastroenterology. Elsevier, Inc; 2019;156: 384–399.
812 doi:10.1053/j.gastro.2018.07.058
813 20. Pult I, Abbott N, Zhang YY, Summers J. Frequency of spontaneous mutations in
814 an avian hepadnavirus infection. J Virol. 2001;75: 9623–32.
815 doi:10.1128/JVI.75.20.9623-9632.2001
816 21. Pereira-Gómez M, Bou JV, Andreu I, Sanjuán R. Lamivudine/Adefovir
817 treatment increases the rate of spontaneous mutation of hepatitis B Virus in
818 patients. PLoS One. 2016;11: 1–12. doi:10.1371/journal.pone.0163363
819 22. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates.
820 J Virol. 2010;84: 9733–48. doi:10.1128/JVI.00694-10
821 23. Pybus OG, Belshaw R, Sanjuan R. Viral mutation and substitution: Units and
822 levels. Curr Opin Virol. 2011;1: 430–435. doi:10.1016/j.coviro.2011.08.004

823 24. Harrison A, Lemey P, Hurles M, Moyes C, Horn S, Pryor J, et al. Genomic
824 analysis of hepatitis B virus reveals antigen state and genotype as sources of
825 evolutionary rate variation. Viruses. 2011;3: 83–101. doi:10.3390/v3020083
826 25. Lythgoe KA, Gardner A, Pybus OG, Grove J. Short-Sighted Virus Evolution and a
827 Germline Hypothesis for Chronic Viral Infections. Trends Microbiol. Elsevier
828 Ltd; 2017;25: 336–348.
829 26. Vrancken B, Suchard MA, Lemey P. Accurate quantification of within- and
830 between-host HBV evolutionary rates requires explicit transmission chain
831 modelling. Virus Evol. 2017;3: 1–9. doi:10.1093/ve/vex028
832 27. Lanfear R, Kokko H, Eyre-walker A. Population size and the rate of evolution.
833 Trends Ecol Evol. Elsevier Ltd; 2014;29: 33–41. doi:10.1016/j.tree.2013.09.009
834 28. Whalley BSA, Murray JM, Brown D, Webster GJM, Emery VC, Dusheiko GM, et
835 al. Kinetics of Acute Hepatitis B Virus Infection in Humans. J Exp Med.
836 2001;193: 847–854.
837 29. Scalia G, Svensson Å, Asikainen T, Giesecke J. Mathematical Biosciences Some
838 model based considerations on observing generation times for communicable
839 diseases. 2010;223: 24–31. doi:10.1016/j.mbs.2009.10.004
840 30. Scholle SO, Ypma RJF, Lloyd AL, Koelle K. Viral substitution rate variation can
841 arise from the interplay between within-host and epidemiological dynamics.
842 Am Nat. 2013;182: 494–513. doi:10.1086/672000
843 31. Kajino K, Jilbert AR, Saputelli J, Aldrich CE, Cullen J, Mason WS. Woodchuck
844 hepatitis virus infections: very rapid recovery after a prolonged viremia and
845 infection of virtually every hepatocyte. J Virol. 1994;68: 5792–803.
846 32. Testoni B, Lebossé F, Scholtes C, Berby F, Miaglia C, Subic M, et al. Serum

847 hepatitis B core-related antigen (HBcrAg) correlates with covalently closed
848 circular DNA transcriptional activity in chronic hepatitis B patients. J Hepatol.
849 European Association for the Study of the Liver; 2019;70: 615–625.
850 doi:10.1016/j.jhep.2018.11.030
851 33. Downs LO, Vawda S, Bester PA, Lythgoe KA, Wang T, McNaughton AL, et al.
852 Bimodal distribution and set point HBV DNA viral loads in chronic infection :
853 retrospective analysis of cohorts from the UK and South Africa. medRxiv.
854 2020; doi:http://dx.doi.org/10.1101/2020.01.08.20016972
855 34. Bonhoeffer S, Funk GA, Gunthard HF, Fischer M, Muller V. Glancing behind
856 virus load variation in HIV-1 infection. Trends Microbiol. 2003;11: 499–504.
857 doi:10.1016/j.tim.2003.09.002
858 35. Müller V, Marée AFM, De Boer RJ. Small variations in multiple parameters
859 account for wide variations in HIV-1 set-points: a novel modelling approach.
860 Proc R Soc B. 2001;268: 235–42. doi:10.1098/rspb.2000.1358
861 36. Lythgoe KA, Pellis L, Fraser C. Large Variations in HIV-1 Viral Load Explained by
862 Shifting-Mosaic Metapopulation Dynamics. PLoS Biol. 2016; 1002567.
863 doi:10.1371/journal.pbio.1002567
864 37. Downs LO, Smith DA, Lumley SF, Patel M, McNaughton AL, Mokaya J, et al.
865 Electronic health informatics data to describe clearance dynamics of hepatitis
866 B surface antigen (HBsAg) and e antigen (HBeAg) in chronic hepatitis B virus
867 infection. MBio. 2019;10: 1–15. doi:10.1128/mBio.00699-19
868 38. Wong DKH, Seto WK, Fung J, Ip P, Huang FY, Lai CL, et al. Reduction of
869 hepatitis B surface antigen and covalently closed circular DNA by nucleos(t)ide
870 analogues of different potency. Clin Gastroenterol Hepatol. 2013;11: 1004-

871 1010.e1. doi:10.1016/j.cgh.2013.01.026

872 39. Dahari H, Shudo E, Ribeiro RM, Perelson AS. Modeling complex decay profiles
873 of hepatitis B virus during antiviral therapy. Hepatology. 2009;49: 32–38.
874 doi:10.1002/hep.22586

875 40. De Crignis E, Carofiglio F, Moulos P, Verstegen MMA, Romal S, Khalid MM, et
876 al. Human liver organoids; a patient-derived primary model for HBV infection
877 and related hepatocellular carcinoma. bioRxiv. 2019; 568147.
878 doi:10.1101/568147

879 41. Addison WR, Walters K-A, Wong WWS, Wilson JS, Madej D, Jewell LD, et al.
880 Half-life of the duck hepatitis B virus covalently closed circular DNA pool in
881 vivo following inhibition of viral replication. J Virol. 2002;76: 6356–6363.
882 doi:10.1128/JVI.76.12.6356-6363.2002

883 42. Zhu Y, Yamamoto T, Cullen J, Saputelli J, Aldrich CE, Miller DS, et al. Kinetics of
884 Hepadnavirus Loss from the Liver during Inhibition of Viral DNA Synthesis. J
885 Virol. 2001;75: 311–322. doi:10.1128/jvi.75.1.311-322.2001

886 43. Volz T, Lutgehetmann M, Wachtler P, Jacob A, Quaas A, Murray JM, et al.
887 Impaired Intrahepatic Hepatitis B Virus Productivity Contributes to Low
888 Viremia in Most HBeAg-Negative Patients. Gastroenterology. 2007;133: 843–
889 852. doi:10.1053/j.gastro.2007.06.057

890 44. Mason WS, Litwin S, Jilbert AR. Immune selection during chronic hepadnavirus
891 infection. Hepatol Int. 2008;2: 3–16. doi:10.1007/s12072-007-9024-3

892 45. Kao JH, Chen PJ, Lai MY, Chen DS. Basal core promoter mutations of hepatitis
893 B virus increase the risk of hepatocellular carcinoma in hepatitis B carriers.
894 Gastroenterology. 2003;124: 327–334. doi:10.1053/gast.2003.50053

895 46. Palmer S, Maldarelli F, Wiegand A, Bernstein B, Hanna GJ, Brun SC, et al. Low-
896 level viremia persists for at least 7 years in patients on suppressive
897 antiretroviral therapy. Proc Natl Acad Sci United States Am. 2008;105: 3879–
898 3884.
899 47. Zhang Y, Zhang B, Theele D, Litwin S, Toll E, Summers J. Single-cell analysis of
900 covalently closed circular DNA copy numbers in a hepadnavirus-infected liver.
901 Proc Natl Acad Sci United States Am. 2003;100: 12372–12377.
902 48. Li M, Sohn JA, Seeger C. Distribution of Hepatitis B Virus Nuclear DNA. J Virol.
903 2017;92: 1–13. doi:10.1128/jvi.01391-17
904 49. Jilbert AR, Wu TT, England JM, Hall PM, Carp NZ, O'Connell AP, et al. Rapid
905 resolution of duck hepatitis B virus infections occurs after massive
906 hepatocellular involvement. J Virol. 1992;66: 1377–1388.
907 50. Lenhoff RJ, Summers J. Construction of avian hepadnavirus variants with
908 enhanced replication and cytopathicity in primary hepatocytes. J Virol.
909 1994;68: 5706–5713.
910 51. Summers J, Smith PM, Horwich AL. Hepadnavirus Envelope Proteins Regulate
911 Covalently Closed Circular DNA Amplification. J Virol. 1990;64: 2819–2824.
912 52. Summers J, Smith PM, Huang MJ, Yu MS. Morphogenetic and regulatory
913 effects of mutations in the envelope proteins of an avian hepadnavirus. J Virol.
914 1991;65: 1310–7.
915 53. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge
916 University Press; 1985.
917 54. Svensson Å. A note on generation times in epidemic models ˚. Mathematic.
918 2007;208: 300–311. doi:10.1016/j.mbs.2006.10.010

919 55. Nishiura H. Time variations in the generation time of an infectious disease:
920 Implications for sampling to appropriately quantify transmission potential.
921 Math Biosci Eng. 2010;7: 851–869. doi:10.3934/mbe.2010.7.851
922 56. Cuevas JM, Domingo-Calap P, Sanjuán R. The Fitness Effects of Synonymous
923 Mutations in DNA and RNA Viruses. Mol Biol Evol. 2011;29: 17–20.
924 doi:10.1093/molbev/msr179
925 57. Lin Y-Y, Liu C, Chien W-H, Wu L-L, Tao Y, Wu D, et al. New Insights into the
926 Evolutionary Rate of Hepatitis B Virus at Different Biological Scales. J Virol.
927 2015;89: 3512–3522. doi:10.1128/JVI.03131-14
928 58. Mathematica Version 10.3. Champaign, Illinois: Wolfram Research, Inc; 2015.
929 59. Sohlenius-Sternbeck AK. Determination of the hepatocellularity number for
930 human, dog, rabbit, rat and mouse livers from protein concentration
931 measurements. Toxicol Vitr. 2006;20: 1582–1586.
932 doi:10.1016/j.tiv.2006.06.003
933