

# Estimating hepatitis B virus cccDNA persistence in chronic infection<sup>†</sup>

Katrina A. Lythgoe,<sup>1,2,\*</sup> Sheila F. Lumley,<sup>3,4</sup> Lorenzo Pellis,<sup>5</sup>  
Jane A. McKeating,<sup>6</sup> and Philippa C. Matthews<sup>3,4,7</sup>

<sup>1</sup>Big Data Institute, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK, <sup>2</sup>Department of Zoology, University of Oxford, Medawar Building, South Parks Road, Oxford OX1 3SY, UK, <sup>3</sup>Nuffield Department of Medicine, University of Oxford, Medawar Building, South Parks Road, Oxford OX1 3SY, UK, <sup>4</sup>Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK, <sup>5</sup>Department of Mathematics, Alan Turing Building, Oxford Rd, Manchester M13 9PL, UK, <sup>6</sup>Nuffield Department of Medicine Research Building, University of Oxford, Oxford OX3 7LF, UK and <sup>7</sup>NIHR Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK

\*Corresponding author: E-mail: katrina.lythgoe@bdi.ox.ac.uk  
Jane A. McKeating and Philippa C. Matthews contributed equally to this study

## 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 (NAs) 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<sup>POS</sup>) infection the cccDNA lifespan is 61 (36–236) days, whereas during the HBeAg<sup>NEG</sup> phase of infection it is only 26 (16–81) days. We found that cccDNA replicative capacity declined by an order of magnitude between HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> 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.

**Key words:** hepatitis B virus; cccDNA; persistence; evolution; dynamics; reservoir; modelling.

<sup>†</sup>Special Issue Santa Fe Institute Workshop on Integrating Critical Phenomena and Multi-Scale Selection in Virus Evolution, supported by the NSF Rules of Life Program grant DEB-1830688

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Author summary**

Nearly 1 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 covalently closed circular 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.

**1. Introduction**

Hepatitis B Virus (HBV) is a global health problem, with more than 240 million chronically infected individuals at risk of developing liver fibrosis, cirrhosis, and hepatocellular carcinoma (WHO 2017). HBV is the prototypic member of the *hepadnaviruses*, a family of small, enveloped hepatotropic viruses with a partially double-stranded relaxed circular DNA (rcDNA) genome that replicates via reverse transcription of pregenomic RNA (pgRNA). Following the infection of hepatocytes, the rcDNA genome is imported to the nucleus and converted to covalently closed circular DNA (cccDNA), that provides the transcriptional template for pregenomic and subgenomic RNAs. Newly synthesised pgRNA is assembled into nucleocapsids that undergo reverse transcription to generate rcDNA, which is subsequently enveloped and released as infectious virions. Alternatively, capsids can be redirected to the nucleus to replenish and maintain the episomal pool of cccDNA and this intracellular amplification pathway, together with the long half-life of cccDNA, contributes to viral persistence (Urban et al. 2010; Ko et al. 2018).

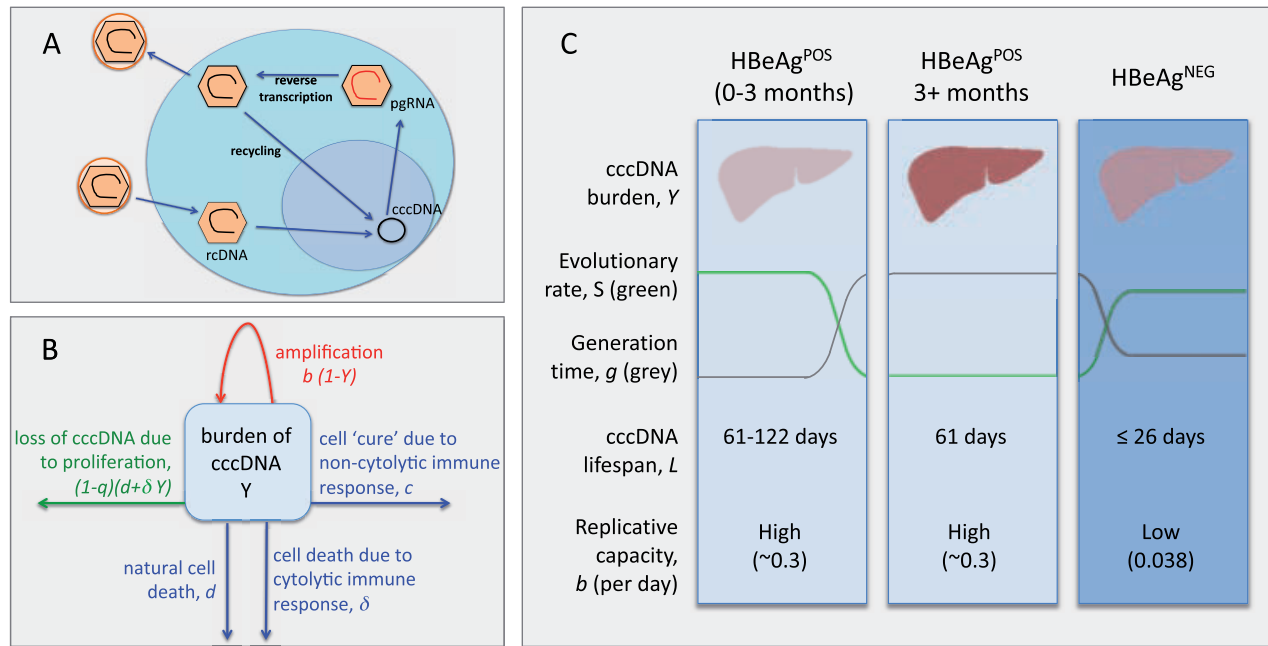
Chronic hepatitis B (CHB) is usually treated with nucleos(t)ide analogues (NAs) that inhibit the reverse transcription of pgRNA to rcDNA. However, these therapies do not directly target the cccDNA reservoir (Lucifora and Protzer 2016) and viremia rebounds once treatment is stopped, even when peripheral levels of viral DNA have remained undetectable for months or years (Nassal 2015). There is a growing impetus to identify curative therapies for HBV (Revell et al. 2019). Despite its central role in the HBV life cycle, our understanding of the viral and host factors that regulate cccDNA abundance and half-life is limited (Laporte and Charlesworth 2002). cccDNA half-life can be defined as the time for the number of copies in the liver to reduce by half and will depend on a number of factors, including cccDNA 'lifespan' (the time an individual cccDNA molecule persists) and the genesis of new cccDNA via extra-cellular virus infection or intra-cellular amplification (Boyd et al. 2016; Tu and Urban 2018) (Fig. 1A). A recent *in vitro* study reported a 40-day half-life of HBV cccDNA in infected HepG2-NTCP cells (Ko et al. 2018), with an estimated lifespan of 58 days. However, the cccDNA lifespan in the human liver is unknown.

The lifespan of cccDNA molecules most likely changes over the course of CHB, influenced by host and viral factors (Locarnini and Zoulim 2010), including the rate of hepatocyte proliferation (Murray and Goyal 2015; Goyal et al. 2017). The natural history of CHB is often classified by four clinical or virological phases of infection and/or hepatitis (according to European Association for the Study of the Liver 2017 (EASL 2017)). However, it is increasingly recognised that this description fails to capture the diversity of viral replication in the liver and misleadingly suggests a linear progression over time from one phase to the next. To avoid difficulties associated with this rigid phenotypic structure of disease, and to focus on viral dynamics rather than disease pathology, we instead consider two states: HBe-antigen positive (HBeAg<sup>POS</sup>) that is associated with high

peripheral HBV DNA levels (viral load—VL), and HBeAg<sup>NEG</sup>, which is associated with lower VLs (Downs et al. 2020b). Early in infection cccDNA is transcriptionally active and translation of pre-core/pgRNA results in detectable levels of hepatitis B e antigen (HBeAg) in the periphery, associated with high VL (Mason et al. 2016). In later stages of infection after HBeAg seroconversion there is a loss of HBeAg that associates with more efficient immune targeting of infected cells (Ribeiro et al. 2010), leading to a reduction in VL and a shortening of cccDNA lifespan. The higher hepatocyte death rates during HBeAg<sup>NEG</sup> CHB infection will induce hepatocyte proliferation (Mason et al. 2016). The extent to which cccDNA is lost during hepatocyte mitosis is uncertain (Tu and Urban 2018), but unless all cccDNA episomes survive mitosis, the increased proliferation rate of infected cells will shorten the average lifespan of cccDNA (Reaiche-Miller et al. 2013; Goyal et al. 2017; Allweiss et al. 2018).

Double-stranded DNA viruses typically have low mutation rates, but since rcDNA is generated via error-prone reverse transcription in the *hepadnaviridae*, they have higher mutation rates than other DNA viruses (Fig. 1A) (McNaughton et al. 2019). The estimated mutation rate for avian hepadnavirus is  $2 \times 10^{-5}$  substitutions per site per infected cell (s/s/c) (Pult et al. 2001), and an upper limit of  $8.7 \times 10^{-5}$  s/s/c has been estimated for HBV (Pereira-Gómez et al. 2016), similar to estimates for RNA retroviruses (Sanjuán et al. 2010). The evolutionary rate measures how quickly mutations become fixed in a population over a period of time (Pybus et al. 2011). Strikingly, the evolutionary rate of HBV is much lower than for RNA viruses with similar mutation rates (Harrison et al. 2011; Lythgoe et al. 2017; Vrancken et al. 2017). Different mechanisms could explain this observation, including the biological constraint of multiple overlapping reading frames in the HBV genome (Harrison et al. 2011), limited viral population size in the liver, or long cccDNA lifespan (Harrison et al. 2011; Lythgoe et al. 2017). Both evolutionary constraint and population size should only influence the rate of evolution of variants that experience selection (Lanfear et al. 2014). For neutral or near neutral mutations, the long cccDNA lifespan provides the simplest explanation for the low evolutionary rate of HBV (Lythgoe et al. 2017).

We propose the within-host evolutionary rate of HBV can be used to estimate cccDNA lifespan and the replicative capacity (a combined measure of extra-cellular viral production and intra-cellular amplification) of the virus. We developed a novel mathematical model that uses published HBV mutation and evolutionary rates (Pult et al. 2001; Harrison et al. 2011; Vrancken et al. 2017), to infer the lifespan and replicative capacity of cccDNA during different phases of CHB. Although we use our model to consider the neutral rate of evolution, it is essentially an ecological model, in that we do not assume any viral phenotypic evolution. To the best of our knowledge, these are the first estimates of cccDNA lifespan in treatment naive subjects and provide important insights into the HBV reservoir that



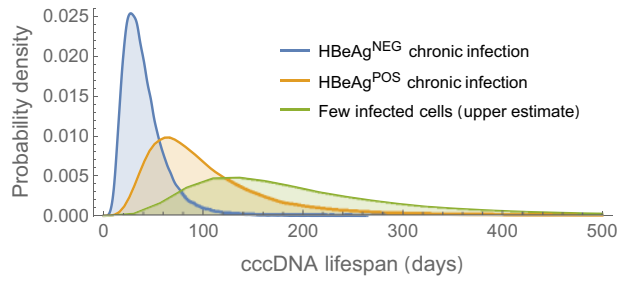
<sup>a</sup>Throughout we use days as the time unit and all rates are per day unless stated otherwise.

will be valuable for the design and evaluation of future treatment interventions.

## 2. Results

We developed a mathematical model describing the number of cccDNA molecules in the liver that is independent of infected

cell frequency, and accounts for intra- and extra-cellular cccDNA amplification and loss of cccDNA during hepatocyte mitosis (Fig. 1B and Section 4, Equations (1) and (2)). Using this model we derived expressions for the viral generation time, defined as the typical time for one cccDNA molecule to generate another cccDNA molecule at time  $t$  since infection,  $g(t)$  (Equation (5)), and the neutral rate of evolution at time  $S(t)$



**Figure 2.** Probability distributions for cccDNA lifespan during different stages of HBV infection. The distributions for cccDNA in stable HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> chronic infection are based on the neutral mutation rate and rate of neutral evolution (orange and blue lines, respectively). If the cccDNA burden during HBeAg<sup>NEG</sup> infection is not stable, but gradually falling (i.e. the basic reproduction number,  $R_0$ , is less than one) the lifespan will be slightly less than inferred here. The upper estimate reflects the maximum likely cccDNA lifespan when few cells are infected, based on the neutral rate of evolution during HBeAg<sup>POS</sup> infection and assuming no cccDNA survives mitosis ( $q = 0$ ; green line).

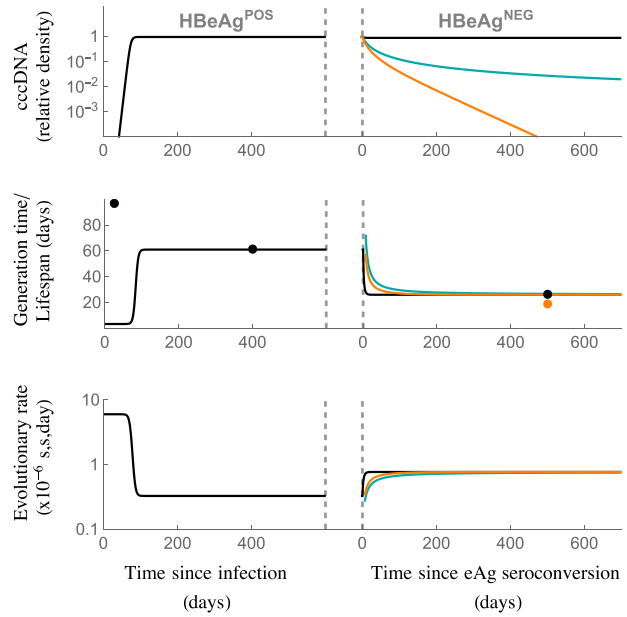
(Equation (6)). At equilibrium, we show that the lifespan of cccDNA,  $L_E$ , is equal to the virus generation time,  $g_E$ , which is given by the neutral mutation rate divided by the neutral rate of evolution,  $\mu/S_E$  (Equation (8)). The notation used throughout is given in Table 1.

### 2.1 Lifespan of cccDNA

From the published estimates for the mutation (Pult et al. 2001) and evolutionary rates of HBV (Harrison et al. 2011; Vrancken et al. 2017), we inferred the probability distributions for cccDNA lifespan during HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> phases of CHB using Equation (8). The predicted lifespan of cccDNA during HBeAg<sup>POS</sup> 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<sup>NEG</sup> infection the lifespan of cccDNA is estimated at only 26 days (16–81 days; Fig. 2, blue line).

The shorter lifespan of cccDNA during HBeAg<sup>NEG</sup> compared to HBeAg<sup>POS</sup> infection can be explained by higher rates of cccDNA clearance (Equation (9)). This may reflect changes in the immune environment due to HBe-antigen seroconversion that is associated with increased cytolytic and non-cytolytic immune responses ( $\delta$  and  $c$ , respectively). Increased host immune responses during HBeAg<sup>NEG</sup> infection could push the basic reproduction number,  $R_0$ , of cccDNA below one (Equation (3)) due to the higher clearance rates of cccDNA molecules, and also due to reduced replicative capacity,  $b$ , of cccDNA. If  $R_0 < 1$ , the number of cccDNA will not reach a stable level but will continually decline. In this non-equilibrium situation the lifespan of cccDNA may be less than our inferred 26 days since the viral generation time will be greater than the lifespan of cccDNA (Fig. 3 and Section 4).

Our model suggests cccDNA lifespan can be up to two times longer when few cells are infected compared to when most cells are infected cells (see Section 4; compare Equations (9) and (10)). When few cells are infected there is less cell death due to cytolytic immune responses, a lower rate of hepatocyte proliferation to maintain the number of hepatocytes, and consequently reduced loss of cccDNA via mitosis of infected cells. This is of more than theoretical interest, because when estimating how long it will take to deplete the cccDNA reservoir on treatment, it is the lifespan of cccDNA when relatively few cells are infected that is important since treatment is known to reduce the cccDNA load. The maximum expected cccDNA lifespan,



**Figure 3.** Population and evolutionary dynamics of cccDNA for the within-host model. We assumed no cccDNA survives mitosis ( $q = 0$ ; see Supplementary Fig. S1 for the case where  $q = 1$ ). The model was parameterised assuming the generation time,  $g$ , during chronic HBeAg<sup>POS</sup> infection is 61 days, and during chronic HBeAg<sup>NEG</sup> infection is 26 days, in line with our predictions for cccDNA generation time *in vivo*. The top panel shows cccDNA burden (relative density), 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 Equations (9) and (10)). The bottom row shows the evolutionary rates. Black line: replicative capacity during HBeAg<sup>NEG</sup> infection remains the same as during HBeAg-positive infection ( $b_{eAg+} = b_{eAg-} = 0.3$  per day). Blue line: replicative capacity falls to  $b_{eAg-} = 0.038$  per day during HBeAg<sup>NEG</sup> infection, and  $R_0 = 1$ . Orange line: replicative capacity falls to  $b_{eAg-} = 0.038$  per day and  $R_0 = 0.7$ . See Table 2 for all other parameters.

**Table 2.** Parameters used to describe the model dynamics.<sup>a,b,c</sup>

	Model 1		Model 2	
	No cccDNA survives mitosis		All cccDNA survives mitosis	
	HBeAg <sup>POS</sup>	HBeAg <sup>NEG</sup>	HBeAg <sup>POS</sup>	HBeAg <sup>NEG</sup>
$q$	0	0	1	1
$g$	$g_E = 61$	$g_E = 26$ ${}^c g_{Y \ll 1} = 26$ ${}^c g_{Y \ll 1} = 26$	$g_E = 61$	$g_E = 26$ ${}^c g_{Y \ll 1} = 26$ ${}^c g_{Y \ll 1} = 26$
$R_0$	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
$\delta$	0.006	0.018 0.034 0.050	0.014	0.036 0.036 0.052
$c$	0	0	0	0
$\mu$	$2 \times 10^{-5}$	$2 \times 10^{-5}$	$2 \times 10^{-5}$	$2 \times 10^{-5}$

<sup>a</sup>All rates are given per day, and generation times are listed as days.

<sup>b</sup>Where three values are given these refer to the alternative parameters used for the different trajectories presented in Fig. 3 and Supplementary Fig. S1 (top parameter, black line; middle parameter, blue line; bottom parameter, orange line).

<sup>c</sup>Since  $R_0 < 1$ , a non-zero equilibrium is not reached.

corresponding to HBeAg<sup>POS</sup> infection, few infected cells, and no cccDNA surviving mitosis, is 123 days (71–472 days; Fig. 2, green line). Reports for duck hepatitis B virus (DHBV) show a high proportion of cccDNA survives mitosis (Reaiche-Miller et al. 2013). In contrast, for HBV recent experimental (Allweiss et al. 2018; Tu and Urban 2018) and modelling (Goyal et al. 2017) results suggest that relatively few cccDNA molecules survive mitosis, making this longer lifespan a reasonable expectation.

## 2.2 Dynamics of the mathematical model

To demonstrate the behaviour of our model we present examples of the dynamics when no cccDNA survives mitosis ( $q=0$ , Fig. 3; see Supplementary Fig. S1 for model dynamics when  $q=1$ ). We used parameters that are compatible with our estimated cccDNA generation times (61 days during HBeAg<sup>POS</sup> infection and 26 days during HBeAg<sup>NEG</sup> infection). Since hepatocytes are long-lived we defined the natural death rate as  $d=0.002$  per hepatocyte per day throughout and, for simplicity, we set  $c=0$  under the assumption that cytolytic responses have greater antiviral activity than non-cytolytic responses. We assume a neutral mutation rate  $\mu = 2 \times 10^{-5}$  s/s/c (Pult et al. 2001). The model dynamics when  $q=1$  are similar to the case where  $q=0$ , apart from the lifespan of cccDNA in the early stages of infection is predicted to be higher if  $q=1$  (see below). A graphical representation of the results is given in Fig. 1C, and a summary of the parameters in Table 2.

## 2.3 HBeAg<sup>POS</sup> 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 (Whalley et al. 2001). The per capita death rate of infected cells due to cytolytic immune responses,  $\delta$ , was determined assuming a cccDNA generation time at equilibrium of 61 days, and solving Equation (9) for  $\delta$  (giving  $\delta=0.006$  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 (Equation (11), Fig. 3, and Supplementary Fig. S1). A recent study estimated an eclipse period of approximately 3 days for a newly infected cell to produce viral particles (Ko et al. 2018), so our estimated viral generation time seems reasonable. This short generation time of cccDNA during early infection contrasts with the long cccDNA lifespan (dots in Fig. 3, Supplementary Fig. S1; note the longer lifespan predicted if  $q=0$  compared to  $q=1$ ). The neutral rate of evolution is also predicted to be high during this early stage of infection due to the short generation time.

As infection progresses, the viral generation time increases due to fewer susceptible target cells (Equation (5)), in line with results in epidemiology (Scalia et al. 2010), and this in turn reduces the evolutionary rate (Equation (6)). This dependency of evolutionary rate on epidemiological dynamics has been noted in a previous simulation study on within-host viral infection (Scholle et al. 2013), but is generally an underappreciated factor influencing evolutionary rates. At equilibrium, the estimated viral generation time and cccDNA lifespan are the same, and it is this equivalency that enables us to determine these parameters from the neutral rate of evolution, independent of the parameters of the model (see Section 4). Due to the long lifespan of

infected hepatocytes, a high cccDNA burden is reached in the model. This is in line with observations that most hepatocytes are infected at peak infection (Kajino et al. 1994).

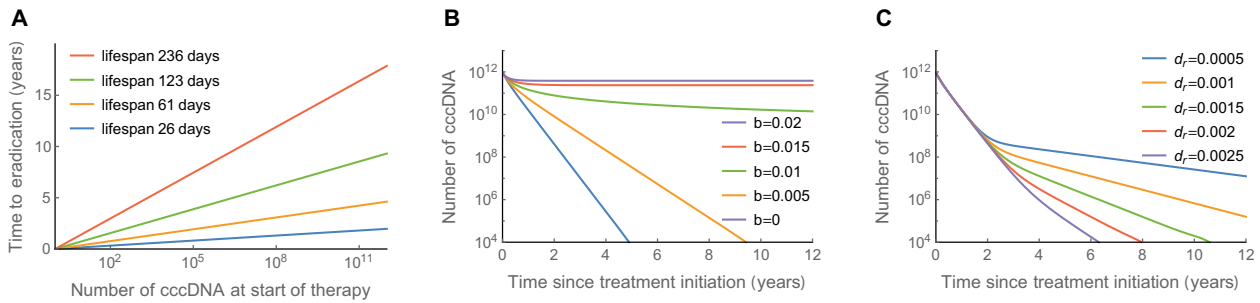
## 2.4 HBeAg<sup>NEG</sup> infection

We assumed the transition from HBeAg<sup>POS</sup> to HBeAg<sup>NEG</sup> occurs after an arbitrary amount of time after HBeAg<sup>POS</sup> equilibrium is reached and associates with a reduced cccDNA generation time from 61 to 26 days. If this reduced generation time is not accompanied by a decrease in replicative capacity, only a modest fall in the cccDNA burden is predicted (Fig. 3, blue line;  $\delta = 0.018$  per day and  $b=0.3$  per day). However, this is inconsistent with *in vivo* infections, where the number of cccDNA molecules and extracellular HBV DNA levels (VL) typically decline by several orders of magnitude after transition to HBeAg<sup>NEG</sup> infection (Testoni et al. 2019; Downs et al. 2020). If the cccDNA burden is low ( $Y \ll 1$ ), then replicative capacity,  $b$ , is estimated by the reciprocal of the generation time ( $b=1/g$ ; Equation (5)). For a generation time of 26 days, this gives  $b=0.038$  per day, leading to an estimated ten-fold reduction in the ability of cccDNA to reproduce during HBeAg<sup>NEG</sup> compared to HBeAg<sup>POS</sup> infection. In Fig. 3, the orange line shows the model dynamics given this decline in  $b$ , and when  $R_0 = 1$  during HBeAg<sup>NEG</sup> infection (i.e.  $\delta = 0.034$  per day and  $b=0.038$  per day). In this case, the cccDNA burden falls at a relatively modest rate. Perhaps more likely is that  $R_0 < 1$  and the number of cccDNA molecules continues to decline. The green line shows the dynamics if  $R_0 = 0.7$  ( $\delta = 0.050$  per day). However, even with this modest increase in  $\delta$ , the number of cccDNA is predicted to fall rapidly.

The difficulty in explaining low but steady VL using standard within-host virus models, and the sensitivity of VL to model parameters when  $R_0$  is close to one, have been acknowledged previously, particularly in relation to HIV-1 infections (Müller et al. 2001; Bonhoeffer et al. 2003; Lythgoe et al. 2016). Possible explanations for the low numbers of cccDNA during HBeAg<sup>NEG</sup> infection and low rates of spontaneous cure include the existence of a small number of hepatocytes that are susceptible to infection, resulting in low numbers of cccDNA molecules even if  $R_0$  is high (Bonhoeffer et al. 2003), or the existence of a metapopulation-type partitioned structure in the liver, which enables the cccDNA to persist when  $R_0$  is low (Lythgoe et al. 2016).

## 2.5 Estimated time to eradicate cccDNA on treatment

When few cells are infected, the inferred cccDNA lifespan is 123 days during HBeAg<sup>POS</sup> infection if  $q=0$ . Even with this longer estimate for cccDNA lifespan, if there are  $10^{12}$  cccDNA molecules at the start of treatment (see Section 4), we would expect the reservoir to be depleted after less than 10 years of treatment (Equation (13), Fig. 4A). Moreover, if treatment is initiated during HBeAg<sup>NEG</sup> CHB the time to eradicate cccDNA is predicted to be even faster (only 1.5 years) with a lifespan of 26 days, and a lower number of cccDNA molecules ( $2 \times 10^9$ ) in the liver at the start of treatment. However, these predictions are in stark contrast to what is observed in the clinic, where a high proportion of individuals remain infected after many years of continuous treatment (Downs et al. 2019) and there is no appreciable difference in treatment mediated cure in HBeAg<sup>NEG</sup> or HBeAg<sup>POS</sup> patients (Wong et al. 2013; Boyd et al. 2016). The discrepancy may arise due to our estimated cccDNA lifespan being too short. An estimated lifespan of 236 days during HBeAg-positive CHB still lies within our 95 per cent confidence interval, and would



**Figure 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,  $\delta=0.006$  per day,  $c=0$ ,  $q=0$ . (C) cccDNA dynamics on treatment, assuming no residual reproduction ( $b=0$ ) but 0.1 per cent cccDNA is long-lived, for different death rates of long-lived cells,  $d_l$ , per day. For normal cells  $d=0.002$  per day,  $\delta=0.006$  per day,  $c=0$ ,  $q=0$ , and for long-lived cells  $\delta_r=0$ ,  $c_r=0$ ,  $q=0$ . 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).

give a time to eradication, and hence sterilising cure between 18 and 36 years (Equation (13)). However, this does not explain the long time to eradicate cccDNA during HBeAg<sup>NEG</sup> infection. Alternative explanations include ongoing (albeit reduced) cccDNA amplification during NA treatment ( $b > 0$ ) (Dahari et al. 2009; Boyd et al. 2016), or the presence of a long-lived subset of infected hepatocytes (Allweiss et al. 2018; De Crignis et al. 2019).

To evaluate these two scenarios, we modelled cccDNA dynamics in CHB patients on treatment assuming different levels of viral replication (Fig. 4B) or a subset of long-lived cells (Fig. 4C and Supplementary Fig. S2). The dynamics of cccDNA are sensitive to the amount of replication, making it unlikely that ongoing amplification alone explains the failure of 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 refractory to immune-killing 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 increased. This makes 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 (Dahari et al. 2009), a combination of factors most likely explains the clinical observations.

### 3. Discussion

We provide a new model to estimate the HBV cccDNA lifespan based on reported mutation and within-host evolutionary rates (Pult et al. 2001; Harrison et al. 2011; Vrancken et al. 2017). The lifespan of cccDNA is an important component of the half-life of the cccDNA reservoir, which describes how the population of cccDNA molecules in an individual declines over time. We predict an average cccDNA lifespan of 61 days during HBeAg<sup>POS</sup> CHB compared to only 26 days in the HBeAg<sup>NEG</sup> phase of infection. Although estimates for the mutation and evolutionary rates for HBV are associated with high levels of uncertainty, our predicted lifespan is in agreement with *in vitro* studies showing a 40-day half-life of HBV cccDNA (Ko et al. 2018) and an estimated half-life of 33–57 days in woodchucks and ducks *in vivo* (Zhu et al. 2001; Addison et al. 2002). As far as we are aware, this is the first time cccDNA lifespan has been estimated during untreated infection. The lower lifespan during HBeAg<sup>NEG</sup> infection is consistent with a study in which VL data during therapy

was fitted to a mathematical model, concluding that the turnover of infected cells is higher if therapy is initiated during HBeAg<sup>NEG</sup> infection (Ribeiro et al. 2010), although our predictions for cccDNA persistence are longer (Ribeiro et al. 2010).

The shorter cccDNA lifespan during HBeAg<sup>NEG</sup> CHB may reflect host immune responses, with our model suggesting a doubling of the clearance rate compared to HBeAg<sup>POS</sup> infection. However, this increased clearance rate is predicted to have a modest effect on the total number of cccDNA molecules. As well as inferring the lifespan of cccDNA, we inferred cccDNA replicative capacity (a combined measure of intra- and extra-cellular amplification). Our results predict an approximate ten-fold reduction in replicative capacity between HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> phases of infection, which may reflect host innate or adaptive immune responses suppressing viral production and/or intra-cellular amplification. This can explain the lower cccDNA levels reported in HBeAg<sup>NEG</sup> CHB (Testoni et al. 2019; Downs et al. 2020), and is consistent with observations that the replicative capacity of cccDNA in the HBeAg<sup>NEG</sup> phase of infection is reduced compared to HBeAg<sup>POS</sup> infection (Testoni et al. 2019). This may reflect immune control at the level of the viral epigenome, but without cell death (Volz et al. 2007).

In our modelling approach, we have assumed that viral dynamics are driven by host-factors, in the absence of viral phenotypic evolution. The HBeAg<sup>NEG</sup> phase of infection is often associated with an emergence of precore mutations that may affect HBV replication and stability of cccDNA (Kao et al. 2003; Mason et al. 2008). Viral phenotypic evolution is likely to have an important role in viral dynamics, and we speculate that these viral phenotypic changes could increase the rate of intra-cellular amplification, but at the expense of killing the host cell, and therefore could represent a form of intra-host short-sighted evolution (Lythgoe et al. 2017).

Our estimates for cccDNA lifespan have implications for curative treatment strategies. If NA therapy inhibits all cccDNA amplification, we would predict HBV to be cured after 1–10 years of continuous treatment. However, this is not observed in the clinic, with only 1 per cent of individuals clearing HBsAg each year (Downs et al. 2019). Possible explanations for this discrepancy are that NAs do not inhibit all intra- and extra-cellular amplification (Dahari et al. 2009; Boyd et al. 2016), or the existence of long-lived infected cells (Allweiss et al. 2018; De Crignis et al. 2019). Our model is consistent with the presence of long-lived infected cells providing the most parsimonious explanation for sustained infection on treatment. There is growing evidence that there is negligible intra-cellular cccDNA amplification in

human HBV infection (Tu and Urban 2018), and since NA treatment will inhibit the genesis of viral particles this will prevent extra-cellular amplification. Furthermore, the dynamics of cccDNA clearance is sensitive to the assumed amplification rates, and therefore if amplification alone explains the dynamics we would expect to see a proportion of individuals clearing infection within 1–2 years of starting treatment. The presence of long-lived HBV infected cells has parallels with the HIV reservoir, where long-lived latent-infected CD4<sup>+</sup> T cells prevent cure (Palmer et al. 2008). Distinguishing between residual amplification and long-lived infected cells will help define the expected impact of treatment strategies that prevent cccDNA replication, compared to those directly targeting cccDNA. As HBV evolution will only occur if there is cccDNA amplification, it may be possible to distinguish between these two mechanisms by measuring the rate of cccDNA evolution whilst on treatment.

Our estimates of cccDNA persistence and amplification provide insights into mechanisms underlying CHB and will inform our understanding of how spontaneous or therapeutic clearance may be achieved. Given different infection profiles among individuals, and limited datasets available for our model, the confidence intervals of our estimations are wide. Our analysis exemplifies the power of modelling as a tool to inform therapeutic interventions and highlights the need for genomic studies to determine HBV evolutionary rates in CHB.

## 4. Methods

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

### 4.1 A within-host model of HBV dynamics

HBV cccDNA can replicate via intra- and extra-cellular routes (Fig. 1A), with a reported copy number between 1 and 50 molecules within a single hepatocyte nucleus (Jilbert et al. 1992; Kajino et al. 1994; Zhang et al. 2003; Li et al. 2017; Allweiss et al. 2018; Ko et al. 2018) (the higher estimates tend to be for DHBV and lower estimates for human HBV (Tu and Urban 2018)). Since cccDNA can be lost during mitosis, we modelled the number of cccDNA copies in the liver, rather than the number of infected cells. To do this, we implicitly assume that viral production is proportional to the number of cccDNA molecules. This is a reasonable assumption since VL has been reported to associate with increasing cccDNA copy numbers (Lenhoff and Summers 1994; Testoni et al. 2019).

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

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

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  $\beta$  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 (Mason et al. 2016), and because evidence suggests there is a maximum number of copies of cccDNA that can persist within each hepatocyte, which is virally controlled (Summers et al. 1990, 1991; Lenhoff and Summers 1994). 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  $\delta$  per day due to cytolytic immune responses, and cccDNA is lost at per capita rate  $c$  per day due to non-cytolytic immune responses.

The final term describes the loss of cccDNA due to cell proliferation. Uninfected and infected hepatocytes are assumed to proliferate at the per capita rate  $\rho(t)$  per day, and hence cccDNA will be exposed to proliferation at rate,  $\rho(t)$ , with a probability  $q$  that a cccDNA molecule will survive mitosis. Since the maximum possible number of cccDNA,  $K$ , is assumed to be constant, proliferation and cell death are balanced, hence:

$$dK + \delta N(t) = \rho(t)K. \quad (1b)$$

Here, the left-hand side represents the loss in carrying capacity due to the natural death rate of all hepatocytes (including infected hepatocytes), and the additional death rate incurred by infected hepatocytes as a direct result of infection. We are making the assumption that cccDNA are randomly distributed among infected hepatocytes, and that the natural death rate of hepatocytes is independent of whether they are infected or not. Thus, for example, a 5 per cent drop in the number hepatocytes results in a 5 per cent drop in cccDNA carrying capacity. The right-hand side of the equation represents consequential gain in cccDNA carrying capacity due to proliferation that is needed for  $K$  to remain constant.

A complete expression for the dynamics of  $N(t)$  can be found by solving Equation (1b) for  $\rho(t)$  and substituting into Equation (1a). To simplify further, we consider the cccDNA burden in the liver,  $Y(t) = N(t)/K$ , rather than the total number of cccDNA molecules, giving us:

$$\frac{dY(t)}{dt} = b [1 - Y(t)]Y(t) - (d + \delta + c)Y(t) - (1 - q)[d + \delta Y(t)]Y(t), \quad (2)$$

where  $b = \beta K$  is a rescaled measure of cccDNA replicative capacity. From this equation we can calculate the basic reproductive rate of cccDNA,  $R_0$ , which is defined as the number of new cccDNA molecules a single cccDNA molecule will produce in a susceptible population of hepatocytes:

$$R_0 = \frac{b}{d(2 - q) + (\delta + c)}. \quad (3)$$

If  $R_0 < 1$ , then the infection cannot be sustained in the long term. At equilibrium, the cccDNA burden is given by:

$$Y_E = \text{Max} \left[ 0, \frac{b - d(2 - q) - (\delta + c)}{b + \delta(1 - q)} \right], \quad (4)$$

which is equivalent to the cccDNA burden during stable chronic infection. Our model considers the number of cccDNA molecules independent of their distribution within cells. This is similar to the ‘single copy’ modelling assumption used in (Reaiche-Miller et al. 2013), in which only one cccDNA molecule can persist in a cell, and which was shown to produce almost identical dynamics to one in which multiple copies of cccDNA are explicitly modelled within infected cells (Reaiche-Miller et al. 2013).

#### 4.2 An expression for the neutral rate of HBV evolution

In a large well-mixed viral population, and in the absence of selection, the rate of evolution at time  $t$  is given by  $S(t) = \mu/g(t)$ , where  $\mu$  is the (neutral) mutation rate, measured per site per viral generation, and  $g(t)$  is the generation time (Kimura 1985). For our within-host model of HBV infection,  $g$  is equivalent to the typical amount of time it takes for one cccDNA molecule to replicate another molecule. This is similar to the meaning of the generation time in demography and epidemiology (Svensson 2007; Nishiura 2010; Scalia et al. 2010). From Equation (2), we can calculate the ‘backwards’ generation time as:

$$g(t) = 1/[b(1 - Y(t))]. \quad (5)$$

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

$$S(t) = \frac{\mu}{g(t)} = \mu b[1 - Y(t)]. \quad (6)$$

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

$$S_E = \frac{\mu b[c + (2 - q)(d + \delta)]}{b + \delta(1 - q)} \text{ for } Y_E > 0. \quad (7)$$

#### 4.3 Lifespan of cccDNA during steady state infection

In our model, at equilibrium the generation time of HBV will be equal to the typical cccDNA lifespan,  $L_E$ . At equilibrium the number of cccDNA molecules remains constant, and therefore the rate at which cccDNA is produced is equal to the rate at which cccDNA is lost due to infected cell death, non-cytolytic clearance of cccDNA, and proliferation of infected cells. Since the reciprocal of the production rate is equal to the generation time, and the reciprocal of the rate cccDNA is lost is the typical lifespan of cccDNA, at equilibrium, viral generation time and cccDNA lifespan are identical ( $g_E = L_E$ ). This relationship holds because of our assumption of constant death rate and hence exponentially distributed lifetimes of cccDNA (Scalia et al. 2010); see (Svensson 2007; Scalia et al. 2010) for how this changes for different distributions.

Using the equivalence of  $g_E$  and  $L_E$ , the lifespan of cccDNA at equilibrium can be determined from the mutation and neutral evolution rates by rearranging the first part of Equation (6):

$$L_E = \frac{\mu}{S_E}. \quad (8)$$

Substituting the expression for  $S_E$  from Equations (7) into (8), we can write an expression for the lifespan of cccDNA at equilibrium based on the model parameters:

$$L_E = \frac{b + \delta(1 - q)}{b[c + (d + \delta)(2 - q)]} \text{ for } Y_E > 0. \quad (9)$$

#### 4.4 The lifespan of cccDNA when few cells are infected

If infection increases the death rate of hepatocytes, then the level of proliferation (to replace eliminated cells) will be larger the more cells are infected. Consequently, the lifespan of cccDNA when few cells are infected (e.g. during early phases of infection or during spontaneous clearance of infection, or after prolonged successful suppressive treatment) may differ from the lifespan during HBeAg<sup>POS</sup> or HBeAg<sup>NEG</sup> steady state infection. By setting  $Y \ll 1$  in Equation (2), we can derive an expression for cccDNA lifespan when the copy number or burden is low:

$$L_{Y \ll 1} \approx \frac{1}{c + d(2 - q) + \delta}. \quad (10)$$

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

As we noted above, the cccDNA lifespan is only equivalent to the generation time at equilibrium. Using Equation (5), when few cells are infected, the generation time is given by:

$$g_{Y \ll 1} \approx \frac{1}{b}. \quad (11)$$

This has also been observed in the epidemiological literature (Scalia et al. 2010). Combining Equations (3), (10), and (11) we see that:

$$R_0 = \frac{L_{Y \ll 1}}{g_{Y \ll 1}}. \quad (12)$$

If  $R_0 > 1$  and few cells are infected (i.e. the number of cccDNA is increasing) the life expectancy of cccDNA will be greater than the viral generation time, whereas if  $R_0 < 1$  the life expectancy will be less than the viral generation time. This might be the case if, for example, increased immune responses associated with HBeAg<sup>NEG</sup> infection push  $R_0$  below one.

#### 4.5 Estimating the generation time and lifespan of cccDNA from within-host evolutionary rates

During stable chronic infection, the lifespan of cccDNA,  $L_E$ , equals the viral generation time,  $g_E$ , with  $g_E = \mu/S_E$  (Equation

(6). Although the mutation rate of HBV has not been determined, for avian hepadnavirus it has been estimated at  $2 \times 10^{-5}$  s/s/c (in the range  $0.8 \times 10^{-5}$  to  $4.5 \times 10^{-5}$ ; Pult et al. 2001). Since we are interested in the neutral rate of evolution, we assume that a third of all mutations in non-overlapping reading frames are synonymous, and that synonymous mutations are neutral or nearly neutral (Cuevas et al. 2012), giving a neutral mutation rate of around  $0.67 \times 10^{-5}$  s/s/c ( $0.3 \times 10^{-5}$  to  $1.5 \times 10^{-5}$ ) in non-overlapping reading frames. To incorporate the uncertainty associated with this estimate, we assumed the probability of the true mutation rate is log-normally distributed with mean  $10^{-5.2}$  and standard deviation (SD)  $10^{0.2}$ .

Using longitudinal HBV sequence data, rates of evolution for non-overlapping regions of the genome were generated using a relaxed clock method, inferring  $16.1 \times 10^{-8}$  ( $8.1 \times 10^{-8}$ ,  $25.5 \times 10^{-8}$ ) substitutions per site per day (s/s/day) for HBeAg<sup>POS</sup> and  $38.9 \times 10^{-8}$  ( $27.2 \times 10^{-8}$ ,  $51.5 \times 10^{-8}$ ) for HBeAg<sup>NEG</sup> chronic infection (the numbers in brackets give the 5 and 95 per cent highest posterior density (HPD) intervals; see Table 5 in Harrison et al. (2011)). In a separate study, using data from (Lin et al. 2015), the synonymous rate of evolution in non-overlapping genomic regions was estimated as half of the overall rate of evolution (Vrancken et al. 2017). Assuming synonymous mutations are neutral, and that the ratio of synonymous to nonsynonymous evolutionary rates is constant during infection, we therefore take the neutral within-host rates of evolution during the HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> phases of infection to be half the rates of evolution reported in (Harrison et al. 2011) for non-overlapping reading frames. This gives a neutral rate of evolution of  $8.0 \times 10^{-8}$  ( $4.0 \times 10^{-8}$ ,  $12.7 \times 10^{-8}$ ) s/s/day during the HBeAg<sup>POS</sup> phase, and  $19.5 \times 10^{-8}$  ( $13.6 \times 10^{-8}$ ,  $25.8 \times 10^{-8}$ ) s/n/day during the HBeAg<sup>NEG</sup> phase. We assumed the probability distributions of these rates are normally distributed, with the SD calculated using the difference between the estimated rate and the 5 per cent HPD.

We randomly sampled from each of the probability distribution functions for the mutation rate and substitution rates and used these values to calculate the generation time of cccDNA during HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> CHB. This was repeated 100,000 times, from which the probability distributions for cccDNA generation time during HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> chronic infection were estimated using the built in SmoothKernelDistribution function in *Mathematica* (Wolfram Research, Inc, 2015). Assuming the number of cccDNA rapidly reaches equilibrium during HBeAg<sup>POS</sup> and HBeAg<sup>NEG</sup> infection, the virus generation time will provide an approximation of the cccDNA lifespan during stable chronic infection (Fig. 3).

#### 4.6 Time to cccDNA eradication on treatment

Apart from when treatment is first initiated, the number of infected cells on treatment will be relatively low. Assuming eradication in our model is achieved when fewer than one cccDNA molecule persists, and there is no cccDNA replication whilst on treatment, the time to eradication can be approximated by:

$$t_{\text{erad}} \approx L_{Y \ll 1} \ln[N_{\text{init}}], \quad (13)$$

where  $N_{\text{init}}$  is the number of cccDNA when therapy is initiated and  $\ln$  is the natural logarithm. To determine reasonable values for  $N_{\text{init}}$ , we multiplied the number of hepatocytes in a human liver by the number of cccDNA per hepatocyte during untreated infection. There are about  $1.4 \times 10^8$  hepatocytes per gram of

human liver (Sohlenius-Sternbeck 2006), and an adult human liver is around 1.5 kg, giving approximately  $2 \times 10^{11}$  hepatocytes in total. In a recent study, an average of 6.3 copies of cccDNA per hepatocyte were found during chronic HBeAg<sup>POS</sup> infection, and 0.01 per hepatocyte during HBeAg<sup>NEG</sup> infection (Testoni et al. 2019), which gives a total of approximately  $1 \times 10^{12}$  copies of cccDNA during HBeAg<sup>POS</sup> infection and  $2 \times 10^9$  copies of cccDNA during HBeAg<sup>NEG</sup> infection.

#### 4.7 Model assuming a subset of long-lived hepatocytes

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

$$\frac{dN(t)}{dt} = b[N(t) + r(t)] \left[ \frac{(1 - \alpha)K - N(t)}{(1 - \alpha)K} \right] - (d + \delta + c)N(t) - \rho_N(t)(1 - q)N(t), \quad (14a)$$

$$\frac{dr(t)}{dt} = b[N(t) + r(t)] \left[ \frac{\alpha K - r(t)}{\alpha K} \right] - (d_r + \delta_r + c_r)r(t) - \rho_r(t)(1 - q)r(t), \quad (14b)$$

where  $d_r$ ,  $\delta_r$ , and  $c_r$  represent the per capita natural death rates, cytolytic death rate, and clearance rates of cccDNA in long-lived cells, and  $\rho_N$  and  $\rho_r$  the per capita proliferation rates of standard- and long-lived cells. Since deaths of standard (or long-lived) cells are balanced by proliferation of standard (or long-lived) cells, we have the relationships:

$$\rho_N(t)(1 - \alpha)K = d(1 - \alpha)K + \delta N(t), \quad (14c)$$

$$\rho_L(t)\alpha K = d_r\alpha K + \delta_r r(t), \quad (14d)$$

As before, we can substitute expressions for  $\rho_N(t)$  and  $\rho_r(t)$  found from solving Equations (14c, d), into (14a, b), convert numbers into proportions, and simplify, to give expressions for the burden of cccDNA in standard-lived cells,  $Y(t)$ , and in long-lived cells,  $Z(t)$ :

$$\frac{dY(t)}{dt} = b \left[ 1 - \frac{Y(t)}{1 - \alpha} \right] [Y(t) + Z(t)] - (d + \delta + c)Y(t) - (1 - q) \left[ d + \frac{\delta Y(t)}{1 - \alpha} \right] Y(t), \quad (15a)$$

$$\frac{dZ(t)}{dt} = b \left[ 1 - \frac{Z(t)}{\alpha} \right] [Y(t) + Z(t)] - (d_L + \delta_L + c_L)Z(t) - (1 - q) \left[ d + \frac{\delta_L Z(t)}{\alpha} \right] Z(t). \quad (15b)$$

### Supplementary data

Supplementary data are available at *Virus Evolution* online.

### Acknowledgements

Thank you to Fabian Zoulim and Helen Fryer for their helpful comments on earlier versions of this work.

## Author contributions

K.A.L.: Conceptualisation, Formal analysis, Funding acquisition, Methodology, Writing—original draft, Writing—Review & Editing, S.F.L.: Formal analysis, Investigation, Writing—original draft, L.P.: Methodology, Writing—Review & Editing, J.A.M.: Conceptualisation, Writing—Review & Editing, P.C.M.: Conceptualisation, Writing—Review & Editing.

## Funding

K.A.L. and L.P. are supported by The Wellcome Trust and the Royal Society grant numbers, 107652/Z/15/Z and 202562/Z/16/Z. J.A.M. is supported by EU 2020 Research and Innovation Programme Consortia HEP-CAR under grant agreement No. 667273, MRC MR/R022011/1, and The Wellcome Trust IA 200838/Z/16/Z. P.C.M. is supported by The Wellcome Trust, grant number 110110/Z/15/Z.

**Conflict of interest:** None declared.

## References

- Addison, W. R. et al. (2002) 'Half-life of the Duck Hepatitis B Virus Covalently Closed Circular DNA Pool *In Vivo* following Inhibition of Viral Replication', *Journal of Virology*, 76: 6356–63.
- Allweiss, L. et al. (2018) 'Proliferation of Primary Human Hepatocytes and Prevention of Hepatitis B Virus Reinfection Efficiently Deplete Nuclear cccDNA *In Vivo*', *Gut*, 67: 542–52.
- Bonhoeffer, S. et al. (2003) 'Glancing behind Virus Load Variation in HIV-1 Infection', *Trends in Microbiology*, 11: 499–504.
- Boyd, A. et al. (2016) 'Decay of ccc-DNA Marks Persistence of Intrahepatic Viral DNA Synthesis under Tenofovir in HIV-HBV Co-infected Patients', *Journal of Hepatology*, 65: 683–91.
- Cuevas, J. M. et al. (2012) 'The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses', *Molecular Biology and Evolution*, 29: 17–20.
- Dahari, H. et al. (2009) 'Modeling Complex Decay Profiles of Hepatitis B Virus during Antiviral Therapy', *Hepatology*, 49: 32–8.
- De Crignis, E. et al. (2019) 'Human Liver Organoids; a Patient-derived Primary Model for HBV Infection and Related Hepatocellular Carcinoma', *bioRxiv*, doi: 10.1101/568147.
- Downs, L. O. et al. (2019) 'Electronic Health Informatics Data to Describe Clearance Dynamics of Hepatitis B Surface Antigen (HBsAg) and e Antigen (HBeAg) in Chronic Hepatitis B Virus Infection', *mBio*, 10: 1–15.
- et al. (2020) 'Bimodal Distribution and Set Point HBV DNA Viral Loads in Chronic Infection: Retrospective Analysis of Cohorts from the UK and South Africa', *Wellcome Open Research*, 5: 113.
- EASL. (2017) 'EASL 2017 Clinical Practice Guidelines on the Management of Hepatitis B Virus Infection', *Journal of Hepatology*, 67: 370–98.
- Goyal, A. et al. (2017) 'The Role of Infected Cell Proliferation in the Clearance of Acute HBV Infection in Humans', *Viruses*, 9: 350.
- Harrison, A. et al. (2011) 'Genomic Analysis of Hepatitis B Virus Reveals Antigen State and Genotype as Sources of Evolutionary Rate Variation', *Viruses*, 3: 83–101.
- Jilbert, A. R. et al. (1992) 'Rapid Resolution of Duck Hepatitis B Virus Infections Occurs after Massive Hepatocellular Involvement', *Journal of Virology*, 66: 1377–88.
- Kajino, K. et al. (1994) 'Woodchuck Hepatitis Virus Infections: Very Rapid Recovery after a Prolonged Viremia and Infection of Virtually Every Hepatocyte', *Journal of Virology*, 68: 5792–803.
- Kao, J. H. et al. (2003) 'Basal Core Promoter Mutations of Hepatitis B Virus Increase the Risk of Hepatocellular Carcinoma in Hepatitis B Carriers', *Gastroenterology*, 124: 327–34.
- Kimura, M. (1985) *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Ko, C. et al. (2018) 'Hepatitis B Virus Genome Recycling and De Novo Secondary Infection Events Maintain Stable cccDNA Levels', *Journal of Hepatology*, 69: 1231–41.
- Langford, R. et al. (2014) 'Population Size and the Rate of Evolution', *Trends in Ecology & Evolution*, 29: 33–41.
- Laporte, V., and Charlesworth, B. (2002) 'Effective Population Size and Population Subdivision in Demographically Structured Populations', *Genetics*, 162: 501–19.
- Lenhoff, R. J., and Summers, J. (1994) 'Construction of Avian Hepadnavirus Variants with Enhanced Replication and Cytopathicity in Primary Hepatocytes', *Journal of Virology*, 68: 5706–13.
- Li, M. et al. (2017) 'Distribution of Hepatitis B Virus Nuclear DNA', *Journal of Virology*, 92: 1–13.
- Lin, Y.-Y. et al. (2015) 'New Insights into the Evolutionary Rate of Hepatitis B Virus at Different Biological Scales', *Journal of Virology*, 89: 3512–22.
- Locarnini, S., and Zoulim, F. (2010) 'Molecular Genetics of HBV Infection', *Antiviral Therapy*, 15: 3–14.
- Lucifora, J., and Protzer, U. (2016) 'Attacking Hepatitis B Virus cccDNA—The Holy Grail to Hepatitis B Cure', *Journal of Hepatology*, 64: S41–8.
- Lythgoe, K. A. et al. (2016) 'Large Variations in HIV-1 Viral Load Explained by Shifting-Mosaic Metapopulation Dynamics', *PLoS Biology*, 14: e1002567.
- et al. (2017) 'Short-sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections', *Trends in Microbiology*, 25: 336–48.
- Mason, W. S. et al. (2008) 'Immune Selection during Chronic Hepadnavirus Infection', *Hepatology International*, 2: 3–16.
- et al. (2016) 'HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant', *Gastroenterology*, 151: 986–98.e4.
- McNaughton, A. L. et al. (2019) 'Insights from Deep Sequencing of the HBV genome—Unique, Tiny, and Misunderstood', *Gastroenterology*, 156: 384–99.
- Müller, V. et al. (2001) 'Small Variations in Multiple Parameters Account for Wide Variations in HIV-1 Set-points: A Novel Modelling Approach', *Proceedings of the Royal Society of London Series B: Biological Sciences*, 268: 235–42.
- Murray, J. M., and Goyal, A. (2015) 'In Silico Single Cell Dynamics of Hepatitis B Virus Infection and Clearance', *Journal of Theoretical Biology*, 366: 91–102.
- Nassal, M. (2015) 'HBV cccDNA: Viral Persistence Reservoir and Key Obstacle for a Cure of Chronic Hepatitis B', *Gut*, 64: 1972–84.
- Nishiura, H. (2010) 'Time Variations in the Generation Time of an Infectious Disease: Implications for Sampling to Appropriately Quantify Transmission Potential', *Mathematical Biosciences and Engineering*, 7: 851–69.
- Palmer, S. et al. (2008) 'Low-level Viremia Persists for at Least 7 Years in Patients on Suppressive Antiretroviral Therapy', *Proceedings of the National Academy of Sciences*, 105: 3879–84.
- Pereira-Gómez, M. et al. (2016) 'Lamivudine/Adefovir Treatment Increases the Rate of Spontaneous Mutation of Hepatitis B Virus in Patients', *PLoS One*, 11: e0163363.

- Pult, I. et al. (2001) 'Frequency of Spontaneous Mutations in an Avian Hepadnavirus Infection', *Journal of Virology*, 75: 9623–32.
- Pybus, O. G. et al. (2011) 'Viral Mutation and Substitution: Units and Levels', *Current Opinion in Virology*, 1: 430–5.
- Reaiche-Miller, G. Y. et al. (2013) 'Duck Hepatitis B Virus Covalently Closed Circular DNA Appears to Survive Hepatocyte Mitosis in the Growing Liver', *Virology*, 446: 357–64.
- Revell, P. A. et al. (2019) 'A Global Scientific Strategy to Cure Hepatitis B', *The Lancet Gastroenterology & Hepatology*, 4: 545–8.
- Ribeiro, R. M. et al. (2010) 'Hepatitis B Virus Kinetics under Antiviral Therapy Sheds Light on Differences in Hepatitis B e Antigen Positive and Negative Infections', *The Journal of Infectious Diseases*, 202: 1309–18.
- Sanjuán, R. et al. (2010) 'Viral Mutation Rates', *Journal of Virology*, 84: 9733–48.
- Scalia, G. et al. (2010) 'Some Model Based Considerations on Observing Generation Times for Communicable Diseases', *Mathematical Biosciences*, 223: 24–31.
- Scholle, S. O. et al. (2013) 'Viral Substitution Rate Variation Can Arise from the Interplay between within-Host and Epidemiological Dynamics', *The American Naturalist*, 182: 494–513.
- Sohlenius-Sternbeck, A. K. (2006) 'Determination of the Hepatocellularity Number for Human, Dog, Rabbit, Rat and Mouse Livers from Protein Concentration Measurements', *Toxicology In Vitro*, 20: 1582–6.
- Summers, J. et al. (1990) 'Hepadnavirus Envelope Proteins Regulate Covalently Closed Circular DNA Amplification', *Journal of Virology*, 64: 2819–24.
- et al. (1991) 'Morphogenetic and Regulatory Effects of Mutations in the Envelope Proteins of an Avian Hepadnavirus', *Journal of Virology*, 65: 1310–7.
- Svensson, Å. (2007) 'A Note on Generation Times in Epidemic Models', *Mathematical Biosciences*, 208: 300–11.
- Testoni, B. et al. (2019) 'Serum Hepatitis B Core-related Antigen (HBcrAg) Correlates with Covalently Closed Circular DNA Transcriptional Activity in Chronic Hepatitis B Patients', *Journal of Hepatology*, 70: 615–25.
- Tu, T., and Urban, S. (2018) 'Virus Entry and Its Inhibition to Prevent and Treat Hepatitis B and Hepatitis D Virus Infections', *Current Opinion In Virology*, 30: 68–79.
- Urban, S. et al. (2010) 'The Replication Cycle of Hepatitis B Virus', *Journal of Hepatology*, 52: 282–4.
- Volz, T. et al. (2007) 'Impaired Intrahepatic Hepatitis B Virus Productivity Contributes to Low Viremia in Most HBeAg-negative Patients', *Gastroenterology*, 133: 843–52.
- Vrancken, B. et al. (2017) 'Accurate Quantification of within- and between-Host HBV Evolutionary Rates Requires Explicit Transmission Chain Modelling', *Virus Evolution*, 3: 1–9.
- Whalley, B. S. A. et al. (2001) 'Kinetics of Acute Hepatitis B Virus Infection in Humans', *The Journal of Experimental Medicine*, 193: 847–54.
- WHO. (2017) *Global Hepatitis Report*.
- Wolfram Research, Inc. (2015) *Mathematica, Version 10.3*. Champaign, IL: Wolfram Research, Inc.
- Wong, D. K. H. et al. (2013) 'Reduction of Hepatitis B Surface Antigen and Covalently Closed Circular DNA by Nucleos(t)ide Analogues of Different Potency', *Clinical Gastroenterology and Hepatology*, 11: 1004–10.e1.
- Zhang, Y. et al. (2003) 'Single-cell Analysis of Covalently Closed Circular DNA Copy Numbers in a Hepadnavirus-infected Liver', *Proceedings of the National Academy of Sciences*, 100: 12372–7.
- Zhu, Y. et al. (2001) 'Kinetics of Hepadnavirus Loss from the Liver during Inhibition of Viral DNA Synthesis', *Journal of Virology*, 75: 311–22.