

Field validation of phylodynamic analytical methods for inference on epidemiological processes in wildlife

Carlo Pacioni¹, Timothy Vaughan², Tanja Strive³, Susan Campbell⁴, David Ramsey⁵, and Alexei Drummond^{6,7}

¹Arthur Rylah Institute for Environmental Research

²ETH Zürich

³CSIRO

⁴Western Australia Department of Primary Industry and Regional Development

⁵Arthur Rylah Institute

⁶University of Auckland

⁷The University of Auckland

July 20, 2020

Abstract

Amongst newly developed approaches to analyse molecular data, phylodynamic models are receiving much attention because of their potential to reveal changes to viral populations over short periods. This knowledge can be very important for understanding disease impacts. However, their accuracy needs to be fully understood, especially in relation to wildlife disease epidemiology, where sampling and prior knowledge may be limited. The release of the rabbit haemorrhagic disease virus (RHDV) as biological control in naïve rabbit populations in Australia in 1996 provides a unique dataset with which to validate phylodynamic models. By comparing the results obtained for RHDV1 with our current understanding of the RHDV epidemiology in Australia, we evaluated the performances of these recently developed models. In line with our expectations, coalescent analyses detected a sharp increase in the virus trajectory in the first few months after the virus release, followed by a more gradual increase. The phylodynamic analyses with a birth-death tree prior generated effective reproductive number estimates (the average number of secondary infections per each infectious case, R_e) larger than one for most of the epochs considered. However, the possible range of the initial R_e included estimates lower than one despite the known rapid spread of RHDV1 in Australia. Furthermore, the analyses that took into account the geographical structuring failed to converge. We argue that the difficulties that we encountered most likely stem from the fact that the samples available from 1996 to 2014 were too sparse with respect to geographic and within outbreak coverage to adequately infer some of the model parameters. In general, while these Bayesian analyses proved to be greatly informative in some regards, we caution that their interpretation may not be straight forward and recommend further research in evaluating the robustness of these models to assumption violations and sensitivity to sampling regimes.

Field validation of phylodynamic analytical methods for inference on epidemiological processes in wildlife

Running title: Phylodynamic methods for inference in wildlife

Carlo Pacioni^{1,2,3}, Timothy G. Vaughan⁴, Tanja Strive^{3,5}, Susan Campbell^{3,6}, David S. L. Ramsey^{1,3} and Alexei J. Drummond⁷

¹Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, Heidelberg, VIC 3084.

²School of Veterinary and Life Sciences, Murdoch University, Murdoch, 6150, Western Australia.

³Centre for Invasive Species Solutions, Bruce, ACT, 2617

⁴Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland 4058.

⁵Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT 2601.

⁶Department of Primary Industries and Regional Development, Albany. WA 6330.

⁷Centre for Computational Evolution, University of Auckland, Auckland, New Zealand 1142.

carlo.pacioni@delwp.vic.gov.au

Summary

Amongst newly developed approaches to analyse molecular data, phylodynamic models are receiving much attention because of their potential to reveal changes to viral populations over short periods. This knowledge can be very important for understanding disease impacts. However, their accuracy needs to be fully understood, especially in relation to wildlife disease epidemiology, where sampling and prior knowledge may be limited. The release of the rabbit haemorrhagic disease virus (RHDV) as biological control in naïve rabbit populations in Australia in 1996 provides a unique dataset with which to validate phylodynamic models. By comparing the results obtained for RHDV1 with our current understanding of the RHDV epidemiology in Australia, we evaluated the performances of these recently developed models.

In line with our expectations, coalescent analyses detected a sharp increase in the virus trajectory in the first few months after the virus release, followed by a more gradual increase. The phylodynamic analyses with a birth-death tree prior generated effective reproductive number estimates (the average number of secondary infections per each infectious case, R_e) larger than one for most of the epochs considered. However, the possible range of the initial R_e included estimates lower than one despite the known rapid spread of RHDV1 in Australia. Furthermore, the analyses that took into account the geographical structuring failed to converge. We argue that the difficulties that we encountered most likely stem from the fact that the samples available from 1996 to 2014 were too sparse with respect to geographic and within outbreak coverage to adequately infer some of the model parameters. In general, while these Bayesian analyses proved to be greatly informative in some regards, we caution that their interpretation may not be straight forward and recommend further research in evaluating the robustness of these models to assumption violations and sensitivity to sampling regimes.

Introduction

Applications of coalescent-based analytical techniques to molecular data allow the reconstruction of demographic changes over time of the studied organisms (Drummond et al., 2005). These techniques have been successfully applied to reconstruct the demographic history of pathogens as well as higher order organisms (Raghwani et al., 2019, Pacioni et al., 2015). In parallel, recent efforts have aimed to extend phylogenetic birth-death models to estimate parameters relevant for epidemiological studies. In these models, the rate of transmission and recovery can be estimated allowing calculation of the basic reproductive number (the number of secondary infections per each infectious case in a naïve population, R_0) (Stadler et al., 2012). Further developments of these models aimed to relax limiting assumptions (e.g. that the sampled infected individuals either die or are removed) while co-inferring evolutionary (i.e. phylogeny) and epidemiological processes (e.g. Gavryushkina et al., 2014, Stadler et al., 2013). Investigations on these research topics constitute the currently very active field of research termed “phylodynamics”.

Viruses are ideal targets for these models as they have very high mutation rates, allowing detection of changes in population dynamics over a very short period (Drummond et al., 2003). The utility of these models for recovering epidemiological processes has been investigated with simulation studies (Stadler et al., 2013)

as well as observational comparisons, mostly with the current understanding of diseases affecting humans (e.g. Stadler et al., 2014). Their applications to wildlife diseases have been limited and their accuracy has been broadly assumed *a priori*. However, several aspects of wildlife disease investigations present additional challenges compared to human infectious disease research. Knowledge of the epidemiology of the pathogen under study is often limited (Wobeser, 2007), making the selection of biologically appropriate models difficult. Compared to human diseases, the sampling regimes in wildlife are generally limited both temporally and spatially. Also, the lack of clear definition of host's epidemiological compartments (groups of individuals with similar characteristics in respect to the pathogen: e.g. susceptible, infected, etc.) due to knowledge gaps, or inadequate sampling presents further challenges. Models that do not consider this additional complexity may fail to accurately represent dynamics of infectious diseases. Furthermore, it is difficult to predict the possible biases of parameter estimates when analysing data with these simplified models. Hence, there is a need for further investigation of phylodynamic models to determine their utility for analysing the epidemiology of wildlife diseases. The analysis of Rabbit Haemorrhagic Disease Virus (RHDV) data from Australia provides a unique opportunity to validate the application of these methods to monitor virus dynamics within a wildlife host. A strong molecular clock-likeness (i.e. the fact that the accumulation of mutations is strictly related with time) of RHDV sequence data has been demonstrated by a number of previous studies (Eden et al., 2015b, Kovaliski et al., 2014, Eden et al., 2015a). This property makes RHDV particularly suitable for the analysis of the heterochronous sequence data that were available for this study. Subsequent to the escape of the virus from quarantine field trials in the mid-1990s, monitoring sites were set up throughout the country to detect the arrival of the virus in different regions. This monitoring allowed a detailed reconstruction of the temporal-spatial spread of the epidemic (e.g. Bruce et al., 2004, Kovaliski, 1998). We, consequently, have a strong prior knowledge of what dynamics we should be able to retrieve from our phylodynamic models.

Since their introduction in 1859, European rabbits (*Oryctolagus cuniculus*) have had a devastating impact on agricultural production and biodiversity in Australia, with competition and land degradation by rabbits listed as key threatening processes under the *Commonwealth's Environmental Protection and Biodiversity Act (1999)*. Prior to the occurrence of Rabbit Haemorrhagic Disease in Australia, rabbit damage was reported to cost Australian agriculture up to \$600 million annually. More recent estimates place the cost of rabbits on agricultural production losses at approximately \$200 million annually, with an additional \$25 million spent each year on management and research costs (Cox et al., 2013). In addition to physical control techniques (e.g. shooting, poisoning, warren ripping and harbour destruction), the myxoma virus (*Myxomatosis cuniculi*) and RHDV constitute the most important landscape scale control strategies for rabbits in Australia, having resulted in excess of \$70 billion in economic savings since the 1950s (Cooke et al., 2013). The RHDV (Czech strain), commonly identified as RHDV1, escaped from Wardang Island, South Australia, in October 1995, where its efficacy as a biological control agent was being tested under natural conditions. Soon after the virus escaped to the mainland, monitoring programs that looked at both rabbit abundance and individuals' serological status, were established in multiple sites across all States and Territories. Through this monitoring, it was demonstrated that the virus became established nationwide within 12 months (Kovaliski, 1998) and it was hypothesised that the rapid spread of the virus was partly due to flies feeding on infected carcasses (Henning et al., 2005, McColl et al., 2002). In experimental trials, death of susceptible rabbits normally occurs within two to three days post infection (Elsworth et al., 2014) and at some field monitoring sites, population reductions of more than 90% were recorded (Mutze et al., 1998, Bruce and Twigg, 2005). However, based on subsequent monitoring, it also became apparent that the efficacy of RHDV was highly variable with average abundance reductions of 28% in high rainfall areas compared to 67% in lower rainfall areas throughout Australia (Bruce et al., 2004, Cooke et al., 2002, Cox et al., 2013).

At some locations, serological analyses detected antibodies cross reacting to RHDV in rabbits prior to the arrival of RHDV, suggesting that RHDV effectiveness was likely impeded by the presence of a related, non-pathogenic calicivirus (Cooke et al., 2002, Robinson et al., 2002a, Nagesha et al., 2000), which was later identified (Strive et al., 2009, Strive et al., 2010, Strive et al., 2013).

Since approximately 2005, rabbit abundance has recovered by 5-10 fold compared to initial post-RHDV1 levels, likely due to the decreasing efficacy of RHDV as control agent (Mutze et al., 2014a). However, the exact

mechanisms involved are not well understood. While Elsworth et al. (2014) demonstrated that field RHDV variants collected between 2007 and 2009 were associated with increased case fatality rates in Australian wild rabbits from the same location, it is generally accepted that other mechanisms may have facilitated the recovery of rabbits. Young rabbits have a physiological resistance to lethal infection until at least 4 weeks of age (Fordham et al., 2014, Abrantes et al., 2012, Robinson et al., 2002b). Mechanisms that enable rabbit populations to recover may therefore include a shift in the age classes that are infected (Mutze et al., 2014b, Elsworth et al., 2014), which results in a lower fatality rate, variability of the doses of infection (Wells et al., 2018), or possibly a combination of both. Further elements that may influence RHDV1 epidemiology are the host population structure, existing (age-related) epidemiological compartments within each rabbit sub-population and the presence/absence of localised myxoma disease epidemics (Barnett et al., 2018).

We aim to analyse available RHDV molecular data to reconstruct the virus dynamics from the first RHDV epidemic in Australia to 2014. As we have a relatively good understanding of the disease dynamics within the first 10 years (1995-2005) since the initial virus release in Australia, we can validate the applied phylogenetic models with field data. A secondary aim is to retrieve additional epidemiological parameters such as effective reproductive number (the average number of secondary infections per each infectious case, R_e) with a particular focus on the detection of recent changes (post 2005) in naturally occurring strains.

Based on general epidemiological theory, we expect the virus to evolve to maximise R_e , with a trade-off in virulence. However, field data suggest that RHDV has not evolved to become more attenuated but appears to be co-evolving with rabbits to maintain very high levels of virulence (Elsworth et al., 2014) - probably because flies that have acquired RHDV from rabbit carcasses are one of the main means for disease transmission over long distances, as opposed to live, diseased animals, which keeps the selection pressure for high levels of virulence that produces dead bodies for effective transmission (Schwensow et al., 2014, Elsworth et al., 2014). Concurrently, as the virus established in Australia, various levels of acquired immunity in animals surviving RHDV infection would have led to a reduction in the pool of susceptible individuals. Our initial hypothesis, based on the available evidence, is that we would expect R_e to be much higher than 1 during the initial outbreak in Australia because of the naïve conditions of Australian rabbit populations (with the exception of a small proportion of rabbits with cross-immunity due to the non-pathogenic calicivirus). As a result of a substantial reduction of the host density, with the number of susceptible hosts further reduced by acquired immunity in animals surviving virus infection, we would expect R_e to decline, although staying above 1 because RHDV1 successfully established itself as endemic in Australia. It is harder to anticipate what the results of the analysis may be in regard to RHDV1 dynamics between 2005 and 2014, when field data indicated a drastic recovery of most rabbit populations, with one possibility being that R_e continued to slowly decrease below one or stayed stable.

Methods

We searched GenBank to obtain sequence data for RHDV1 (Table S1) before different virus strains were known to occur in Australia (Hall et al., 2015, Mahar et al., 2017). In fact, up until 2014, descendants of the originally released RHDV strain were the only virulent RHDV strains known to circulate in Australia (Kovaliski et al., 2014). It is important to highlight that, in all RHDV molecular studies conducted in Australia, sampling was opportunistic, i.e. sampling occurred from rabbit carcasses opportunistically encountered after a disease outbreak. Hence, while we had a substantial temporal coverage, we could only obtain data for one or two individuals within each site and year.

We retained only sequences where, in the original publication, the number of screened samples were stated or where it was confirmed with the author(s) that no bias was introduced towards highly variable sequences. Time of collection of the samples was included in the analysis as fractional years before present. We aligned the sequences in Geneious (Kearse et al., 2012) and, as most sequences found were from the hypervariable gene that encodes for the capsid protein VP60, we trimmed the alignment to include only data for this gene, leaving 1740 nt for analysis.

Using the collated dataset, we conducted coalescent and newly developed phylodynamic analyses and compared results with known epidemiological data collected in the field, thereby providing an indication of the accuracy of these analytical methods implemented in a natural context. Specifically, we initially used the Bayesian sampled ancestor birth-death skyline tree prior (SABDSKY, Gavryushkina et al., 2014). We considered this the most flexible model and used it to test the fit of the data to a strict molecular clock, or alternatively a lognormal or exponential relaxed molecular clock (Drummond et al., 2006) by using the nested sampling approach (Russel et al., 2018) and comparing the Bayes Factor between models (Baele et al., 2012). For each model, we gradually increased the subchain length until we obtained two approximately equal maximum likelihood estimates, between consecutive analyses. We then conducted a second run with the longest subchain to ensure consistent results. We used a very tight lognormal prior around the origin parameter (log mean=2.95 and log stdev=0.01), with a hard lower and upper bound of 18.47 and 19.8 years before the most recent sample, as Australia was RHDV free before its release and we considered it unlikely for outbreaks to go unnoticed for a period longer than a few months. We used a lognormal prior also for R_e (log mean=0.5 and log stdev=1) and the total removal rate δ (Gavryushkina et al., 2014) (log mean=2 and log stdev=2). We used a uniform distribution bounded between 0 and 1 for the sampling proportion parameter, but fixed this parameter to zero prior to 1995 because no samples were collected prior to this date. We also fixed the removal probability parameter to zero because we sampled very few individuals across the many that had been infected, and even though in some instances we removed the whole carcass at sampling (therefore reducing the probability of transmitting the disease), it is likely that they already had opportunities to infect other rabbits, so we considered appropriate to make the assumption that sampling did not affect RHDV1 outbreaks. We initially modelled R_e and δ with a piecewise constant function through time, allowing different values for these parameters for each five year “epoch” (i.e. a total of four time intervals after 1995). We considered this sufficient to detect changes in these parameters and to specifically test the hypothesis that there had been a change after 2005. However, on inspection of preliminary results, we noted an inverse correlation between R_e and δ . Therefore, for all analyses we used a common δ across all epochs while letting R_e to vary. Furthermore, to better reflect the sampling times, we fixed the sampling proportion parameter to zero in the years where we did not have any samples as an alternative approach.

In all analyses, we used bModelTest (Bouckaert and Drummond, 2017) to take into account the uncertainty of the substitution model, and a gamma prior (shape=1, scale=0.001) for the clock rate, which encompasses the expected range for an RNA virus (Biek et al., 2015). After defining the molecular clock, we repeated the analysis with a Birth-Death skyline (BDSKY, Stadler et al., 2013), a Bayesian skyline (BSP, Drummond et al., 2005) and a time-aware Gaussian Markov random field Bayesian skyride coalescent tree priors (Skyride, Minin et al., 2008). We allowed the BSP analysis to consider six population size changes, and applied a lognormal population size prior (with precision equal to 1) to reconstruct changes over time of the virus population size. The mean of the lognormal population size prior for each epoch is the estimated population size of the previous epoch, with a uniform prior (between 0 and 3.8×10^5) for the oldest epoch. For the few sequences for which the exact date of collection was unknown (accession number: GU373617, GU373618, EU650679, EU650680), we put a lognormal prior on the tip date to match the possible range and reflect this uncertainty (the season of collection for these sequences was known, Table S1). It should be noted that the tip dates sampler operator, which is commonly used for this purpose, cannot be used in the SABDSKY model (Gavryushkina pers. comm.), as it is incompatible with the extended state space implied by that model. Instead, we replaced this with the purposely designed ‘Sampled Node Date Random Walker’ operator developed and described by Gavryushkina et al. (2014). Lastly, we attempted to take into account geographical compartmentalisation using the multitype Birth-Death model (aka: Birth-Death with Migration Model, BDMM) (Kühnert et al., 2016). For this analysis we removed the (few) sequences from Western Australia and Tasmania, and limited the analysis to two demes: central-south Australia and east Australia (all priors were kept as above). All analyses were conducted in BEAST2 (Bouckaert et al., 2014) except the skyride analysis, which was run in BEAST1 (Drummond and Rambaut, 2007) with a GTR + Γ + I. Tracer (Rambaut and Drummond, 2007) was used to ensure convergence of the two MCMC that were run for each model (that is, that independent analyses sampled the same parameter space towards the end of the MCMC), and to generate the data for the BSP and the skyride plot. The two runs were then combined for

final analysis. Plots for the BDSKY and SABDSKY analyses were generated with the R package `bdskytools` (<https://github.com/laduplessis/bdskytools>). All BEAST input files are available from the corresponding author.

Results

The most supported molecular clock model was the lognormal relaxed molecular clock, with a Bayes factor with the next most supported model (strict molecular clock) equal to 225, therefore we used this molecular clock prior in all subsequent analyses. Under the SABDSKY model, most parameter estimates were very similar, so we report them jointly unless otherwise specified.

The mutation rate posterior median was 4.46×10^{-3} per year (95% HPD: 3.72×10^{-3} - 5.36×10^{-3} per year), with a coefficient of variance of 0.5 (95% HPD: 0.28-0.9). The root of the tree was estimated to be in 1995.36 (95% HPD: 1995.09 – 1995.58) and the origin in 1995.23 (95% HPD: 1994.92 – 1995.52). The sequence KT006745 (sampled in South Australia in 2009) had a sampled ancestor posterior probability of 38.3% and it was always paired, when sampled ancestor, with the sequence KF494930 (sampled in South Australia in 2010).

Across the four time intervals after 1995, the median estimated R_e had a slightly increasing trend until 2010 followed by a dramatic decline (Figure 1). However, based on the 95% Highest Probability Density (HPD) intervals, no significant difference was detected across the first three epochs in the R_e estimate. The range of values for R_e in the first five years after the release of the virus was very low (0.76 – 1.4), successively increasing to 0.9 – 1.4 and then declining to below 1 after 2010 (0.2 – 0.9) (Figure 1). The range of values for δ was 0.9 – 5.6 (median=2.6) and the sampling proportion parameter was estimated to be 0.008 (0.0005 – 0.0277) in the SABDSKY model with a common sampling proportion parameter across all the epochs (after 1995), while δ was narrower (0.6 – 2.2, median=1.3) and the sampling proportion showed a declining trend (Figure 2) in the SABDSKY model with a variable sampling proportion parameter. Results from the BDSKY analyses were similar to SABDSKY although with wider 95% HPD (Figure 1), with the exception that δ estimate was much smaller (0.5, 0.17 – 1.8) and the sampling proportion parameter had a much wider 95% HPD (0.02 – 0.9) in the BDSKY model with a common sampling proportion parameter across all the epochs as well as in the BDSKY model with a variable sampling proportion parameter (Figure 2).

The BSP analysis, correctly detected the initial dramatic increase in the virus population size at the beginning of the initial outbreak (Figure 3), despite the paucity of sequence data in the first few years since the release of the virus (Figure S1). The analysis suggests a gradual increase in the virus population size until 2007 and another substantial increase between 2007 and 2009. The virus population size then stabilised in the last five years or so of the analysis. The results of the skyride analysis were similar, and, although they had larger HPD intervals and the second increase in the population size between 2007 and 2009 was not as evident (Figure S2), these are not discussed further.

The analysis with the BDMM did not converge. We further attempted to reduce the time interval for the estimation of R_e and fixed δ across the different epochs to reduce the number of parameters to be estimated with no success. The estimation of the sampling proportion and type change events proved particularly difficult and, even with runs in excess of 1.6×10^9 iterations the latter parameter failed to converge.

Discussion

We evaluated the performances of recently developed phylodynamic and coalescent-based analyses using sequence data for RHDV and compared these results with current understanding of RHDV field epidemiology and our current understanding of the disease in wild rabbits in Australia.

Phylodynamic analyses

The two phylodynamic analyses (SABDSKY and BDSKY) were very similar, probably because there was only one sample ancestor in the analysis. In relative terms, these analyses agreed with our expectations as the median R_e estimates were larger than one for most of the analysed time windows. However, it was surprising that the possible range of R_e values (i.e. the HPD) in the first five years since RHDV release included estimates only marginally larger than one, with a median actually lower than one in some analysis, in spite of the known rapid spread of the disease in Australia and the approximate 90% rabbit abundance reduction observed at some sites (Cox et al., 2013). Furthermore, R_e estimates were not substantially higher than the successive epochs (and possibly not as high as we would have expected). It is conceivable that R_e rapidly reduced as soon as the first outbreak wave was over, while in our analysis we are attempting to estimate a R_e value that encompasses the first five years since the release of the virus in Australia, which may force the analysis to a lower value. We also argue that the limited coverage of the data that we had access to, possibly contributed to this result. The data from GenBank was very sparse because we only obtained data from approximately one individual per outbreak, for a very small fraction of the total number of outbreaks that have occurred in Australia and, most likely, our analysis lacks the capacity to estimate R_e in a classical epidemiological sense (i.e. the number of infections created by an infected individual). In other words, rather than operating at an individual basis, we argue that, given the resolution of our temporal and spatial sampling, our inference is on parameters more related to the rate of spread and recovery of local outbreaks, rather than individuals. In this view, R_e could be considered the number of virus field variants that are generated by a virus lineage; δ represents the rate at which lineages are not detectable anymore; and the sampling proportion refers to the proportion of outbreaks detected. Since the infectious period can be estimated as $1/\delta$, if our interpretation is correct, this would be equivalent to the time a variant is detectable in the field. In our analysis, this would correspond to approximately four months to two years on average. We recommend further work to evaluate how sampling regime (i.e. sampling effort) could influence the parameter estimations and affect these analyses.

An additional challenge that we faced in these analyses was that geographical heterogeneity is likely to be an important element when considering RHDV dynamics. In fact, it is believed that rabbit susceptibility, due to heritable, acquired or abiotic factors, is not homogeneous throughout Australia, which may cause R_e values to vary greatly in different regions (Henzell et al., 2003). Therefore, not all populations recover equally from RHDV infections and field data suggest that climatic and other environmental variables may also influence RHDV transmission (Henzell et al., 2003). All these factors can make estimating mean R_e values that adequately describe the overall (i.e. nationwide) virus dynamic difficult because neither BDSKY or SABDSKY cater for such further structuring. It would be particularly useful if future studies improved our capacity to predict the extent and the direction of possible biases when model assumptions are not met. Models that take into account structuring (Kühnert et al., 2016) or more complex disease dynamics (Volz, 2012) exist, and we attempted the implementation of one such model. However, these analyses failed. It is well known that analyses that take into consideration geographical structuring need adequate sampling within each location (and possibly, all the relevant demes sampled) and we argue that the lack of a rigorous geographical coverage of the historic data available to us is the most likely cause of the difficulties we encountered in the BDMM analysis. Unfortunately, logistical and resource issues that limit the collection of samples that would be suitable for these more complex models are a common occurrence in wildlife research.

Both phylodynamic analyses (BDSKY and SABDSKY) showed a sharp decrease in R_e since 2010 and indeed, inspecting the phylogenetic tree (Figure S3), we noted that most of post-2010 samples had very long branches. As mentioned above, the mechanisms of the rabbit recovery in the last 10 years are not quite clear. Field data indicate that the frequency of RHDV outbreaks and the prevalence of the disease in the monitored sites did not change over time (Mutze et al., 2014b, Wells et al., 2018). Hence, this finding was somewhat unexpected, but possibly points to an alternative mechanism that caused the recovery of the rabbit populations. In 2010, weather changes caused by La Niña broke a prolonged drought that started in 1996 and encompassed the whole continent, known in Australia as ‘The Millennium Drought’. It is possible that this drastic change in environmental conditions altered RHDV epidemiology either directly or by altering

vector activities, reducing the long-distance transmission of the virus. The heavy rainfall may also have increased the competition between the benign variant (RCV-A1) and RHDV1 in the areas where this virus was active. It is known that the former virus' spread is facilitated in wet conditions (Liu et al., 2014). Further development of rabbit genetic resistance to the virus infection could also be a contributing factor, and has been suggested as a possible explanation for reduced RHDV mortality rates and recovery of rabbit populations in some parts of Australia {Mutze, 2014 #2683}. We considered RHDV1 sequences up until 2014 because it is believed that alternative virus variants now known from Australia were not circulating before this date. For this reason, we judge it highly unlikely that new variants, especially RHDV2, which caused a substantial rabbit mortality wave in 2015/16 {Ramsey, 2020 #3137}, would go undetected and were not competing with RHDV1 between 2010 and 2014. While all these are possible explanations of the quite drastic reduction of R_e since 2010, and it is possible that field work focused in sites where RHDV1 remained active rather than being an unbiased, representative selection across the whole country, these remain to be proved.

Coalescent based analyses

The coalescent based analyses (BSP and skyride) were very informative and generally supported our current knowledge of the RHDV dynamics in Australia. For example, both analyses correctly detected a sharp initial increase in the virus population size, which subsequently slowed down from 1997-1998 onwards. The further increase of the virus population size between 2007 and 2009, detected by the coalescent-based analyses, was unexpected. We note that strains isolated in the field in this period were demonstrated to be of higher virulence compared to the original strain released in Australia (Elsworth et al., 2014). It has been hypothesised that virus selection tends to occur in clusters (Gog and Grenfell, 2002) and it may be possible that this is what we are witnessing here. Indeed, the VP60 capsid is under selection pressure from the host immune system and it is possible that obtaining data from other regions of the genomes could help to provide some evidence to this end. However, we were not able to confirm this hypothesis with the data available and it is also possible that other conditions that affect the virus' life cycle (such the ones mentioned above in regard to possible changes in transmission rate) could artificially inflate the virus population size.

Conclusion

Our analyses were informative, and we propose that these methods can potentially be applied as tools for monitoring the efficacy of the RHDV strains in the field, as well as for other viruses in natural contexts. However, this study highlights potential complexities in parameter interpretation. To the best of our knowledge, we indicated possible explanations for our parameter estimates, but we warn the users of such methods that the interpretations of their results may not be straightforward. We recommend further research to better evaluate biases caused by the spatial distribution and intensity of sampling, and the robustness of these models to improper assumptions and incorrect parameterisation, which are all likely to occur when there is limited knowledge as commonly occurs in the fields of wildlife and invasive species research.

Data availability statement

The data that support the findings of this study are available in Genbank and their reference numbers are reported in Table S1.

Acknowledgements

We would like to thank J. Kovaliski for clarification on the data used in his publication. We are grateful to E. Holmes, J. Mahar and J. Kovaliski for the useful and stimulating conversations on this study and

K. Helbig and Lindy Lumsden for comments on the draft of the manuscript. We would like to thank D. Kühnert, A. Gavryushkina and R. Bouckaert for clarifications on important aspects of their models and their implementation in BEAST2. This project was partly funded by the Centre for Invasive Species Solutions (project PO1-B-002).

Computer simulations were partly supported by CIPRES (<http://www.phylo.org>)

Ethical statement

The authors confirm that the ethical policies of the journal, as noted on the journal’s author guidelines page, have been adhered to. No ethical approval was required as this study used data available in public database.

Conflict of interest statement

The authors declare that there are no potential conflicts of interest.

References

- Abrantes, J., W. Van Der Loo, J. Le Pendu and P. J. Esteves, 2012: Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review. *Vet. Res.*, 43, 12.
- Baele, G., P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard and A. V. Alekseyenko, 2012: Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 29, 2157-2167.
- Barnett, L. K., T. A. Prowse, D. E. Peacock, G. J. Mutze, R. G. Sinclair, J. Kovaliski, B. D. Cooke and C. J. Bradshaw, 2018: Previous exposure to myxoma virus reduces survival of European rabbits during outbreaks of rabbit haemorrhagic disease. *Journal of Applied Ecology*, 55, 2954-2962.
- Biek, R., O. G. Pybus, J. O. Lloyd-Smith and X. Didelot, 2015: Measurably evolving pathogens in the genomic era. *Trends in Ecology & Evolution*, 30, 306-313.
- Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut and A. J. Drummond, 2014: BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10, e1003537.
- Bouckaert, R. R. and A. J. Drummond, 2017: bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17, 42.
- Bruce, J. S. and L. E. Twigg, 2005: The reintroduction, and subsequent impact, of rabbit haemorrhagic disease virus in a population of wild rabbits in south-western Australia. *Wildlife Research*, 32, 139-150.
- Bruce, J. S., L. E. Twigg and G. S. Gray, 2004: The epidemiology of rabbit haemorrhagic disease, and its impact on rabbit populations, in south-western Australia. *Wildlife Research*, 31, 31-49.
- Cooke, B., P. Chudleigh, S. Simpson and G. Saunders, 2013: The economic benefits of the biological control of rabbits in Australia, 1950–2011. *Australian Economic History Review*, 53, 91-107.
- Cooke, B. D., S. McPhee, A. J. Robinson and L. Capucci, 2002: Rabbit haemorrhagic disease: does a pre-existing RHDV-like virus reduce the effectiveness of RHD as a biological control in Australia? *Wildlife Research*, 29, 673-682.
- Cox, T., T. Strive, G. Mutze, P. West and G. Saunders, 2013: *Benefits of Rabbit Biocontrol in Australia*. Invasive Animals Cooperative Research Centre, Canberra.

- Drummond, A. and A. Rambaut, 2007: BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214.
- Drummond, A. J., S. Y. W. Ho, M. J. Phillips and A. Rambaut, 2006: Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol*, 4, e88.
- Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg and A. G. Rodrigo, 2003: Measurably evolving populations. *Trends in Ecology & Evolution*, 18, 481-488.
- Drummond, A. J., A. Rambaut, B. Shapiro and O. G. Pybus, 2005: Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185-1192.
- Eden, J.-S., J. Kovaliski, J. A. Duckworth, G. Swain, J. E. Mahar, T. Strive and E. C. Holmes, 2015a: Comparative Phylodynamics of Rabbit Hemorrhagic Disease Virus in Australia and New Zealand. *Journal of Virology*, 89, 9548-9558.
- Eden, J.-S., A. J. Read, J. A. Duckworth, T. Strive and E. C. Holmes, 2015b: Resolving the origin of rabbit Hemorrhagic disease virus: insights from an investigation of the viral stocks released in Australia. *Journal of Virology*, 89, 12217-12220.
- Elsworth, P., B. D. Cooke, J. Kovaliski, R. Sinclair, E. C. Holmes and T. Strive, 2014: Increased virulence of rabbit haemorrhagic disease virus associated with genetic resistance in wild Australian rabbits (*Oryctolagus cuniculus*). *Virology*, 464-465, 415-423.
- Fordham, D. A., K. T. Shoemaker, N. H. Schumaker, H. R. Akçakaya, N. Clisby and B. W. Brook, 2014: How interactions between animal movement and landscape processes modify local range dynamics and extinction risk. *Biology Letters*, 10, 20140198.
- Gavryushkina, A., D. Welch, T. Stadler and A. J. Drummond, 2014: Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput Biol*, 10, e1003919.
- Gog, J. R. and B. T. Grenfell, 2002: Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 17209-17214.
- Hall, R. N., J. E. Mahar, S. Haboury, V. Stevens, E. C. Holmes and T. Strive, 2015: Emerging rabbit hemorrhagic disease virus 2 (RHDVb), Australia. *Emerging Infectious Diseases*, 21, 2276-2278.
- Henning, J., J. Meers, P. Davies and R. Morris, 2005: Survival of rabbit haemorrhagic disease virus (RHDV) in the environment. *Epidemiol. Infect.*, 133, 719-730.
- Henzell, R. P., R. B. Cunningham and H. M. Neave, 2003: Factors affecting the survival of Australian wild rabbits exposed to rabbit haemorrhagic disease. *Wildlife Research*, 29, 523-542.
- Jahnke, M., E. C. Holmes, P. J. Kerr, J. D. Wright and T. Strive, 2010: Evolution and phylogeography of the nonpathogenic calicivirus RCV-A1 in wild rabbits in Australia. *Journal of Virology*, 84, 12397-12404.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz and C. Duran, 2012: Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647-1649.
- Kinnear, M. and C. C. Linde, 2010: Capsid gene divergence in rabbit hemorrhagic disease virus. *Journal of General Virology*, 91, 174-181.
- Kovaliski, J., 1998: Monitoring the spread of rabbit hemorrhagic disease virus as a new biological agent for control of wild European rabbits in Australia. *Journal of Wildlife Diseases*, 34, 421-428.
- Kovaliski, J., R. Sinclair, G. Mutze, D. Peacock, T. Strive, J. Abrantes, P. J. Esteves and E. C. Holmes, 2014: Molecular epidemiology of rabbit haemorrhagic disease virus in Australia: when one became many. *Molecular Ecology*, 23, 408-420.

- Kühnert, D., T. Stadler, T. G. Vaughan and A. J. Drummond, 2016: Phylodynamics with migration: a computational framework to quantify population structure from genomic data. *Molecular Biology and Evolution*, 33, 2102-2116.
- Liu, J., D. A. Fordham, B. D. Cooke, T. Cox, G. Mutze and T. Strive, 2014: Distribution and prevalence of the Australian non-pathogenic rabbit calicivirus Is correlated with rainfall and temperature. *PLoS One*, 9, e113976.
- Mahar, J., A. Read, X. Gu, N. Urakova, R. Mourant, M. Piper, S. Haboury, E. Holmes, T. Strive and R. Hall, 2017: Detection and circulation of a novel RHDVa in Australia. *Emerging Infectious Diseases*, 24, 22-31.
- McCull, K., C. Morrissy, B. Collins and H. Westbury, 2002: Persistence of rabbit haemorrhagic disease virus in decomposing rabbit carcasses. *Australian veterinary journal*, 80, 298-299.
- Minin, V. N., E. W. Bloomquist and M. A. Suchard, 2008: Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, 25, 1459-1471.
- Mutze, G., P. Bird, S. Jennings, D. Peacock, N. de Preu, J. Kovaliski, B. Cooke and L. Capucci, 2014a: Recovery of South Australian rabbit populations from the impact of rabbit haemorrhagic disease. *Wildlife Research*, 41, 552-559.
- Mutze, G., B. Cooke and P. Alexander, 1998: The initial impact of rabbit hemorrhagic disease on European rabbit populations in South Australia. *Journal of Wildlife Diseases*, 34, 221-227.
- Mutze, G. J., R. Sinclair, D. Peacock, L. Capucci and J. Kovaliski, 2014b: Is increased juvenile infection the key to recovery of wild rabbit populations from the impact of rabbit haemorrhagic disease? *European Journal of Wildlife Research*, 60, 489-499.
- Nagesha, H., K. McCull, B. Collins, C. Morrissy, L.-F. Wang and H. Westbury, 2000: The presence of cross-reactive antibodies to rabbit haemorrhagic disease virus in Australian wild rabbits prior to the escape of virus from quarantine. *Archives of Virology*, 145, 749-757.
- Pacioni, C., H. Hunt, M. E. Allentoft, T. G. Vaughan, A. F. Wayne, A. Baynes, D. Haouchar, J. Dortch and M. Bunce, 2015: Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular Ecology*, 24, 5813-5828.
- Raghwani, J., C.-H. Wu, C. K. Y. Ho, M. De Jong, R. Molenkamp, J. Schinkel, O. G. Pybus and K. A. Lythgoe, 2019: High-resolution evolutionary analysis of within-host hepatitis C virus infection. *The Journal of Infectious Diseases*, 219, 1722-1729.
- Rambaut, A. and A. J. Drummond, 2007: TRACER. 1.6 edn. <http://beast.bio.ed.ac.uk/Tracer>
- Robinson, A. J., P. Kirkland, R. Forrester, L. Capucci, B. D. Cooke and A. Philbey, 2002a: Serological evidence for the presence of a calicivirus in Australian wild rabbits, *Oryctolagus cuniculus*, before the introduction of rabbit haemorrhagic disease virus (RHDV): its potential influence on the specificity of a competitive ELISA for RHDV. *Wildlife Research*, 29, 655-662.
- Robinson, A. J., P. T. M. So, W. J. Iler, B. D. Cooke and L. Capucci, 2002b: Statistical models for the effect of age and maternal antibodies on the development of rabbit haemorrhagic disease in Australian wild rabbits. *Wildlife Research*, 29, 663-671.
- Russel, P. M., B. J. Brewer, S. Klaere and R. R. Bouckaert, 2018: Model selection and parameter inference in phylogenetics using Nested Sampling. *Systematic biology*, 68, 219-233.
- Schwensow, N. I., B. Cooke, J. Kovaliski, R. Sinclair, D. Peacock, J. Fickel and S. Sommer, 2014: Rabbit haemorrhagic disease: virus persistence and adaptation in Australia. *Evolutionary Applications*, 7, 1056-1067.
- Stadler, T., R. Kouyos, V. von Wyl, S. Yerly, J. Böni, P. Bürgisser, T. Klimkait, B. Joos, P. Rieder, D. Xie, H. F. Günthard, A. J. Drummond, S. Bonhoeffer and the Swiss HIV Cohort Study, 2012: Estimating the

basic reproductive number from viral sequence data. *Molecular Biology and Evolution*, 29, 347-357.

Stadler, T., D. Kühnert, S. Bonhoeffer and A. J. Drummond, 2013: Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, 110, 228-233.

Stadler, T., D. Kühnert, D. A. Rasmussen and L. du Plessis, 2014: Insights into the early epidemic spread of ebola in Sierra Leone provided by viral sequence data. *PLoS Currents*, 6.

Strive, T., P. Elsworth, J. Liu, J. D. Wright, J. Kovaliski and L. Capucci, 2013: The non-pathogenic Australian rabbit calicivirus RCV-A1 provides temporal and partial cross protection to lethal rabbit haemorrhagic disease virus infection which is not dependent on antibody titres. *Vet. Res.*, 44, 51.

Strive, T., J. Wright, J. Kovaliski, G. Botti and L. Capucci, 2010: The non-pathogenic Australian lagovirus RCV-A1 causes a prolonged infection and elicits partial cross-protection to rabbit haemorrhagic disease virus. *Virology*, 398, 125-134.

Strive, T., J. D. Wright and A. J. Robinson, 2009: Identification and partial characterisation of a new lagovirus in Australian wild rabbits. *Virology*, 384, 97-105.

Volz, E. M., 2012: Complex population dynamics and the coalescent under neutrality. *Genetics*, 190, 187-201.

Wells, K., D. A. Fordham, B. W. Brook, P. Cassey, T. Cox, R. B. O’Hara and N. I. Schwensow, 2018: Disentangling synergistic disease dynamics: Implications for the viral biocontrol of rabbits. *Journal of Animal Ecology*, 87, 1418-1428.

Wobeser, G. A., 2007: *Disease in wild animals: investigation and management* . Springer Verlag, Germany, Berlin Heidelberg.

Figure legends

Figure 1: Effective reproductive number (R_e) as a function of time (years). Each considered epoch was approximately five years. Shaded area indicates the 95% highest probability density. First row are parameter estimates under the Bayesian Sampled Ancestor Birth-Death Skyline model, second row under the Bayesian Birth-Death Skyline model. On the left-hand side, with common δ and sampling proportion parameter across all epochs, while on the right-hand side with a variable sampling proportion parameter.

Figure 2: Violin plots of sampling proportion parameter (s) estimates for different time intervals (x-axis) under the Bayesian Sampled Ancestor Birth-Death Skyline (left) and the Bayesian Birth-Death Skyline (right) models. Dots indicate the medians.

Figure 3: Bayesian Skyline Plot of virus demographic changes (y-axis in log scale) over time. Shaded area indicates the 95% highest probability density.

Table

Table S1. Details of the sequences used in this study, including the year of collection and fractional year used in the analysis.

Accession number	State	Year	Fractional Year	Reference
KT006731	NSW	2007	2007.94	(Eden et al., 2015a)
KT006732	NSW	2014	2014.24	(Eden et al., 2015a)
KT006733	WA	1999	1999.40	(Eden et al., 2015a)
KT006734	SA	1999	1999.81	(Eden et al., 2015a)

Accession number	State	Year	Fractional Year	Reference
KT006735	WA	2000	2000.12	(Eden et al., 2015a)
KT006736	SA	2000	2000.39	(Eden et al., 2015a)
KT006737	SA	2002	2002.79	(Eden et al., 2015a)
KT006738	SA	2003	2003.71	(Eden et al., 2015a)
KT006739	SA	2004	2004.67	(Eden et al., 2015a)
KT006740	SA	2004	2004.49	(Eden et al., 2015a)
KT006741	SA	2005	2005.76	(Eden et al., 2015a)
KT006742	SA	2006	2006.59	(Eden et al., 2015a)
KT006743	TAS	2007	2007.33	(Eden et al., 2015a)
KT006744	SA	2008	2008.79	(Eden et al., 2015a)
KT006745	SA	2009	2009.65	(Eden et al., 2015a)
KT006746	WA	2012	2012.73	(Eden et al., 2015a)
KT006747	NSW	2013	2013.51	(Eden et al., 2015a)
KT006748	SA	2013	2013.72	(Eden et al., 2015a)
KT344772	SA	1995	1995.78	(Eden et al., 2015b)
GU373617	ACT	2009	2009.15-2009.93	(Jahnke et al., 2010)
GU373618	ACT	2009	2009.15-2009.93	(Jahnke et al., 2010)
EU650679	NSW	2005	2005.67-2005.83	(Kinnear and Linde, 2010)
EU650680	NSW	2006	2006.67-2006.83	(Kinnear and Linde, 2010)
KF494906	ACT	2009	2009.25	(Kovaliski et al., 2014)
KF494907	ACT	1998	1998.75	(Kovaliski et al., 2014)
KF494908	ACT	1998	1998.75	(Kovaliski et al., 2014)
KF494909	ACT	2010	2010.17	(Kovaliski et al., 2014)
KF494910	ACT	1998	1998.75	(Kovaliski et al., 2014)
KF494911	ACT	2011	2011.67	(Kovaliski et al., 2014)
KF494912	NSW	2007	2007.75	(Kovaliski et al., 2014)
KF494913	NSW	2007	2007.25	(Kovaliski et al., 2014)
KF494914	NSW	2011	2011.33	(Kovaliski et al., 2014)
KF494915	NSW	2011	2011.33	(Kovaliski et al., 2014)
KF494916	NSW	2011	2011.25	(Kovaliski et al., 2014)
KF494917	NSW	2011	2011.25	(Kovaliski et al., 2014)
KF494918	NSW	2011	2011.25	(Kovaliski et al., 2014)
KF494919	NSW	2011	2011.58	(Kovaliski et al., 2014)
KF494920	NSW	2011	2011.17	(Kovaliski et al., 2014)
KF494921	SA	1995	1995.78	(Kovaliski et al., 2014)
KF494922	SA	2008	2008.83	(Kovaliski et al., 2014)
KF494923	SA	2009	2009.67	(Kovaliski et al., 2014)
KF494924	SA	1995	1995.82	(Kovaliski et al., 2014)
KF494925	SA	2008	2008.75	(Kovaliski et al., 2014)
KF494926	SA	2008	2008.5	(Kovaliski et al., 2014)
KF494927	SA	2008	2008.83	(Kovaliski et al., 2014)
KF494928	SA	2003	2003.67	(Kovaliski et al., 2014)
KF494929	SA	1995	1995.92	(Kovaliski et al., 2014)
KF494930	SA	2010	2010.83	(Kovaliski et al., 2014)
KF494931	SA	2010	2010.67	(Kovaliski et al., 2014)
KF494932	SA	1995	1995.83	(Kovaliski et al., 2014)
KF494933	SA	2011	2011.75	(Kovaliski et al., 2014)
KF494934	SA	2011	2011.67	(Kovaliski et al., 2014)
KF494935	SA	1999	1999.83	(Kovaliski et al., 2014)
KF494936	SA	1999	1999.83	(Kovaliski et al., 2014)

Accession number	State	Year	Fractional Year	Reference
KF494937	SA	2006	2006.92	(Kovaliski et al., 2014)
KF494938	SA	2006	2006.92	(Kovaliski et al., 2014)
KF494939	SA	2008	2008.83	(Kovaliski et al., 2014)
KF494940	SA	2011	2011.67	(Kovaliski et al., 2014)
KF494943	SA	2008	2008.83	(Kovaliski et al., 2014)
KF494944	SA	2008	2008.75	(Kovaliski et al., 2014)
KF494945	SA	2009	2011.67	(Kovaliski et al., 2014)
KF494946	SA	2010	2010.42	(Kovaliski et al., 2014)
KF494947	SA	1995	1995.92	(Kovaliski et al., 2014)
KF494948	TAS	2006	2006.67	(Kovaliski et al., 2014)
KF494949	TAS	2007	2007.42	(Kovaliski et al., 2014)
KF494950	WA	2010	2010.75	(Kovaliski et al., 2014)
KF494951	WA	2010	2010.75	(Kovaliski et al., 2014)
KF494952	WA	2010	2010.92	(Kovaliski et al., 2014)



