# Frequent intergenotypic recombination between the non-structural and structural genes is a major driver of epidemiological fitness in caliciviruses

Jackie E. Mahar,[1,2,†] Maria Jenckel,[2] Nina Huang,[2,‡] Elena Smertina,[2,4] Edward C. Holmes,[1,§] Tanja Strive,[2,3] and Robyn N. Hall[2,3,**,*]

[1]Marie Bashir Institute for Infectious Disease and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia, [2]Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Black Mountain, ACT 2601, Australia, [3]Centre for Invasive Species Solutions, University of Canberra, Bruce, ACT 2617, Australia and [4]Faculty of Science and Technology, University of Canberra, Bruce, ACT 2617, Australia

[†]https://orcid.org/0000-0003-2687-9810
[‡]https://orcid.org/0000-0003-4202-2825
[§]https://orcid.org/0000-0001-9596-3552
[**]https://orcid.org/0000-0002-5836-2990
*Corresponding author: E-mail: robyn.hall@csiro.au

## Abstract

The diversity of lagoviruses (*Caliciviridae*) in Australia has increased considerably in recent years. By the end of 2017, five variants from three viral genotypes were present in populations of Australian rabbits, while prior to 2014 only two variants were known. To understand the evolutionary interactions among these lagovirus variants, we monitored their geographical distribution and relative incidence over time in a continental-scale competition study. Within 3 years of the incursion of rabbit haemorrhagic disease virus 2 (RHDV2, denoted genotype GI.1bP-GI.2 [polymerase genotype]P-[capsid genotype]) into Australia, two novel recombinant lagovirus variants emerged: RHDV2-4e (genotype GI.4eP-GI.2) in New South Wales and RHDV2-4c (genotype GI.4cP-GI.2) in Victoria. Although both novel recombinants contain non-structural genes related to those from benign, rabbit-specific, enterotropic viruses, these variants were recovered from the livers of both rabbits and hares that had died acutely. This suggests that the determinants of host and tissue tropism for lagoviruses are associated with the structural genes, and that tropism is intricately connected with pathogenicity. Phylogenetic analyses demonstrated that the RHDV2-4c recombinant emerged independently on multiple occasions, with five distinct lineages observed. Both the new RHDV2-4e and -4c recombinant variants replaced the previous dominant parental RHDV2 (genotype GI.1bP-GI.2) in their respective geographical areas, despite sharing an identical or near-identical (i.e. single amino acid change) VP60 major capsid protein with the parental virus. This suggests that the observed replacement by these recombinants was not driven by antigenic variation in VP60, implicating the non-structural genes as key drivers of epidemiological fitness. Molecular clock estimates place the RHDV2-4e recombination event in early to mid-2015, while the five RHDV2-4c recombination events occurred from late 2015 through to early 2017. The emergence of at least six viable recombinant variants within a 2-year period highlights the high frequency of these events, detectable only through intensive surveillance, and demonstrates the importance of recombination in lagovirus evolution.

Key words: calicivirus; recombination; genomic epidemiology; lagovirus; RHDV2; fitness

## 1. Introduction

Caliciviruses are an important group of vertebrate-infecting viruses. As well as noroviruses, the major cause of gastroenteritis in humans worldwide (Parra et al., 2017), this family also includes the genus *Lagovirus* that contains viruses infecting rabbits and hares. While some lagoviruses are hepatotropic and cause an acute fulminant viral hepatitis with a case fatality rate exceeding 90 per cent, others are enterotropic and thought to be entirely benign (Abrantes et al., 2012). These are referred to as rabbit haemorrhagic disease viruses (RHDVs) and rabbit caliciviruses (RCVs), respectively.

Lagoviruses are hierarchically classified principally by their major capsid protein (VP60) type, and less frequently by polymerase type, into genogroups (e.g. GI, GII), genotypes (e.g. GI.1, GI.2, GI.4), and variants (e.g. GI.1a, GI.1b, GI.1c) (Le Pendu et al., 2017). The first lagoviruses identified, from hares in Europe in the early 1980s, were those of the GII.1 genotype (European brown hare syndrome virus) (Gavier-Widen and Morner, 1991). Subsequently, mortality events in *Oryctolagus* rabbits in China in 1984 led to the discovery of RHDV1, a genotype GI.1 virus (Liu et al., 1984). As the diversity of GI.1 viruses increased, this genotype was further subdivided into several variants based on the VP60
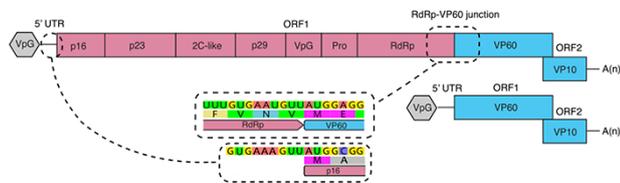
**Figure 1.** Lagovirus genome organisation. gRNAs encode the non-structural proteins (pink) and structural proteins (blue), while sgRNAs encode only the structural proteins. Both RNAs are linked to VpG at their 5′-end and are polyadenylated at their 3′-end. Inset boxes demonstrate the nucleotide sequence identity between the 5′-untranslated region of the gRNA and the RdRp–VP60 junction.

phylogeny. In 1997, antigenic variants of GI.1 viruses, now classified as RHDVa (GI.1a), were described in Italy (Capucci et al., 1998) and these spread throughout Europe and more distantly. Then, in 2010, RHDV2 (GI.2) first emerged in France (Dalton et al., 2012; Le Gall-Recule et al., 2011a, 2013). RHDV2 rapidly spread globally, triggering epizootics worldwide in wild and domestic lagomorph populations. As surveillance efforts and technologies improved, non-pathogenic lagoviruses RCV-E1 (genotype GI.3), RCV-E2 (GI.4), and RCV-A1 (also GI.4) were identified in the 1990s, first in Europe (Capucci et al., 1996, 1997, Le Gall-Recule et al., 2011b, 2018) and then in Australia (Strive, Wright, and Robinson 2009) and New Zealand (Nicholson et al., 2017). Non-pathogenic hare lagoviruses, HaCVs (GII.2, GII.3, GII.4, and GII.5), have also been reported recently, in 2014 in Europe (Cavadini et al., 2020; Droillard et al., 2020) and 2019 in Australia (Mahar et al., 2019).

Lagoviruses, like other caliciviruses, are small, non-enveloped viruses containing a monopartite positive-sense single-stranded RNA genome approximately 7.5 kb in length (Fig. 1) (Abrantes et al., 2012). Virus particles contain both genomic RNA (gRNA) and 3′-co-terminal sub-genomic RNAs (sgRNA) approximately 2.5 kb in length (Ohlinger et al., 1990). In lagoviruses, the non-structural (NS) genes are situated upstream of the major capsid protein, VP60, all of which are encoded as a single large polyprotein (ORF1), while the 3′-terminal ORF2 encodes a minor structural protein that is essential for infectivity and is presumed to be important for genome release during infection (Sosnovtsev et al., 2005; Conley et al., 2019). Lagoviruses encode seven NS proteins, including a helicase (2C-like), a viral genome-linked protein (VpG), a protease (Pro), an RNA-dependent RNA polymerase (RdRp), and three proteins of unknown functions (p16, p23, and p29). Both gRNA and sgRNA are linked to VpG at their 5′-end and are polyadenylated at their 3′-end (Meyers, Wirblich, and Thiel 1991). There is considerable homology between the 5′-terminal nucleotide sequences of the gRNA and sgRNA (Fig. 1) (Meyers, Wirblich, and Thiel 1991), enabling recombination at this site. The RdRp–VP60 junction is also highly conserved within and between genogroups, facilitating intergenotypic template switching at this region.

Recombination is an important evolutionary mechanism in many RNA viruses (Worobey et al., 1999). In the *Caliciviridae*, recombination frequently occurs at the junction between the NS and structural genes, effectively mixing a set of structural genes with an entirely new set of NS genes (Bull et al., 2005; Ludwig-Begall, Mauroy, and Thiry 2018; Abrantes et al., 2020). Recombinant lagoviruses are defined by the nomenclature [RdRp genotype]P-[capsid genotype], and include combinations of two pathogenic RHDVs (e.g. GI.1bP-GI.2) as well as the seemingly more common combination of a benign RCV and a pathogenic RHDV (e.g. GI.4eP-GI.1a, GI.4eP-GI.2, GI.3P-GI.2) (Lopes et al., 2015,

2019; Dalton et al., 2018; Hall et al., 2018; Mahar et al., 2018b; Silverio et al., 2018; Abrantes et al., 2020). Phylogenetic analyses have demonstrated that all genotype GI.2 viruses described so far are recombinants, implying that GI.2 is an orphan capsid-type (Abrantes et al., 2020). Intergenogroup recombinants between GI (rabbit and hare) and GII (hare) viruses have also been reported, with two genotype GII.1P-GI.2 viruses recovered from hares in Germany in 2014 and 2019 (Szillat et al., 2020).

Prior to 2016, five GI lagovirus variants had been reported in Australia—RCV-A1 (GI.4), RHDV1 (GI.1c), RHDVa-K5 (GI.1a), RHDVa-Aus (GI.4eP-GI.1a), and RHDV2 (GI.1bP-GI.2). These variants differ in their host and tissue tropism and in pathogenicity (Fig. 2) (Mahar et al., 2018a). RCV-A1 (GI.4) is a benign enterotropic virus that has circulated in wild and domestic rabbits since at least the 1950s (Strive, Wright, and Robinson 2009; Mahar et al., 2016). RHDV1 (GI.1c) and RHDVa-K5 (GI.1a) are pathogenic viruses, both deliberately released as biocontrols (Kerr, Hall, and Strive 2021). RHDVa-Aus (GI.4eP-GI.1a) and RHDV2 (GI.1bP-GI.2) were both exotic incursions, first detected through syndromic surveillance activities in January 2014 and May 2015, respectively (Hall et al., 2015; Mahar et al., 2018a). Both exotic incursions are themselves recombinant viruses, and, like the orphan GI.2 structural genes, the NS genes of RHDVa-Aus (GI.4eP-GI.1a) form an orphan clade within the diversity of GI.4 sequences that was designated as genotype GI.4e (Mahar et al., 2018b). Importantly, the orphan nature of this clade is likely due to a lack of lagovirus genomic surveillance in China, the likely source country of RHDVa-Aus. Phylogenetic analyses suggest that both incursions arose from single point source introductions with subsequent ongoing transmission in Australian rabbit populations (Mahar et al., 2018a,b). Phylodynamic analyses place the incursion of RHDV2 (GI.1bP-GI.2) during early 2014, while serological analyses suggest the incursion occurred slightly later during the spring of 2014 (September to November) (Mahar et al., 2018a; Ramsey et al., 2020). Both analyses indicated that RHDV2 (GI.1bP-GI.2) rapidly replaced RHDV1 (GI.1c) within 18 months of its arrival. Insufficient RHDVa-Aus (GI.4eP-GI.1a) sequences are available for phylodynamic analyses to reliably estimate the time of incursion. In mid-2017, a novel GI.4eP-GI.2 recombinant virus (designated here as RHDV2-4e) was detected in Australia, comprising the NS genes of the RHDVa-Aus virus (GI.4eP) and structural genes of RHDV2 (GI.2) (Hall et al., 2018).

Australia is an isolated landmass with a large rabbit population that is distributed across most of the country, making it an ideal setting for understanding the evolution, recombination, and epidemiological interactions of caliciviruses. The genetic diversity of Australian lagoviruses arose from a small number of incursion events, and the parental sequence as well as the timing of the introduction is known in several instances. Since there is little movement of rabbits into the country, the study of lagoviruses can occur in a relatively controlled environment with a defined number of pre-existing variants. Additionally, a robust surveillance system for these viruses has been in place since 2014. Prior to this, syndromic surveillance and convenience sampling were used to investigate rabbit mortality events since the 1990s, providing a broad understanding of lagovirus presence and epidemiology in Australia. The previous lack of diversity of the Australian lagovirus population and sudden diversification of lagovirus variants in 2014–5 created ideal natural experimental conditions for monitoring the epidemiological and evolutionary dynamics of existing lagovirus variants, while determining the frequency and importance of successful recombination events between these viruses.
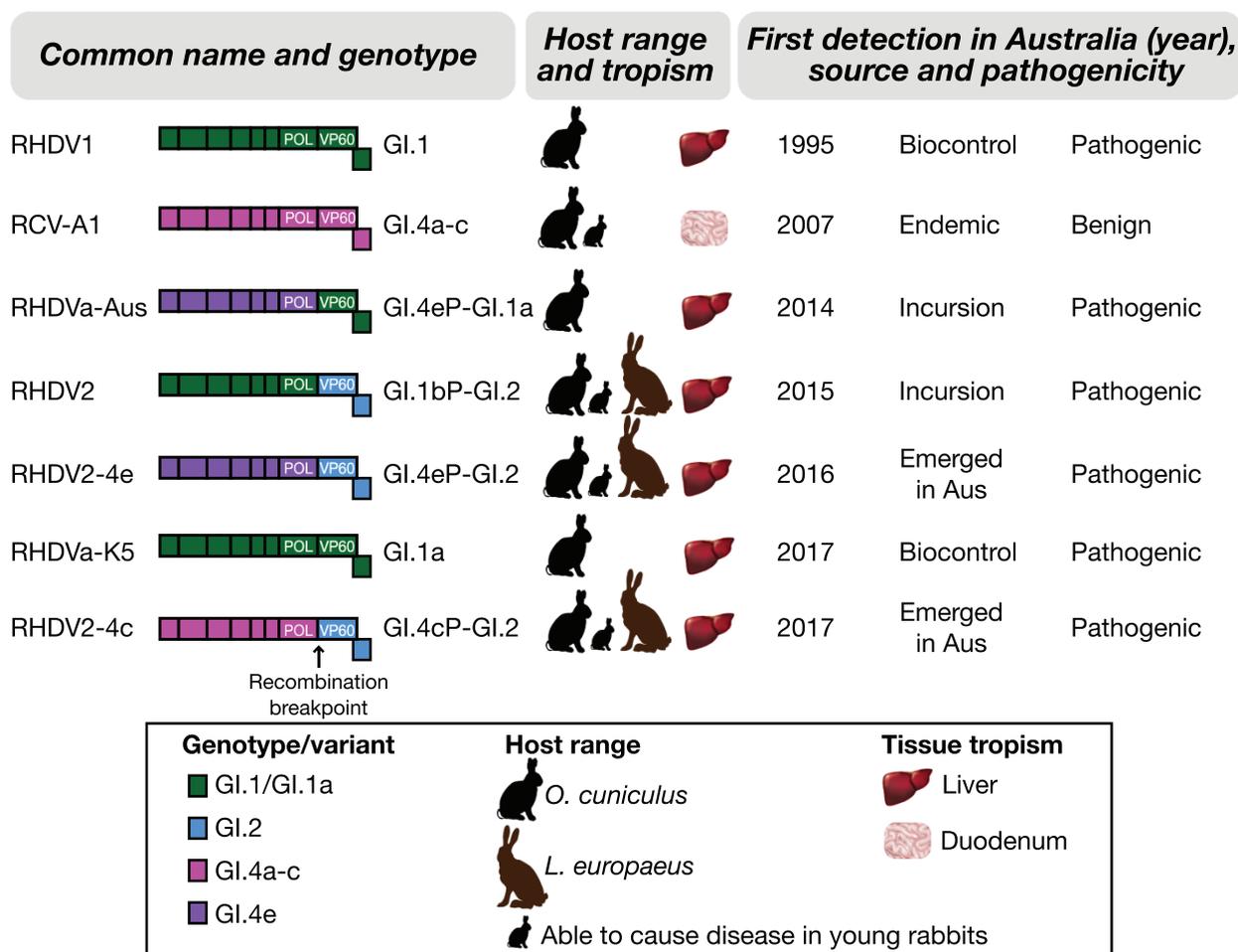
**Figure 2.** GI lagoviruses present in Australia. There are seven known GI lagovirus variants circulating in Australian leporids. Two of these were introduced as biocontrol agents, two were exotic incursions, two emerged in Australia via recombination, and one is benign and endemic in Australian European rabbit populations. Variants differ in host range, pathogenicity, propensity to cause disease in young rabbits, and in the primary site of viral replication (tissue tropism). Protein-coding regions are coloured by the genotype in the genome schematic; recombinant variants are indicated by different coloured non-structural and structural coding segments.

## 2. Materials and methods

### 2.1 Sample submission

Samples from dead domestic and wild lagomorphs (i.e. rabbits and hares) were submitted either directly or via RabbitScan (https://www.feralscan.org.au/rabbitscan/) to the Commonwealth Scientific and Industrial Research Organisation (CSIRO) by veterinarians, pet owners, landholders, and members of the public as part of ongoing opportunistic lagovirus surveillance. No animal ethics approvals are required for sampling rabbits that are found dead in Australia. Samples were provided either fresh-frozen or stored in an RNA stabilisation solution (Hall et al., 2018). Additionally, 42 positive samples submitted to the Elizabeth Macarthur Agricultural Institute for RHDV diagnostic testing were forwarded to CSIRO, either as fresh-frozen tissue or as a swab of the cut tissue stored in phosphate-buffered gelatin saline (PBGS) (Kirkland and Frost 2020). Samples for this study were restricted to those collected between 1 January 2014 (based on the estimated time of incursion of RHDV2 [GI.1bP-GI.2] into Australia [Mahar et al., 2018a; Ramsey et al., 2020]) and 30 September 2020 and were received from the Australian Capital Territory (ACT, $n = 139$), New South Wales (NSW, $n = 401$), the Northern Territory (NT, $n = 7$), Queensland (QLD, $n = 86$), South Australia (SA, $n = 227$), Tasmania

(TAS, $n = 221$), Victoria (VIC, $n = 649$), and Western Australia (WA, $n = 194$) (total $n = 1,924$).

### 2.2 Initial testing

RNA was extracted from 20 to 30 mg of tissue, predominantly the liver or bone marrow (if carcasses were desiccated), but occasionally the spleen, kidneys, muscle, ear tips, eyelids, or blowfly maggots retrieved from the carcass where the liver was not available. After homogenisation with glass beads using a Precellys 24-dual tissue homogenizer (Bertin Technologies), RNA was extracted using either the Maxwell 16 LEV simplyRNA tissue kit (Promega, Alexandria, NSW) or RNeasy mini kit (Qiagen, Chadstone Centre, VIC) as per manufacturers' directions. For swab samples, RNA was extracted from 200 μl PBGS using the Purelink viral RNA/DNA mini kit (ThermoFisher Scientific). Samples were screened for known lagoviruses using a broadly reactive lagovirus SYBR-green-based. This has been amended to reverse-transcription quantitative polymerase chain reaction (RT-qPCR) targeting VP60 and sgRNA, as described previously (Hall et al., 2018).

### 2.3 Variant identification

Of the samples that were positive by screening RT-qPCR ($n = 1,209$), 175 were sequenced previously for other studies (Eden et al., 2015;

Hall et al., 2015; Hall et al., 2018; Mahar et al., 2018a,b). The remaining samples were screened for recombination by RT-PCR and sequencing. RT-PCR primers were designed that spanned the RdRp–VP60 junction, generating a 555-bp amplicon (Rec2 RT-PCR; Supplementary Table S1). Primers were subsequently modified by the 5′-addition of Illumina pre-adapter and index primer-binding sequences to enable high-throughput Illumina sequencing (Rec2.tailed RT-PCR; Supplementary Table S1). Assays were experimentally validated to confirm reactivity with all known Australian lagoviruses. For a subset of samples (prior to the development of the Rec2 RT-PCR), amplification over the RdRp–VP60 junction was performed with alternative primer sets (Supplementary Table S1). Briefly, each 25 μl PCR reaction comprised 1× OneStep Ahead RT-PCR master mix (Qiagen), 1× OneStep Ahead RT mix (Qiagen), 0.5 μM each primer and 1 μl of extracted RNA diluted 1/10 in nuclease-free water. Cycling conditions were: 45 °C for 15 min, 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s, 55 °C for 15 s, 68 °C for 2 min, with a final extension of 68 °C for 5 min.

Amplicons were purified and either sent for Sanger sequencing at the Biomolecular Resource Facility, The John Curtin School of Medical Research, Australian National University, or were indexed, pooled, and sequenced on an Illumina MiSeq (300-cycle v2 chemistry) according to the manufacturer's directions. After quality trimming, sequences were partitioned into regions upstream and downstream of the RdRp–VP60 junction. Each partition was aligned with representative sequences of known Australian lagovirus variants using MAFFTv7.450 (Katoh et al., 2013) in Geneious Prime 2020.0.4, and an RdRp and VP60 variant type was assigned for each sample. Four isolates were identified as mixed infections based on sequencing. Alignments of partial sequences are available at https://doi.org/10.25919/c9cs-b885.

## 2.4 Molecular epidemiological analyses

To investigate the geographical distribution and relative incidence of recombinant variants over time, positive samples were restricted to those reported after 1 January 2016; i.e. 6 months prior to the first detection of a novel recombinant ($n = 1,139$). Samples identified as RHDVa-K5 (GI.1a; $n = 121$) were excluded from further analyses, since these were all associated with release sites of this approved biocontrol virus; that is, this variant was not transmitting through rabbit populations and was therefore not competing epidemiologically with other virus variants that were circulating naturally (and not associated with releases) (Ramsey et al., 2020). Samples from the ACT were grouped with those from NSW, since the ACT is a small jurisdiction (2,358 km$^2$) located within NSW and was considered epidemiologically to be part of NSW.

For the RHDV2-4c (GI.4cP-GI.2) recombinant lineage designation, 169 RHDV2-4c sequences spanning the RdRp–VP60 junction were aligned with 77 RHDV2-4c full genomes and 26 publicly available RCV-A1 (GI.4) genomes using MAFFTv7.450 (Katoh and Standley 2013). The taxa included in the final data set are detailed in Supplementary Table S2. The alignment was trimmed to nucleotide positions 5,115–5,619, based on the M67473.1 DEU/FRG/1988.50 reference sequence (https://www.ncbi.nlm.nih.gov/genbank/). The remaining 109 RHDV2-4c (GI.4cP-GI.2) sequences obtained in this study were not suitable for inclusion in the final alignment because they were generated either using 150-bp paired-end sequencing or with an alternative primer set and thus did not span the complete 504-nt region of the alignment. Model selection and maximum likelihood (ML) tree inference were conducted using IQ-TREEv1.6.11 (Minh et al., 2020), and branch

supported was estimated by 1,000 replicates of the SH-like approximate likelihood ratio test (Guindon et al., 2010). The tree was rooted on the internal branch leading to the RCV-A1 (GI.4) lineage.

## 2.5 Full genome sequencing

Of the 1,034 samples typed to the variant level, a subset of 224 were selected for full genome sequencing (Supplementary Table S3). These samples were selected to be geographically and temporally representative, with a focus on the novel RHDV2-4e (GI.4eP-GI.2) and RHDV2-4c (GI.4cP-GI.2) recombinants. Viral genomes were amplified in overlapping fragments, and DNA libraries were prepared and sequenced using Illumina Miseq technology as described previously (Eden et al., 2015; Mahar et al., 2016). Primers used for the amplification of the overlapping fragments are detailed in Supplementary Table S1. Consensus sequences were constructed by mapping cleaned reads to the lagovirus GI reference sequence (GenBank accession M67473.1/RHDV1/DEU/FRG/1988.50) using Geneious Prime 2020.0.4. Sequences were deposited in GenBank under accession numbers MW460205–MW460242.

## 2.6 Recombination analysis

To further characterise the putative novel RHDV2-4c variant (GI.4cP-GI.2) and the newly sequenced RHDV2-4e (GI.4eP-GI.2) recombinants, recombination analyses were conducted on full genome sequences using the RDP4 software (Martin et al., 2015). The complete genome alignment included lagovirus GI potential parent sequences from GenBank (https://www.ncbi.nlm.nih.gov/genbank) ($n = 384$; 7,309 nt; Supplementary Table S4). Recombination was considered *bona fide* if detected by at least two of three primary scanning methods (RDP, MaxChi and GENECONV), with a highest acceptable *P*-value of 0.05 with Bonferroni multiple comparison correction. The BootScan, 3Seq, CHIMAERA, and SciScan methods were used to verify recombination events identified using the primary methods.

## 2.7 Phylogenetic analyses

Newly sequenced lagovirus genomes were aligned with representative GI lagovirus sequences from GenBank (https://www.ncbi.nlm.nih.gov/genbank) using MAFFTv7.271 (Katoh and Standley 2013). The complete genome alignment ($n = 240$ sequences; Supplementary Table S5) was split into two data sets representing (1) the NS genes (5,231 nt) and (2) the VP60 structural gene (1,740 nt). An additional VP60 alignment was constructed that contained the VP60 of the newly sequenced viruses along with all published Australian RHDV2 (GI.2) VP60 sequences ($n = 332$; 1,737 nt; Supplementary Table S6). Model selection and ML tree estimation was performed using IQ-TREEv1.6.11 (Minh et al., 2020), as described above. Ancestral state reconstruction to identify non-synonymous changes along individual branches was carried out using the ML method in PAMLv4.9 (Yang 2007).

## 2.8 Phylogeographic analyses

A Bayesian Markov chain Monte Carlo (MCMC) approach was employed to infer time-scaled phylogenies, and these were used to infer the temporal pattern and most likely geographic location of internal nodes (i.e. phylogeography). All data sets were initially screened using TempEst (Rambaut et al., 2016) to ensure that sufficient temporal signal was present, using ML phylogenies as input to construct linear regressions of root-to-tip genetic distances against the sampling time (Supplementary Fig. S1). Due to recombination between the NS and VP60 genes, two data sets either side of the recombination breakpoint were analysed separately: (1) a

VP60 gene data set containing all published Australian GI.2 VP60 sequences (RHDV2) together with the VP60 genes of the recombinants sequenced here ($n = 332$; 1,737 nt; Supplementary Table S6) and (2) an NS gene data set of all published Australian GI.4 (RCV-A1 and RHDVa-Aus) sequences together with the NS genes of the recombinants sequenced here ($n = 188$; 5,218 nt; Supplementary Table S7). The Bayesian Evolutionary Analysis by Sampling Trees (BEAST) software v1.8 (Drummond and Rambaut 2007) was used to conduct Bayesian MCMC analysis of each data set, using substitution models inferred by ModelFinder as implemented in IQ-TREEv1.6.11 (VP60, SYM + G4; NS, SYM + I + G4) (Zharkikh 1994; Minh et al., 2020). A discrete trait partition indicating the sampling location (Australian state) was included to facilitate the ancestral state reconstruction of location (utilizing the symmetric substitution model, inferring the social network with Bayesian Stochastic Search Variable Selection, and a strict clock model). All analyses were run twice to convergence (defined as an effective sample size >200) to confirm consistency. Marginal likelihood estimation using path sampling/stepping-stone sampling was used to assess the most appropriate clock model prior (strict vs uncorrelated log-normally distributed [UCLD]) and tree prior (Gaussian Markov random field Bayesian skyride model vs constant size coalescent vs exponential coalescent) for the nucleotide partition.

The NS data set had a larger sampling window and a better temporal signal ($R^2$ value of 0.86 on linear regression of the root-to-tip genetic distance against the sampling time) compared to the VP60 data set ($R^2$ value of 0.64) and was therefore considered the most informative (linear regressions presented in Supplementary Fig. S1). The UCLD clock model and constant population size coalescent tree prior had the highest marginal likelihood for this data set.

Figures 1–6 and S2 were constructed using Inkscape 1.1 and/or R v4.0.3 (R Core Team 2020) using the following packages: 'ggtree v2.2.1' (Yu et al., 2017), 'treeio v1.12.0' (Wang et al., 2020), 'scales v1.1.1' (Wickham and Seidel 2020), 'tidyverse v1.3.0' (Wickham et al., 2019), 'ape v5.4' (Paradis and Schliep 2019), 'phytools v0.7–47' (Revell 2012), 'phangorn v2.5.5' (Schliep 2011), 'cowplot v1.0.0.9000' (Wilke 2020), 'lubridate v1.7.8' (Grolemund and Wickham 2011), 'plyr v1.8.6' (Wickham 2011), and 'ozmaps v0.3.6' (Sumner 2020). Figure 5B was generated in GeneiousPrime 2020.2.4 using the 3-D structure viewer.

## 3. Results

### 3.1 Phylogenetic analyses reveal five independent recombination events in RHDV2-4c viruses

A recombinant lagovirus variant, RHDV2-4e (GI.4eP-GI.2), consisting of RHDVa-Aus NS genes (GI.4eP) and RHDV2 structural genes (GI.2), was previously detected in Australia using a multiplex RT-PCR assay. As this assay is only able to detect specific recombinant variants (Hall et al., 2018), to comprehensively screen for lagovirus recombination events we sequenced the RdRp–VP60 junction region (i.e. the typical recombination breakpoint) of 1,034 lagoviruses collected between January 2014 and September 2020 (inclusive) from wild and domestic lagomorphs found dead across Australia. This screen identified six distinct lagovirus variants circulating in Australia during the study period; four known variants: RHDV1 (GI.1cP-GI.1c, $n = 18$), RHDV2 (GI.1bP-GI.2, $n = 390$), RHDVa-K5 (GI.1aP-GI.1a, $n = 121$), RHDVa-Aus (GI.4eP-GI.1a, $n = 12$); the previously reported RHDV2-4e recombinant (GI.4eP-GI.2, $n = 133$); and a novel putative recombinant designated RHDV2-4c (GI.4cP-GI.2, $n = 356$); along with four mixed

infections. The novel RHDV2-4c recombinant comprised RHDV2 structural genes (GI.2) and non-structural genes from the benign RCV-A1 (GI.4cP).

We selected a subset of novel RHDV2-4c recombinant viruses ($n = 77$), recently emerged RHDV2-4e recombinant viruses ($n = 53$), and parental RHDV2 (GI.1bP-GI.2) viruses ($n = 73$) for full genome sequencing, selecting isolates that were representative of the temporal and geographical spread of these variants (Supplementary Table S3). Recombination analyses of newly sequenced RHDV2-4e viruses confirmed these as recombinants with a breakpoint at the RdRp–VP60 junction, supporting our previous finding (Table 1, Supplementary Table S4) (Hall et al., 2018). Recombination analyses of the novel putative RHDV2-4c viruses also detected a recombination breakpoint at the typical RdRp–VP60 junction (Table 1). The most likely parental variants for RHDV2-4c, as determined by the RDP4 software, were benign Australian RCV-A1 (GI.4c) lagoviruses in the NS region (GenBank accession KX357699) and Australian RHDV2 (GI.2) viruses in the structural region (GenBank accession MF421577), strongly suggesting that this recombination event occurred within Australia after the arrival of RHDV2 (GI.1bP-GI.2) in the late 2014. Notably, both Australian recombinant variants contained non-structural genes related to benign RCV-A1-like viruses of the GI.4 clade.

Subsequently, the nucleotide alignment was partitioned either side of the putative breakpoint and ML phylogenetic trees were inferred for the NS genes and the VP60 gene separately. There was clear incongruence between the NS and VP60 phylogenies, supporting the results of the RDP4 analysis (Fig. 3). Strikingly, ML phylogenies based on the VP60 region revealed five distinct lineages of RHDV2-4c, suggesting multiple independent emergences of this variant (Fig. 4A). In contrast, the NS gene sequences from this variant formed a monophyletic group (collapsed in Fig. 3A). All RHDV2-4e sequences fell within a single clade in both trees (Figs 3 and 4).

### 3.2 Rapid epidemiological replacement of parental RHDV2 by the newly emerged RHDV2-4c and RHDV2-4e recombinants

Detections of the new RHDV2-4c (GI.4cP-GI.2) recombinant variant consistently outnumbered those of the parental RHDV2 (GI.1bP-GI.2) in VIC by April 2018, 14 months after the first detection in Alvie, VIC (5 February 2017), suggesting that this variant was outcompeting previously circulating viruses (Fig. 5). This pattern of emergence and replacement was also observed in TAS (Fig. 5). This variant was detected sporadically in NSW/ACT from October 2017 where it co-circulated with the locally dominant RHDV2-4e (GI.4eP-GI.2) recombinant. Notably, the proportional incidence of RHDV2-4c has increased in this region during 2020. Interestingly, a single detection of RHDV2-4c was identified in WA in November 2018 in two domestic rabbits recently imported from TAS as the breeding stock. The sequence of this WA virus clearly nested within the genetic diversity found in TAS (Fig. 5A, lineage v), suggesting a single incursion into WA with no ongoing local transmission. During the study period, the novel RHDV2-4c was recovered from several juvenile animals less than 800 g body-weight and as young as 4-weeks old, demonstrating pathogenicity in young rabbits, like all other variants characterized so far with a GI.2 VP60 gene (Fig. 2). It was also identified in one European brown hare (*Lepus europaeus*).

We also examined the distribution of the previously reported RHDV2-4e recombinant variant (Hall et al., 2018). Most detections ($n = 140/151$) were identified in NSW/ACT. The earliest detection
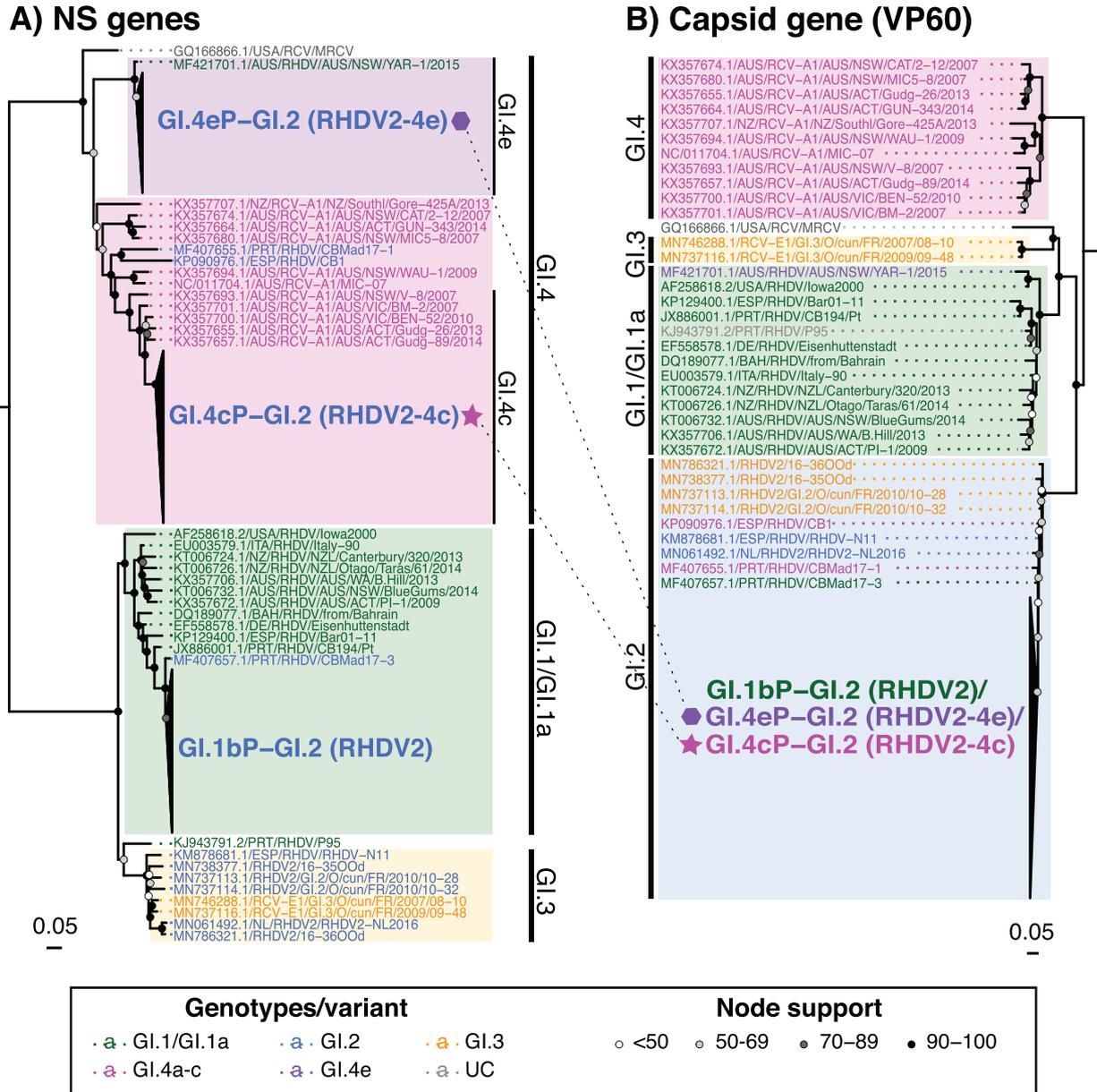
**Figure 3.** ML phylogenetic trees of the (A) NS genes and (B) the VP60 gene of representative lagovirus sequences and the Australian recombinant sequences. Discrepancies between the highlight and text colour indicate recombinant sequences. Clades are highlighted according to the genotype of the genetic segment analysed in the tree, while taxa names are coloured by the genotype of the alternative segment (i.e. taxa names in the NS gene tree are coloured by the genotype of the VP60 gene and vice versa). UC, unclassified. Major clades have been collapsed for visualisation purposes. All Australian recombinant sequences from this study fall within the collapsed clades indicated by the star (GI.4cP-GI.2) or hexagon (GI.4eP-GI.2) symbols. The dotted lines between the two trees link the Australian recombinant clades. Phylogenies were midpoint rooted and the scale bars represent the number of substitutions per site. Bootstrap support values (1,000 replicates) are indicated by shaded circles at the nodes. Associated metadata including taxon names are provided in Supplementary Table S5.

was in Tubbul, NSW, on 13 July 2016, and by March 2017 it had mostly replaced the parental RHDV2 (GI.1bP-GI.2) to become the dominant lagovirus variant in this region (Fig. 5B). More recently, in the late 2019 and throughout 2020, it has co-circulated with RHDV2-4c in NSW/ACT. Although most detections were in adult rabbits, RHDV2-4e was also recovered from three European brown hares and from several juvenile rabbits down to 250 g bodyweight, suggesting a similar host tropism to the parental RHDV2 (GI.1bP-GI.2) virus and the novel RHDV2-4c recombinant (Fig. 2).

## 3.3 All recombination events occurred within a 2-year timeframe in eastern Australia

To determine the temporal dynamics, the location of emergence, and the rate of viable recombination events in Australian lagoviruses, we conducted time-structured phylogenetic analyses for the NS and VP60 genes of recombinant viruses and related sequences (i.e. Australian genotype GI.2 VP60 and genotype GI.4 NS sequences). Regression analysis indicated that data from the GI.4 NS data set were more accurate, and therefore, this is presented in Fig. 5. Based on these analyses, RHDV2-4e (GI.4eP-GI.2)
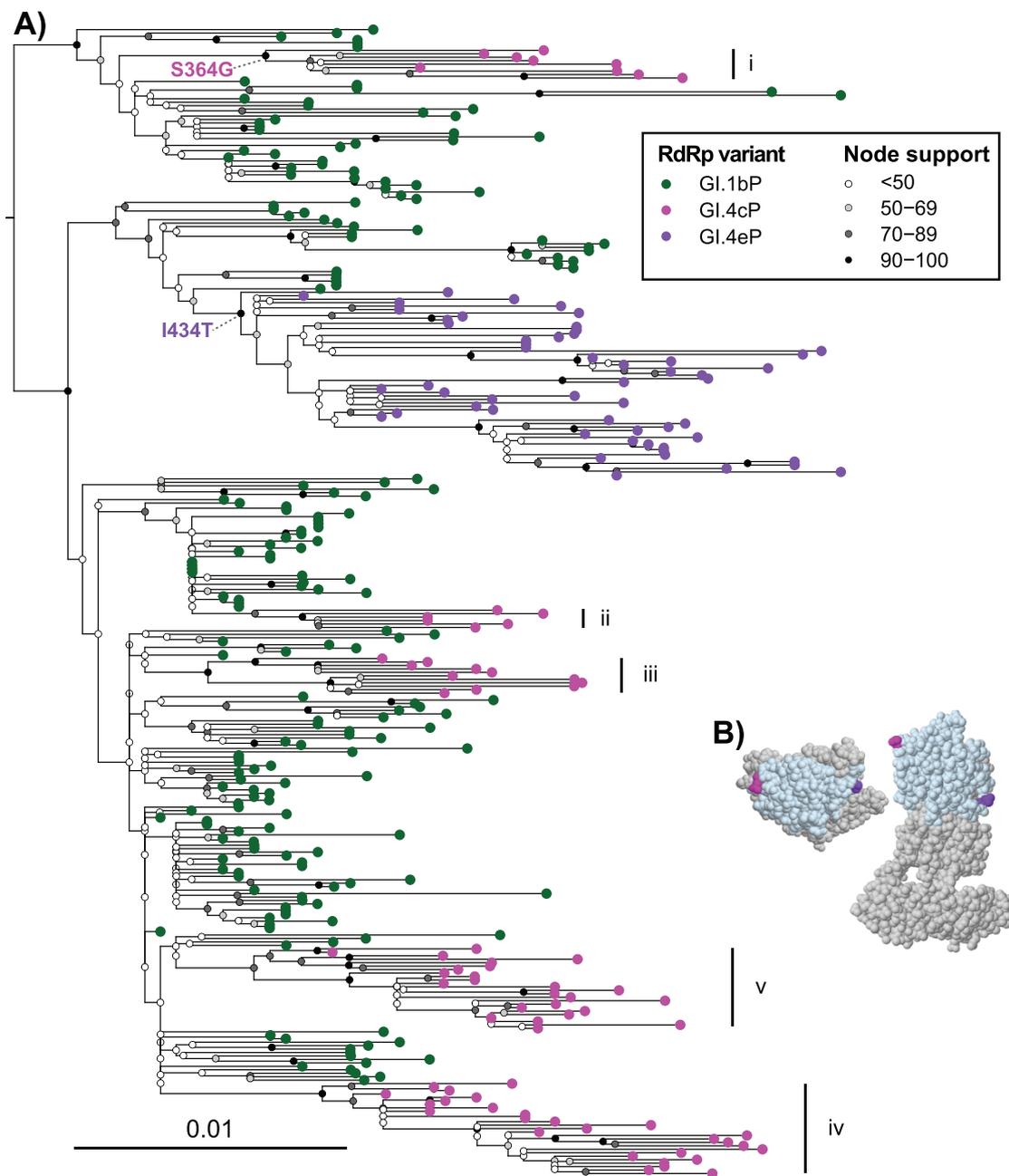
**Figure 4.** ML phylogenetic tree of Australian GI.2 VP60 sequences and amino acid changes associated with the emergence of RHDV2-4c and RHDV2-4e recombinant variants. (A) A phylogeny was inferred from the VP60 gene sequences ($n = 332$) of newly sequenced Australian GI.2 lagoviruses (including recombinants) along with representative published sequences. Coloured tip points represent the polymerase (RdRp) genotype: GI.1bP, green; GI.4cP, pink; GI.4eP, purple. GI.4cP lineages are labelled (i through v). The phylogeny was midpoint rooted and the scale bar represents nucleotide substitutions per site. Bootstrap support values (1,000 replicates) are indicated by shaded circles at the nodes. Associated metadata including taxon names are provided in Supplementary Table S6. All reconstructed amino acid changes occurring on branches leading to the GI.4cP and GI.4eP lineages are labelled at the relevant node. (B) These changes were mapped to a GI lagovirus VP60 monomer (PDB accession: 3J1P). Both a top-down view (left) and a side view (right) of the monomer is shown. The P2 domain is coloured in light blue. Amino acid 364, associated with the change in RHDV2-4c Lineage i, is coloured pink and 434, associated with the RHDV2-4e, is coloured purple.

was the first variant to emerge, with the most recent common ancestor of this lineage dated to early to mid-2015 (95 per cent highest posterior density [HPD] NS dataset, 2014.8–2015.4) (Fig. 5A). For RHDV2-4c (GI.4cP-GI.2) recombinants, the 95 per cent HPD intervals of the time to most recent common ancestor (TMRCA) overlapped between lineages, making it difficult to confidently define the exact timings and order of emergence of each lineage (Fig. 5A). However, the data clearly show that these five recombination events all occurred within the space of 2 years. This

2-year timeframe was consistently observed regardless of whether the analysis was conducted on the NS genes GI.4 data set or the VP60 gene GI.2 data set. Our phylogeographic analysis (Fig. 5A) suggests that all the RHDV2-4c recombinant lineages emerged in VIC (probability >0.99 for all lineages), with subsequent spread to other states. Notably, this was observed even for lineage v, which was first detected in Tasmania. In contrast, RHDV2-4e clearly emerged in NSW/ACT (probability >0.99, Fig. 5A); this is further supported by our epidemiological data, where we found
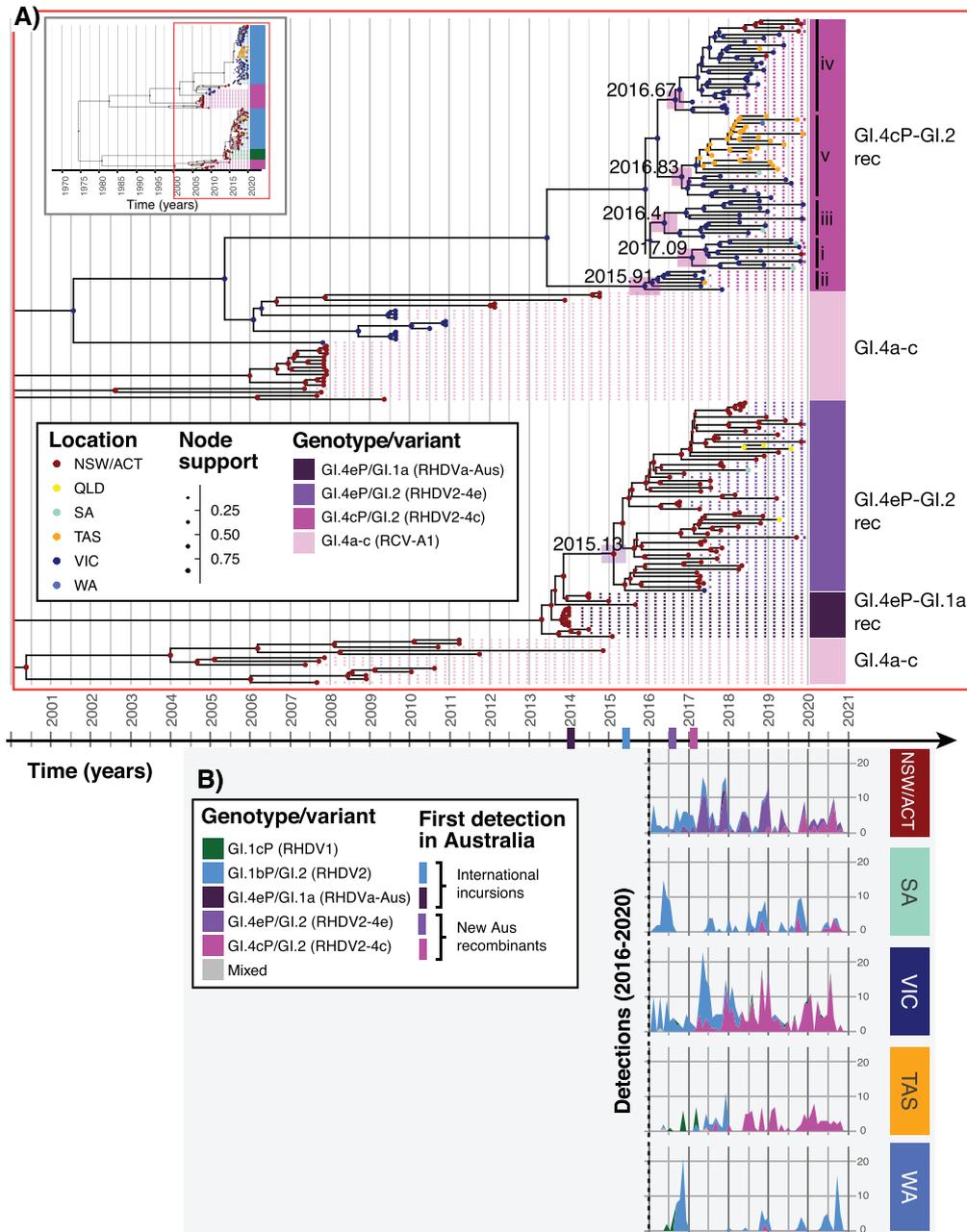
**Figure 5.** Timing and location of emergence of the Australian lagovirus recombinants based on time-structured phylogeographic reconstruction and diagnostic detections. Panels A and B are scaled to the same x-axis (time in years, shown between the two panels). (A) Time-structured phylogeographic analysis of Australian genotype GI.4 lagovirus NS genes (UCLD clock model, constant tree prior). The inset indicates the section of the phylogeny shown. Node and tip points are coloured according to the sampling location (Australian state) and node points are sized according to posterior support for that clade. Genotype/variant of each clade is indicated by clade labels (rec = recombinant) and coloured blocks beside the labels. The RHDV2-4c (GI.4cP-GI.2) lineages (i–v) are also labelled. The mean TMRCA for the RHDV2-4e (GI.4eP-GI.2) and each RHDV2-4c (GI.4cP-GI.2) recombinant lineage is indicated at the respective internal node, and horizontal coloured bars at these nodes represent the 95 per cent HPD. (B) Lagovirus-positive samples collected in NSW/ACT, SA, VIC, TAS, and WA from 2016 to 2020 (n = 739) were genotyped to the variant level by sequencing either side of the typical calicivirus recombination breakpoint. The number of detections of each variant by month are shown for each geographical region as an area plot, with the plotted area coloured by variants.

only limited detections of this variant in other states (QLD, VIC, and SA).

## 3.4 Continued co-circulation of RHDV2-4c recombinant lineages

The RHDV2-4c (GI.4cP-GI.2) recombinant viruses were initially assigned to lineages based on the placement of full genome sequences within the VP60 phylogeny (Fig. 4A). For RHDV2-GI.4c

viruses where a full genome sequence was not available, we first inferred a phylogeny from a 504-nt region spanning the RdRp–VP60 junction (Fig. 6B); this phylogeny was annotated using taxa for which the lineage was definitively known from the full-length VP60 phylogeny. While sequences from each RHDV2-4c lineage (as defined from the complete genome sequences) did not form monophyletic clades in the RdRp–VP60 junction phylogeny, they did form distinct groups with visible genetic distance between
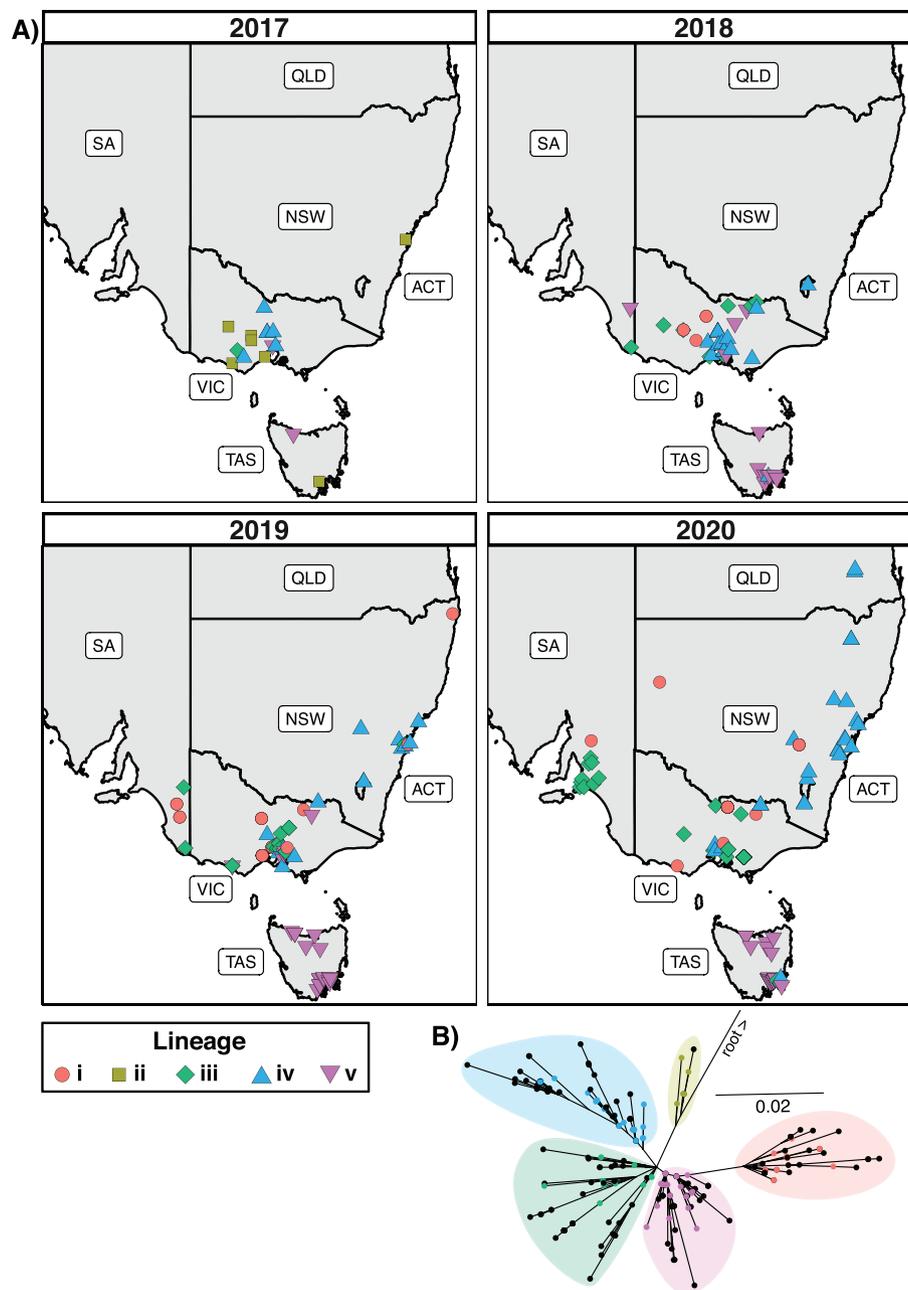
**Figure 6.** Geographic distribution of RHDV2-4c recombinant lagoviruses in Australia, 2017 to 2020. (A) Sampling locations for RHDV2-4c (GI.4cP-GI.2) lineages were mapped for each year following the emergence of this variant in 2017. (B) RHDV2-4c viruses were allocated to a lineage based on the ML phylogenetic grouping of the partial RdRp–VP60 sequence (504 nt) with reference sequences. Reference sequences were annotated based on full genome sequencing and are indicated in the tree by the coloured tip points. The phylogeny was rooted on the internal branch leading to the RCV-A1 (GI.4) lineage and the scale bar represents nucleotide substitutions per site.

**Table 1.** Recombination analyses.

| Variant | | Putative parent[b] | | P-value for recombination event by detection method | | |
|---|---|---|---|---|---|---|
| Representative recombinant sequence | Breakpoint position[a] | NS parent | S parent | RDP | Maxchi | GENE-CONV |
| GI.4eP-GI.2/O.cun/AUS/ NSW/2017-03-17/CAR-8 | 5306 | KY628315/ GI.4eP-GI.1a | MF421585/ GI.1bP-GI.2 | 4E-194 | 3E-43 | 3E-189 |
| GI.4cP-GI.2/O.cun/AUS/ VIC/2017-02-05/ALV-1 | 5307 | KX357699/ GI.4c | MF421577/ GI.1bP-GI.2 | 3E-148 | 2E-39 | 1E-154 |

A more detailed version of this table is supplied as Supplementary Table S4.
[a]Nucleotide position of the putative breakpoint according to M67473.1/DEU/FRG/1988.50 reference sequence numbering.
[b]NS refers to putative parent providing the non-structural genes; S refers to putative parent providing the structural genes.

them (Fig. 6B). These groups were distinct enough to assign an RHDV2-4c lineage type to those samples with partial sequence only. After assigning lineages to all RHDV2-4c recombinants, we explored the interactions between the different lineages by mapping their sampling locations over time. We found that four of the five lineages were already circulating in 2017 (Fig. 6). Lineage ii was not detected after 2017 while the other four lineages continued to be detected throughout 2018, 2019, and 2020. In VIC, where all lineages are circulating, there was no clear indication of geographical clustering of the different lineages. In TAS, Lineages ii, iii, and iv were detected sporadically while lineage v has remained dominant over time. Similarly, in NSW/ACT Lineage iv has remained dominant, while Lineages i, ii, and iii were detected intermittently.

## 3.5 The emergence of RHDV2-4c recombinant variants was not associated with antigenic changes in the VP60 major capsid protein

To determine whether the observed epidemiological replacement of the parental RHDV2 (GI.1bP-GI.2) virus by (1) the novel RHDV2-4c (GI.4cP-GI.2) recombinants in VIC and TAS and (2) the RHDV2-4e (GI.4eP-GI.2) recombinant in NSW/ACT was associated with non-synonymous, potentially antigenic, changes in the VP60 major capsid protein, we performed ancestral state reconstruction using our previous Australian GI.2 VP60 phylogeny. Notably, four of the five RHDV2-4c recombinant lineages (classified based on nucleotide sequences) showed no lineage-defining amino acid changes in VP60 relative to the common ancestor of all Australian GI.2 VP60 sequences, suggesting that the emergence of these lineages was not associated with changes in antigenicity (Fig. 4A). However, lineage-defining non-synonymous mutations were identified in the RHDV2-4c Lineage i and in the RHDV2-4e recombinant viruses (Fig. 4A). When we mapped these mutations back to an atomic model of lagovirus GI VP60, both mutations were within the protruding domain, specifically the P2 subdomain that is known to contain immunodominant epitopes (Fig. 4B) (Wang et al., 2013). The RHDV2-4c Lineage i was associated with an S364G mutation, while an I434T mutation was present in all RHDV2-4e recombinants (Fig. 4A). Thus, it is possible that the epidemiological fitness of RHDV2-4e in NSW/ACT may be attributable to antigenic changes in the capsid; however, this cannot explain the replacement of parental RHDV2 (GI.1bP-GI.2) by RHDV2-4c variants in VIC or TAS, or the continued co-circulation of RHDV2-4c lineages without changes in VP60. Furthermore, in the NSW/ACT the RHDV2-4e recombinant variant, despite the I434T change in VP60, appears to be currently undergoing replacement by RHDV2-4c. Ancestral state reconstruction using a phylogeny constructed on the NS coding sequences showed multiple amino acid changes between the different Australian RHDV2 variants within all mature peptide-coding regions (Supplementary Fig. S2).

## 4. Discussion

In a global landscape where population immunity to RHDV1 (GI.1) and RHDVa (GI.1a) viruses was widespread in lagomorphs, it is unsurprising that RHDV2 (GI.1bP-GI.2) viruses were so epidemiologically successful (Mahar et al., 2018a). We were interested to determine whether, after establishment, RHDV2 (GI.1bP-GI.2) viruses would continue to evolve via immune selection and antigenic drift or by other mechanisms. We were also interested to document whether variants without the GI.2 capsid would emerge as population immunity increased against this genotype. Instead, we found that the GI.2 structural genes were retained and that

the emerging variants acquired alternative NS genes via recombination. It is becoming increasingly clear that recombination is a major driver of calicivirus diversity, facilitating the emergence of new variants including those with pandemic and panzootic potential (Eden et al., 2013; Ludwig-Begall, Mauroy, and Thiry 2018; Parra 2019; Abrantes et al., 2020). However, the frequency at which viable recombination occurs and the genetic drivers of epidemiological fitness are still poorly understood. In this study, a continent-scale natural competition experiment of lagoviruses in Australia, we found that viable recombination occurs at an extremely high frequency given the right agent, host, and environmental circumstances.

### 4.1 Six viable recombination events in Australian lagoviruses in a two-year period

Shortly after detecting two exotic lagovirus incursions in Australia, RHDVa-Aus (GI.4eP-GI.1a) in 2014 and RHDV2 (GI.1bP-GI.2) in 2015 (Hall et al., 2015; Mahar et al., 2018b), we reported the emergence of a recombinant variant of these two viruses, RHDV2-4e (GI.4eP-GI.2), detected in July 2016 in NSW (Hall et al., 2018). This finding prompted the development of improved molecular tests for the detection of recombination, and the subsequent retrospective and prospective screening of lagovirus-positive samples for further recombination events. This screen identified an additional novel recombinant lagovirus, designated RHDV2-4c (GI.4cP-GI.2), first detected in VIC in February 2017. Further analysis determined that the VP60 genes of RHDV2-4c separated into five distinct clades (Fig. 4A), suggesting that this variant emerged on separate occasions following at least five independent recombination events. Recombination was already known to occur in lagoviruses. For example, in the last decade, the lagovirus GI.2 capsid has been detected in combination with five different sets of NS genes (Lopes et al., 2015; Hall et al., 2018; Abrantes et al., 2020; Szillat et al., 2020). Our study demonstrates that viable recombination during natural infection occurs even more frequently than previously appreciated but may often go undetected if different events result in closely related progeny. Since the first detection of RHDV2 (GI.1bP-GI.2) in Australia in mid-2015 (Hall et al., 2015), these structural genes have recombined and produced epidemiologically viable variants at least six times in a 2-year period.

The previously reported RHDV2-4e recombinant arose from a recombination event between RHDVa-Aus (GI.4eP-GI.1a) and RHDV2 (GI.1bP-GI.2), both recombinant viruses themselves that were exotic incursions detected in Australia in January 2014 and May 2015, respectively, although the source of these viruses remains unknown (Hall et al., 2015; Hall et al., 2018; Mahar et al., 2018a,b). Detections of the parental RHDVa-Aus (GI.4eP-GI.1a) were predominantly restricted to the Sydney basin region of NSW, with two cases detected in QLD and one further case detected in a wild rabbit in the ACT (Mahar et al., 2018b). Therefore, it is not surprising that the RHDV2-4e (GI.4eP-GI.2) recombinant emerged in NSW/ACT, according to our epidemiological data and phylogeographic analysis (Fig. 5). The RHDV2-4c (GI.4cP-GI.2) recombinants arose from recombination events between RHDV2 (GI.1bP-GI.2) and endemic RCV-A1 (GI.4c) viruses (Strive, Wright, and Robinson 2009; Hall et al., 2015; Mahar et al., 2016). Emergence of the RHDV2-4c recombinant variants in the south-eastern states of Australia is consistent with high rabbit population densities and a high seroprevalence to RCV-A1 (GI.4) viruses in these regions compared to more arid regions of Australia (Liu et al., 2014).

Despite the VP60 phylogeny suggesting multiple recombination events (Fig. 4A), only a single unique recombination event

was detected by the RDP4 program (Table 1). This is not surprising given that the breakpoint locations are indistinguishable, and these recombinants (and their true parents) are very closely related. Indeed, it was only possible to infer a more accurate number of recombination events because of the high sampling frequency of parental RHDV2 VP60 sequences (GI.2), which has been maintained for this and previous studies. Had the genetic diversity of parental RHDV2 (GI.1bP-GI.2) VP60 sequences been undersampled, the RHDV2-4c recombinant VP60 sequences would have formed a monophyletic clade and a single recombination event assumed. This is demonstrated by the NS phylogeny (Fig. 5A), in which the RHDV2-4c recombinant NS sequences form a single clade within the greater GI.4c clade. Importantly, within this clade the sequences still group into the same five lineages as seen in the VP60 tree (Supplementary Fig. S2). Compared to pathogenic RHDV2 (GI.2) viruses, benign RCV-A1 (GI.4) viruses are greatly under-sampled in Australia, with the most recent GI.4 sequences (n = 6) derived from samples collected from the ACT in 2012 and 2014 (Mahar et al., 2016). Aside from three sequences sampled in 2014, no other GI.4 (RCV-A1) sequences were available for the entirety of the study period. Consequently, a close relative of the GI.4cP parental virus was not identified, suggesting a hidden diversity of unsampled GI.4 viruses in Australian rabbits. This study provides a lower bound to the rate of recombination in these viruses; it is possible that additional recombination events have occurred with RCV-A1 (specifically variant GI.4c) viruses that are not observable against the current background of available GI.2 VP60 sequences.

## 4.2 Lagoviruses show frequent recombination between the GI.2 capsid and GI.4 NS sequences

Interestingly, all the recombination events detected during the study period involved the pairing of GI.2 (RHDV2) structural genes and the NS genes of a GI.4 (RCV-A1-like) virus. Despite screening over 800 lagovirus-positive samples, we found no evidence of recombination events involving either structural or NS sequences of GI.1 viruses (beyond the original GI.1bP-GI.2 and GI.4eP-GI.1a incursion variants).

For recombination to occur, several criteria must be met (Worobey and Holmes 1999). First, there must be co-infection of an individual host. This is influenced by host tropism, the prevalence of each parental virus in the population, and the duration of infection. Second, there must be co-infection of a single cell, both viruses must replicate within this cell, and precise template switching must occur to generate viable gRNA. Finally, the resultant variant must be epidemiologically competitive. That is, it must be able to successfully transmit, to establish infection in new hosts, and to outcompete both the parental and other circulating variants.

Following the epizootic incursion of an antigenically novel virus, RHDV2 (GI.1bP-GI.2), into a naïve population, the prevalence of this variant was extremely high in Australian rabbits, reflected by the estimated population-wide mortality rate of 60 per cent following its arrival (Ramsey et al., 2020). The detection of four RHDV1 (GI.1c) and RHDV2 (GI.1bP-GI.2) co-infections in this study between 2015 and 2017 demonstrates that both viruses were circulating at sufficiently high prevalence for mixed infections to occur. However, this does not explain the predilection for RCV-A1-like (GI.4) NS sequences. The discovery of RHDV2-4c (GI.4cP-GI.2) recombinants was surprising, since the parental RCV-A1 (GI.4c) is a benign enterotropic lagovirus (Strive, Wright, and Robinson 2009), while RHDV2 (GI.1bP-GI.2) and RHDV1 (GI.1) viruses are virulent and hepatotropic (Fig. 2). This would seem to preclude

co-infection of individual cells with both viruses. Yet, several recombinants of RHDV2 (GI.2 structural sequences) and benign enterotropic RCVs (GI.3 NS sequences) have also been reported from Europe (Lopes et al., 2015, 2019; Dalton et al., 2018; Silverio et al., 2018; Lopes et al., 2019; Abrantes et al., 2020). This suggests that active replication of both RCVs and RHDV2 must be occurring in the same target cell. Extrapolating from our current understanding of human and murine norovirus tropism, macrophages or other immune cells may be likely candidates (Wobus et al., 2004). This is further supported by the detection of plus- and minus-strand viral RNAs in splenic and alveolar macrophages of rabbits experimentally infected with RHDV1 (Kimura et al., 2001). Additionally, RCVs (GI.3 and GI.4) typically infect young rabbits early in life (Capucci, Nardin, and Lavazza 1997; Strive, Wright, and Robinson 2009). Although young rabbits can be infected with RHDV1 (GI.1) viruses, robust innate immunity limits the extent of viral replication in an age-dependent manner (Neave et al., 2018). Thus, mixed infections with RHDV1 and RCVs are probably infrequent. This age-dependent resistance is not observed with RHDV2 (GI.2) infection (Neave et al., 2018). Furthermore, the duration of infection is longer for benign RCVs compared to pathogenic variants, where infected individuals typically die within 48–72 h post-infection (Abrantes et al., 2012). These factors may at least partly explain why RCV-RHDV2 recombinants appear to emerge more frequently than RHDV1-RHDV2 recombinants.

## 4.3 Epidemiological drivers of lagovirus emergence and spread

Both the RHDV2-4e (GI.4eP-GI.2) and RHDV2-4c (GI.4cP-GI.2) recombinants rapidly replaced the dominant circulating parental RHDV2 (GI.1bP-GI.2) in NSW/ACT and VIC/TAS, respectively. This replacement, at least for several RHDV2-4c lineages, occurred without any associated antigenic changes in the VP60 capsid protein, demonstrating that the replacement was not driven by antigenic escape. This shows that NS sequence variation is an important driver of epidemiological fitness in lagoviruses, complementing similar findings in human and murine noroviruses (Ruis et al., 2017; Tohma et al., 2017; Ao et al., 2018). For example, the pandemic GII.P16/GII.4 Sydney 2012 norovirus, which does not contain unique substitutions in the capsid, has substitutions within the RdRp that were proposed to increase transmissibility (Ruis et al., 2017). This same GII.P16 P-type (with a different GII.2 capsid) was also shown to have mutations on the surface of the polymerase that correlated with an increased evolutionary rate (Ao et al., 2018). Yet in contrast, another study found that the evolutionary rate does not vary between genotypes or across genomic regions (Tohma et al., 2021). Infection with the norovirus GII.P16-GII.2 variant also resulted in increased viral shedding compared to other norovirus genotypes, as measured by higher faecal viral loads (Cheung et al., 2019). Previous studies have shown that RdRp fidelity and intra-host viral diversity affect transmissibility of murine norovirus *in vivo*, with high-fidelity variants being less efficiently transmitted than the wild-type variant (Arias et al., 2016). Taken together, these studies demonstrate that NS proteins, particularly the RdRp, are important drivers of calicivirus fitness. Interestingly, the *in vitro* polymerase replication activity of the cloned RCV-A1 (GI.4c) RdRp was previously shown to be at least two times that of the RHDV1 (GI.1c) RdRp (Urakova et al., 2016), although comparison to the GI.1b RdRp, the specific 'competing' variant in the current study, was not reported. Analogous to findings in murine norovirus, we propose that a higher replication rate of the RCV-A1 (GI.4) RdRp may lead to increased intra-host viral diversity and transmissibility of GI.4 recombinants

(Arias et al., 2016). However, we cannot rule out that other NS proteins may also contribute to the observed high fitness of the recombinant variants, given the many non-synonymous changes present throughout the NS-coding region (Supplementary Fig. S2).

Within the RHDV2-4c (GI.4cP-GI.2) recombinants, there was no evidence of dominance of any one lineage over time, demonstrating that different lagoviruses can co-circulate at relative equilibrium over extended periods. The apparent dominance of Lineage iv in NSW/ACT and lineage v in TAS is consistent with founder effects in both of these states, where viruses from competing lineages without substantial fitness benefits were not able to outcompete pre-existing dominant variants. Onward transmission and early virus establishment would be highly stochastic. For example, infections in large meat or show rabbitries would have a much higher probability of onward transmission than an infection in a single pet rabbit. Hence, early dominance cannot be attributed to enhanced fitness without evidence of replacement. Given the observed rapid replacement of parental RHDV2 (GI.1bP-GI.2) by GI.4P recombinant variants in this study, this further supports a fitness advantage of these variants, conferred by the NS proteins and possibly associated with the RdRp.

## 4.4 The lagovirus structural proteins govern host and tissue tropism and are correlated with virulence

Both the newly identified RHDV2-4c (GI.4cP-GI.2) recombinant and the previously emerged RHDV2-4e (GI.4eP-GI.2) recombinant are virulent, hepatotropic viruses that were recovered from the livers of both rabbits and hares and from rabbits of all ages. This tropism mimics that seen with other RHDV2 (GI.2) viruses (Fig. 2) (Puggioni et al., 2013; Camarda et al., 2014; Dalton et al., 2014; Hall et al., 2017; Le Gall-Recule et al., 2017; Velarde et al., 2017; Neimanis et al., 2018; Buehler et al., 2020). In stark contrast, the parental RCV-A1 (GI.4) is a benign, enterotropic virus that has only been recovered from rabbits (Strive, Wright, and Robinson 2009), while the RHDVa-Aus (GI.4eP-GI.1a) variant, although virulent and hepatotropic, has only been found in adult rabbits (Fig. 2). Our findings suggest that it is the lagovirus structural sequences (VP60 and VP10, plus 3′-untranslated regions [UTRs]) that confer both host and tissue tropism, and that tissue tropism is correlated with virulence.

With the broader host tropism conferred by the genotype GI.2 structural proteins, there is increased potential for the emergence of novel epizootic lagovirus variants through both intragenotypic and intergenotypic recombination. Hares are known to carry their own, presumed benign, caliciviruses (Droillard et al., 2018, 2020; Mahar et al., 2019; Cavadini et al., 2020). Indeed, the first intergenotypic lagovirus recombinants were recently reported from Germany (Szillat et al., 2020). Since the incursion of RHDV2 (GI.3P-GI.2) into North America in 2020, this variant has also been reported to infect several *Sylvilagus* species (cottontail rabbits) (Lankton et al., 2021). While endemic *Sylvilagus* calciviruses have never been reported, very limited sampling has been conducted in this species. It remains to be seen whether North American leporids may be a new reservoir for the emergence of novel lagoviruses with the panzootic potential.

This highlights a critical need for sustained surveillance efforts and full genetic characterisation of lagoviruses to facilitate the early detection of novel variants of interest. Indeed, this approach is of importance for many emerging and endemic viruses of significance to both animal and human health. Our finding that the NS genes strongly influence epidemiological fitness in the lagoviruses and that recombination happens so frequently, even

in acute viral infections, concurs with similar findings on human noroviruses and emphasizes the importance of considering the full viral genome when undertaking surveillance for variants of concern. Caliciviruses, with a single major recombination breakpoint, provide an excellent model system to demonstrate this; however, this is likely true for many virus families. While conventionally the focus has been on detecting changes in the structural gene sequences, it is becoming clear that a deeper exploration of the genes encoding the NS proteins is necessary to better understand the complex drivers of viral epidemiological fitness.

## Data availability

## Supplementary data

Supplementary data is available at *Virus Evolution* online.

## Acknowledgements

## Funding

## References

Abrantes, J. et al. (2020) 'Recombination at the Emergence of the Pathogenic Rabbit Haemorrhagic Disease Virus *Lagovirus europaeus*/GI.2', *Scientific Reports*, 10: 14502.

———— et al. (2012) 'Rabbit Haemorrhagic Disease (RHD) and Rabbit Haemorrhagic Disease Virus (RHDV): A Review', *Veterinary Research*, 43: 12.

Ao, Y. et al. (2018) 'Genetic Analysis of Reemerging GII.P16-GII.2 Noroviruses in 2016-2017 in China', *Journal of Infectious Diseases*, 218: 133–43.

Arias, A. et al. (2016) 'Norovirus Polymerase Fidelity Contributes to Viral Transmission *in Vivo*', *mSphere*, 1: e00279–16.

Buehler, M. et al. (2020) '*Lagovirus Europeus* GI.2 (Rabbit Hemorrhagic Disease Virus 2) Infection in Captive Mountain Hares (*Lepus Timidus*) in Germany', *Veterinary Research*, 16: 166.

Bull, R. A. et al. (2005) 'Norovirus Recombination in ORF1/ORF2 Overlap', *Emerging Infectious Diseases*, 11: 1079–85.

Camarda, A. et al. (2014) 'Detection of the New Emerging Rabbit Haemorrhagic Disease Type 2 Virus (RHDV2) in Sicily from Rabbit (*Oryctolagus Cuniculus*) and Italian Hare (*Lepus Corsicanus*)', *Research in Veterinary Science*, 97: 642–5.

Capucci, L. et al. (1998) 'A Further Step in the Evolution of Rabbit Hemorrhagic Disease Virus: The Appearance of the First Consistent Antigenic Variant', *Virus Research*, 58: 115–26.

—— et al. (1996) 'Detection and Preliminary Characterization of a New Rabbit Calicivirus Related to Rabbit Hemorrhagic Disease Virus but Nonpathogenic', *Journal of Virology*, 70: 8614–23.

Capucci, L., Nardin, A., and Lavazza, A. (1997) 'Seroconversion in an Industrial Unit of Rabbits Infected with a Non-pathogenic Rabbit Haemorrhagic Disease-like Virus', *Veterinary Record*, 140: 647–50.

Cavadini, P. et al. (2020) 'Widespread Occurrence of the Non-pathogenic Hare Calicivirus (Hacv Lagovirus GII.2) In Captive-reared and Free-living Wild Hares in Europe', *Transboundary and Emerging Diseases*, 68: 509–18.

Cheung, S. K. C. et al. (2019) 'Higher Viral Load of Emerging Norovirus GII.P16-GII.2 Than Pandemic GII.4 And Epidemic GII.17, Hong Kong, China', *Emerging Infectious Diseases*, 25: 119–22.

Conley, M. J. et al. (2019) 'Calicivirus VP2 Forms a Portal-like Assembly following Receptor Engagement', *Nature*, 565: 377–81.

Dalton, K. P. et al. (2018) 'Conventional and Real Time RT-PCR Assays for the Detection and Differentiation of Variant Rabbit Hemorrhagic Disease Virus (Rhdvb) and Its Recombinants', *Journal of Virological Methods*, 251: 118–22.

—— et al. (2014) 'Spread of New Variant RHDV in Domestic Rabbits on the Iberian Peninsula', *Veterinary Microbiology*, 169: 67–73.

—— et al. (2012) 'Variant Rabbit Hemorrhagic Disease Virus in Young Rabbits, Spain', *Emerging Infectious Diseases*, 18: 2009–12.

Droillard, C. et al. (2018) 'First Complete Genome Sequence of a Hare Calicivirus Strain Isolated from *Lepus Europaeus*', *Microbiology Resource Announcements*, 7: e01224–18.

—— et al. (2020) 'Genetic Diversity and Evolution of Hare Calicivirus (Hacv), a Recently Identified Lagovirus from *Lepus Europaeus*', *Infection, Genetics and Evolution*, 82: 104310.

Drummond, A. J., and Rambaut, A. (2007) 'BEAST: Bayesian Evolutionary Analysis by Sampling Trees', *Evolutionary Biology*, 7: 214.

Eden, J. S. et al. (2015) 'Comparative Phylodynamics of Rabbit Hemorrhagic Disease Virus in Australia and New Zealand', *Journal of Virology*, 89: 9548–58.

—— et al. (2013) 'Recombination within the Pandemic Norovirus GII.4 Lineage', *Journal of Virology*, 87: 6270–82.

Gavier-Widen, D., and Morner, T. (1991) 'Epidemiology and Diagnosis of the European Brown Hare Syndrome in Scandinavian Countries: A Review', *Revue Scientifique Et Technique*, 10: 453–8.

Grolemund, G., and Wickham, H. (2011) 'Dates and Times Made Easy with Lubridate', *Journal of Statistical Software*, 40: 1–25.

Guindon, S. et al. (2010) 'New Algorithms and Methods to Estimate Maximum-likelihood Phylogenies: Assessing the Performance of PhyML 3.0', *Systematic Biology*, 59: 307–21.

Hall, R. N. et al. (2015) 'Emerging Rabbit Hemorrhagic Disease Virus 2 (Rhdvb), Australia', *Emerging Infectious Diseases*, 21: 2276–8.

—— et al. (2018) 'A Strain-specific Multiplex RT-PCR for Australian Rabbit Haemorrhagic Disease Viruses Uncovers A New Recombinant Virus Variant in Rabbits and Hares', *Transboundary and Emerging Diseases*, 65: e444–e56.

—— et al. (2017) 'Detection of RHDV2 in European Brown Hares (*Lepus Europaeus*) in Australia', *Veterinary Research*, 180: 121.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Kerr, P. J., Hall, R. N., and Strive, T. (2021) 'Viruses for Landscape-scale Therapy: Biological Control of Rabbits in Australia', *Methods in Molecular Biology*, 2225: 1–23.

Kimura, T. et al. (2001) 'Distribution of Rabbit Haemorrhagic Disease Virus RNA in Experimentally Infected Rabbits', *Journal of Comparative Pathology*, 124: 134–41.

Kirkland, P. D., and Frost, M. J. (2020) 'The Impact of Viral Transport Media on PCR Assay Results for the Detection of Nucleic Acid from SARS-CoV-2', *Pathology*, 52: 811–4.

Lankton, J. S. et al. (2021) 'Pathology of *Lagovirus Europaeus* GI.2/RHDV2/b (Rabbit Hemorrhagic Disease Virus 2) in Native North American Lagomorphs', *Journal of Wildlife Diseases*, 57: 694–700.

Le Gall-Recule, G. et al. (2013) 'Emergence of a New Lagovirus Related to Rabbit Haemorrhagic Disease Virus', *Veterinary Research*, 44: 81.

—— et al. (2017) 'Large-scale Lagovirus Disease Outbreaks in European Brown Hares (*Lepus Europaeus*) in France Caused by RHDV2 Strains Spatially Shared with Rabbits (*Oryctolagus Cuniculus*)', *Veterinary Research*, 48: 70.

—— et al. (2011a) 'Detection of a New Variant of Rabbit Haemorrhagic Disease Virus in France', *Veterinary Research*, 168: 137–8.

—— et al. (2011b) 'Characterisation of a Non-pathogenic and Non-protective Infectious Rabbit Lagovirus Related to RHDV', *Virology*, 410: 395–402.

Le Pendu, J. et al. (2017) 'Proposal for a Unified Classification System and Nomenclature of Lagoviruses', *Journal of General Virology*, 98: 1658–66.

Lemaitre, E. et al. (2018) 'First Complete Genome Sequence of a European Non-pathogenic Rabbit Calicivirus (Lagovirus GI.3)', *Archives of Virology*, 163: 2921–4.

Liu, J. et al. (2014) 'Distribution and Prevalence of the Australian Non-pathogenic Rabbit Calicivirus Is Correlated with Rainfall and Temperature', *PLoS One*, 9: e113976.

Liu, S. J. et al. (1984) 'A New Viral Disease in Rabbits', *Journal of Veterinary Diagnostic Investigation*, 16: 253–5.

Lopes, A. M. et al. (2015) 'Full Genomic Analysis of New Variant Rabbit Hemorrhagic Disease Virus Revealed Multiple Recombination Events', *Journal of General Virology*, 96: 1309–19.

—— et al. (2019) 'GI.1b/GI.1b/GI.2 Recombinant Rabbit Hemorrhagic Disease Virus 2 (*Lagovirus europaeus*/GI.2) In Morocco, Africa', *Archives of Virology*, 164: 279–83.

Ludwig-Begall, L. F., Mauroy, A., and Thiry, E. (2018) 'Norovirus Recombinants: Recurrent in the Field, Recalcitrant in the Lab—a Scoping Review of Recombination and Recombinant Types of Noroviruses', *Journal of General Virology*, 99: 970–88.

Mahar, J. E. et al. (2018a) 'Rabbit Hemorrhagic Disease Virus 2 (RHDV; GI.2) Is Replacing Endemic Strains of RHDV in the Australian Landscape within 18 Months of Its Arrival', *Journal of Virology*, 92: e01374–17.

—— et al. (2019) 'The Discovery of Three New Hare Lagoviruses Reveals Unexplored Viral Diversity in This Genus', *Virus Evolution*, 5: vez005.

—— et al. (2016) 'Benign Rabbit Caliciviruses Exhibit Evolutionary Dynamics Similar to Those of Their Virulent Relatives', *Journal of Virology*, 90: 9317–29.

—— et al. (2018b) 'Detection and Circulation of a Novel Rabbit Hemorrhagic Disease Virus in Australia', *Emerging Infectious Diseases*, 24: 22–31.

Martin, D. P. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.

Meyers, G., Wirblich, C., and Thiel, H. J. (1991) 'Genomic and Subgenomic RNAs of Rabbit Hemorrhagic Disease Virus are Both Protein-linked and Packaged into Particles', *Virology*, 184: 677–86.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Neave, M. J. et al. (2018) 'Robust Innate Immunity of Young Rabbits Mediates Resistance to Rabbit Hemorrhagic Disease Caused by *Lagovirus Europaeus* GI.1 But Not GI.2', *Viruses*, 10: 512.

Neimanis, A. S. et al. (2018) 'Overcoming Species Barriers: An Outbreak of *Lagovirus Europaeus* GI.2/RHDV2 in an Isolated Population of Mountain Hares (*Lepus Timidus*)', *Veterinary Research*, 14: 367.

Nicholson, L. J. et al. (2017) 'Benign Rabbit Calicivirus in New Zealand', *Applied and Environmental Microbiology*, 83: e00090–17.

Ohlinger, V. F. et al. (1990) 'Identification and Characterization of the Virus Causing Rabbit Hemorrhagic Disease', *Journal of Virology*, 64: 3331–6.

Paradis, E., and Schliep, K. (2019) 'Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.

Parra, G. I. (2019) 'Emergence of Norovirus Strains: A Tale of Two Genes', *Virus Evolution*, 5: vez048.

—— et al. (2017) 'Static and Evolving Norovirus Genotypes: Implications for Epidemiology and Immunity', *PLoS Pathogens*, 13: e1006136.

Puggioni, G. et al. (2013) 'The New French 2010 Rabbit Hemorrhagic Disease Virus Causes an RHD-like Disease in the Sardinian Cape Hare (*Lepus Capensis Mediterraneus*)', *Veterinary Research*, 44: 96.

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rambaut, A. et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Ramsey, D. S. L. et al. (2020) 'Emerging RHDV2 Suppresses the Impact of Endemic and Novel Strains of RHDV on Wild Rabbit Populations', *Journal of Applied Ecology*, 57: 630–41.

Revell, L. J. (2012) 'Phytools: An R Package for Phylogenetic Comparative Biology (And Other Things)', *Methods in Ecology and Evolution*, 3: 217–23.

Ruis, C. et al. (2017) 'The Emerging GII.P16-GII.4 Sydney 2012 Norovirus Lineage Is Circulating Worldwide, Arose by Late-2014 and Contains Polymerase Changes that May Increase Virus Transmission', *PLoS One*, 12: e0179572.

Schliep, K. P. (2011) 'Phangorn: Phylogenetic Analysis in R', *Bioinformatics*, 27: 592–3.

Silverio, D. et al. (2018) 'Insights into the Evolution of the New Variant Rabbit Haemorrhagic Disease Virus (GI.2) And the Identification of Novel Recombinant Strains', *Transboundary and Emerging Diseases*, 65: 983–92.

Sosnovtsev, S. V. et al. (2005) 'Feline Calicivirus VP2 Is Essential for the Production of Infectious Virions', *Journal of Virology*, 79: 4012–24.

Strive, T., Wright, J. D., and Robinson, A. J. (2009) 'Identification and Partial Characterisation of a New Lagovirus in Australian Wild Rabbits', *Virology*, 384: 97–105.

Sumner, M. (2020) ozmaps: Australia maps.

Szillat, K. P. et al. (2020) 'Full-genome Sequencing of German Rabbit Haemorrhagic Disease Virus Uncovers Recombination between RHDV (GI.2) And EBHSV (GII.1)', *Virus Evolution*, 6: veaa080.

Tohma, K. et al. (2017) 'Phylogenetic Analyses Suggest that Factors Other than the Capsid Protein Play a Role in the Epidemic Potential of GII.2 Norovirus', *mSphere*, 2: e00187–17.

—— et al. (2021) 'Genome-wide Analyses of Human Noroviruses Provide Insights on Evolutionary Dynamics and Evidence of Coexisting Viral Populations Evolving under Recombination Constraints', *PLoS Pathogens*, 17: e1009744.

Urakova, N. et al. (2016) 'Purification and Biochemical Characterisation of Rabbit Calicivirus RNA-dependent RNA Polymerases and Identification of Non-nucleoside Inhibitors', *Viruses*, 8: 100.

Velarde, R. et al. (2017) 'Spillover Events of Infection of Brown Hares (*Lepus Europaeus*) with Rabbit Haemorrhagic Disease Type 2 Virus (RHDV2) Caused Sporadic Cases of an European Brown Hare Syndrome-like Disease in Italy and Spain', *Transboundary and Emerging Diseases*, 64: 1750–61.

Wang, L. G. et al. (2020) 'Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data', *Molecular Biology and Evolution*, 37: 599–603.

Wang, X. et al. (2013) 'Atomic Model of Rabbit Hemorrhagic Disease Virus by Cryo-electron Microscopy and Crystallography', *PLoS Pathogens*, 9: e1003132.

Wickham, H. (2011) 'The Split-apply-combine Strategy for Data Analysis', *Journal of Statistical Software*, 40: 1–29.

—— et al. (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4: 1686.

Wickham, H., and Seidel, D. (2020) 'Scales: Scale Functions for Visualization'. R package version 1.1.1.

Wilke, C. (2020) Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'.

Wobus, C. E. et al. (2004) 'Replication of Norovirus in Cell Culture Reveals a Tropism for Dendritic Cells and Macrophages', *PLoS Biolology*, 2: e432.

Worobey, M., and Holmes, E. C. (1999) 'Evolutionary Aspects of Recombination in RNA Viruses', *Journal of General Virology*, 80: 2535–43.

Yang, Z. (2007) 'PAML 4: Phylogenetic Analysis by Maximum Likelihood', *Molecular Biology and Evolution*, 24: 1586–91.

Yu, G. et al. (2017) 'Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data', *Methods in Ecology and Evolution*, 8: 28–36.

Zharkikh, A. (1994) 'Estimation of Evolutionary Distances between Nucleotide Sequences', *Journal of Molecular Evolution*, 39: 315–29.