



OPEN

Recombination at the emergence of the pathogenic rabbit haemorrhagic disease virus *Lagovirus europaeus*/GI.2

Joana Abrantes^{1,2✉}, Clément Droillard³, Ana M. Lopes^{1,4}, Evelyne Lemaitre³, Pierrick Lucas⁵, Yannick Blanchard⁵, Stéphane Marchandeu⁶, Pedro J. Esteves^{1,2} & Ghislaine Le Gall-Reculé³

Rabbit haemorrhagic disease is a viral disease that emerged in the 1980s and causes high mortality and morbidity in the European rabbit (*Oryctolagus cuniculus*). In 2010, a new genotype of the rabbit haemorrhagic disease virus emerged and replaced the former circulating *Lagovirus europaeus*/GI.1 strains. Several recombination events have been reported for the new genotype *Lagovirus europaeus*/GI.2, with pathogenic (variants GI.1a and GI.1b) and benign (genotype GI.4) strains that served as donors for the non-structural part while GI.2 composed the structural part; another recombination event has also been described at the p16/p23 junction involving GI.4 strains. In this study, we analysed new complete coding sequences of four benign GI.3 strains and four GI.2 strains. Phylogenetic and recombination detection analyses revealed that the first GI.2 strains, considered as non-recombinant, resulted from a recombination event between GI.3 and GI.2, with GI.3 as the major donor for the non-structural part and GI.2 for the structural part. Our results indicate that recombination contributed to the emergence, persistence and dissemination of GI.2 as a pathogenic form and that all described GI.2 strains so far are the product of recombination. This highlights the need to study full-genomic sequences of lagoviruses to understand their emergence and evolution.

Since the 1980s, European rabbits worldwide, either domestic or wild, have been affected by rabbit haemorrhagic disease¹. This highly contagious and fatal disease is caused by the rabbit haemorrhagic disease virus (RHDV), a single-stranded positive-sense RNA virus that belongs to the family *Caliciviridae*, genus *Lagovirus*. Benign rabbit caliciviruses that confer more or less protection against pathogenic strains, as well as moderately pathogenic strains, including Michigan rabbit calicivirus (MRCV), have also been described²⁻⁷.

In 2010, a new pathogenic lagovirus was identified in France⁸, formerly designated as RHDV2 or RHDVb and now as *Lagovirus europaeus*/GI.2 according to a proposal for a unified nomenclature for lagoviruses⁹. GI.2 caused unusual mortalities in rabbits vaccinated against GI.1 (former G1-G6) strains^{8,10}. Further analysis revealed genetic differences that were reflected in its positioning in a phylogenetic tree based on capsid protein (VP60) sequences. GI.2 constituted a new phylogenetic group^{8,11} with more than 15% divergence from all known benign and pathogenic lagoviruses. Other unique characteristics include its ability to fatally infect rabbits younger than two months (previously resistant to the disease)^{11,12} and several hare species (*Lepus* spp.)¹³⁻¹⁷. Differences in disease duration, mortality rates, and in the frequency of occurrence of subacute/chronic forms have also been

¹CIBIO/InBio-UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Research Network in Biodiversity and Evolutionary Biology, Universidade do Porto, Campus de Vairão, Rua Padre Armando Quintas, 4485-661 Vairão, Portugal. ²Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal. ³Unité de Virologie, Immunologie, Parasitologie, Aviaires et Cunicoles, Laboratoire de Ploufragan-Plouzané-Niort, Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail (Anses), Ploufragan, France. ⁴Instituto de Ciências Biomédicas Abel Salazar (ICBAS)/Unidade Multidisciplinar de Investigação Biomédica (UMIB), Universidade do Porto, Porto, Portugal. ⁵Unité de Génétique Virale et Biosécurité, Laboratoire de Ploufragan-Plouzané-Niort, Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail (Anses), Ploufragan, France. ⁶Unité Petite Faune Sédentaire et Espèces Outre-Mer, Direction de la Recherche et de l'Appui Scientifique, Office Français de la Biodiversité (OFB), Nantes, France. ✉email: jabras@ cibio.up.pt

documented¹¹; moreover, since 2016, virulence of some GI.2 strains has increased to reach that of GI.1 viruses¹⁸. As for the other lagoviruses, the origin of GI.2 is still unknown.

Following its detection in France, GI.2 rapidly migrated through Europe, but also reached more distant places such as Oceania, Africa, and North America¹⁹, revealing an efficient dispersal and establishment. Replacement of the former circulating GI.1 strains was also reported^{11, 20–23}. In the Iberian Peninsula, emergence of GI.2 was accompanied by several recombination events^{24, 25}. These events involved the pathogenic GI.1b variant that circulated only in Iberian rabbit populations, and benign GI.4 strains (RCV-A1). In Australia, a recombinant GI.1a/GI.2 strain was recently characterised in rabbit and hare samples²⁶. A recurrent recombination breakpoint was found at the boundary of the non-structural and structural encoding genes (between RdRp and VP60). A second recombination breakpoint was further detected at the junction of the two genes encoding the non-structural proteins p16 and p23²⁵. While the biological implications of the different genomic composition of the recombinant strains remains to be assessed, it is clear that recombination plays a central role in the evolution of GI.2. Recombination might have also precipitated the emergence of GI.1 as a pathogenic form²⁷, although this still remains under debate. Indeed, two hypotheses had been suggested for the emergence of pathogenic lagoviruses: either from a pre-existing benign lagovirus or following a species jump^{28, 29}. Both hypotheses are sustained by multiple pieces of evidence and are not mutually exclusive, but they require further confirmation, especially following the recent emergence of the new pathogenic genotype GI.2.

In this study, and in order to disclose the emergence of pathogenic rabbit haemorrhagic disease virus *Lagovirus europaeus*/GI.2, we sequenced the full-length genomic sequences of two GI.2 strains collected in the earliest outbreaks¹¹ and the coding sequences of two GI.2 strains collected more recently. We obtained also four new full coding sequences of benign GI.3 (RCV-E1) strains collected in France between late 2007 and early 2009, since GI.3 appears to have a closer relationship with GI.2³⁰. Maximum-likelihood (ML) analyses were performed considering the recombination breakpoint identified by recombination detection methods^{24, 25}.

Results

In this study, we obtained eight new genome sequences: four for GI.3 and four for GI.2. These included four strains previously identified based on their capsid sequences as GI.3 (JA10/08-10, BO25/08-133, JA34/09-48 and CHA20/09-100; GenBank accession numbers: LT708119, LT708122, LT708127 and LT708128, respectively) and two strains as GI.2 (10-28 and 10-32; GenBank accession numbers: HE800531 and HE800532, respectively)^{9, 11}. We obtained the full-length genomic sequences of strains 10-28 and 10-32, the complete coding sequences (CDS) of strains 16-35 and 16-36, and the CDS along with the 3' untranslated region (UTR) of the strains JA10/08-10, BO25/08-133, JA34/09-48 and CHA20/09-100.

The genomes of the earliest GI.2 strains 10-28 and 10-32 were 7,448 nucleotides (nt) long and shared 94.6% nucleotide identity. The length of the 5' UTR sequences is 9 nt and that of the 3' UTR is 70 nt. The closest strains were GI.3 strains with an average nucleotide identity of 93.8% (94.3% for the non-structural part of the genomes), while the average identity with GI.1 strains was 85% (86.4% for the non-structural part of the genomes). The genomes of the GI.3 strains were 7,427 nt long (excluding the 5' UTR), except the CHA20/09-100 strain that was 7,424 nt long. The 3' UTR sequences had the same length (64 nt). When including the GI.3 06-11 sequence (GenBank accession number MN737115), the average nucleotide identity between the five GI.3 strains was 97.1%. The CDS of the four GI.2 showed between 93.2% and 94.5% of nucleotide identity (93.2% between 16-35 and 16-36 strains), and when compared to GI.1 CDS presented no deletions or insertions (7,369 nt long), while the GI.3 CDS presented a deletion of six consecutive nucleotides within the capsid gene (7,363 nt long instead of 7,369 nt). This deletion is present in the benign strains 06-11, Ashington and the Italian RCV (GenBank accession numbers: EF558587 and X96868)^{2, 4, 7} and comprises a codon putatively under positive selection in GI.1 strains³¹. The strain CHA20/09-100 had a further three nucleotide deletion in the minor structural protein VP10 (positions 193–195, amino acid 65).

The complete dataset (221 sequences; 7,368 nucleotides) was screened for recombination with RDP (Recombination Detection Program)³². The different methods available in RDP detected as recombinants, with strong statistical support (p-values < 0.05; Table 1), all the strains from the early GI.2 outbreaks (strains 10-28, 10-32 and N11), all the GI.2 strains previously identified as non-recombinants (strains CBVal16, Zar11-11, Rij06-12, Tar06-12, Zar06-12 and Seg08-12) and GI.2 strains more recently detected (16PLM1, NL2016, Canada2016, 16-35, 16-36). A single recombination breakpoint was determined at the non-structural/structural parts boundary (Table 1). For most of these strains, the GI.3 strain 06-11 was identified as the most likely donor for the non-structural part, with the exception of strain 16-36 for which the most likely donor is CHA20/09-100. A GI.2 strain (7-13_Barrancos_Portugal_2013 or SOS155_Portugal_2015, GenBank accession numbers KF442963 and MG763946, respectively) was the most likely donor for the structural part. Thus, according to RDP all these sequences (GI.2 strains from the early outbreaks and GI.2 non-recombinant strains) are GI.3/GI.2 recombinants. The tanglegram based on the alignments for the non-structural and the structural genome partitions also evidenced as recombinants the strains identified by RDP (Fig. 1).

The phylogenetic analysis was performed taking into account the location of the recombination breakpoint detected by RDP at the junction of the non-structural and structural parts. In the ML tree of the structural part, which includes VP60 and VP10 (Fig. 2i), the GI.3 strains formed a strongly supported group (bootstrap value 100), that was a sister group to GI.1 (bootstrap value 100). The MRCV strain, a GI.4-like/GI.3-like recombinant³³, also grouped with the GI.3 strains (bootstrap value 77). According to the nomenclature recently proposed for lagoviruses, MRCV does not meet the criteria to be considered as a member of the GI.3 group and is currently unclassified⁹. Indeed, the genetic distance between MRCV and GI.3 VP60 gene sequences is > 15% (data not shown) and no other similar strains (from independent outbreaks) had been identified. As for the four GI.2 sequences obtained in this study, they clustered within a highly supported group (bootstrap value 100) formed

Strains	Recombination breakpoint (nucleotide positions) ^a	Most likely donor strain		Methods and average p-value						
		Non-structural proteins	Structural proteins	RDP	GENECONV	BootScan	MaxChi	Chimaera	SiScan	3Seq
10-28 10-32 N11 CBVal16 Zar11-11 Tar06-12 Zar06-12 Rij06-12 Seg08-12 16PLM1 NL2016 Canada2016 16-35	5,242–5,319	06-11 (GI.3; accession number MN737115)	7-13 Barrancos (GI.2; accession number KF442963)	1.061×10^{-126}	8.713×10^{-70}	6.006×10^{-125}	2.539×10^{-27}	9.056×10^{-36}	1.075×10^{-44}	1.604×10^{-10}
16-36	5,231–5,423	CHA20/09-100 (GI.3; accession number LT708128)	SOS155 (GI.2; accession number MG763946)	2.991×10^{-108}	2.061×10^{-91}	–	1.196×10^{-32}	3.575×10^{-32}	3.432×10^{-39}	3.401×10^{-10}

Table 1. Results of the RDP recombination analysis. ^a99% confidence interval.

by all the strains previously identified as having a GI.2 capsid, and that is more closely related to GI.1 and GI.3 than to GI.4. Two strains collected in the Netherlands³⁴ and Canada in 2016 and that were recently made available (GenBank accession numbers MN061492 and KY235675, respectively) were also part of the GI.2 cluster.

Regarding the ML tree of the non-structural part except p16 (Fig. 2ii), the four GI.2 strains obtained in this study 10-28, 10-32, 16-35 and 16-36 clustered with the GI.3 strains. This highly supported cluster (bootstrap value 100) further includes strains SOS148, SOS149, SOS151, SOS173, SOS404 and SOS468 previously identified as GI.4/GI.2/GI.2 (GI.4 for p16, and GI.2 for the remaining genome regions)²⁵, GI.2 strains previously considered as non-recombinants (marked with an *), and the 2016 strains from Canada (Canada2016) and the Netherlands (NL2016).

The ML tree for the non-structural protein p16 (see supplementary information), recapitulated most of the results observed in the ML tree for the non-structural part with the exception of the positioning of the Portuguese strains SOS137, SOS164, SOS148, SOS149, SOS151, SOS173, SOS404 and SOS468, that appear clustered with GI.4. Strains SOS137 and SOS164 were previously shown to be triple recombinants between GI.4, GI.1b and GI.2 (GI.4 for p16, GI.1b for the remaining non-structural proteins and GI.2 for the structural proteins)²⁵ which is in agreement with the results obtained in this study. As for strains SOS148, SOS149, SOS151, SOS173, SOS404 and SOS468, the tree suggests that their genome also originated from three different strains: GI.4 for p16, GI.3 for the remaining non-structural proteins and GI.2 for the structural proteins. In this case, it most likely involved recombination between a GI.4 strain and a GI.3/GI.2 recombinant strain. A tanglegram analysis and RDP further confirmed these results (see supplementary information).

Discussion

GI.2 was identified as a novel pathogenic form of lagovirus in France in 2010⁸. Field observations and further characterisation of the strains revealed unique characteristics in comparison with former strains such as the ability to cross the species boundaries^{13–17} and to kill young rabbits^{11,12}. Previous studies also revealed the occurrence of recombination of GI.2 strains with non-pathogenic (GI.4) and pathogenic (GI.1a and GI.1b) strains^{24–26}, showing an important role of recombination in generating diversity in GI.2 and confirming the high capacity of recombination within lagoviruses.

Our results show that the strains circulating at the time of the first noticed GI.2 outbreaks were already recombinants between a European non-pathogenic strain (GI.3), that was donor for the non-structural part, and the new lagovirus genotype, GI.2, that formed the structural part. Indeed, in our ML tree for the non-structural part, the strains from the first GI.2 outbreaks (10-28 and 10-32) grouped with the non-pathogenic GI.3 group, while for the structural part these strains clustered with all the strains of the newly emerged GI.2 genotype. Strains from the earliest outbreaks in Spain and Portugal in 2011–2012²⁴, and strains collected in the Canary Islands (Spain)³⁵, France, the Netherlands³⁴ and Canada in 2016 are also GI.3/GI.2 recombinants. These results, that were confirmed by the RDP, the ML trees and the tanglegram analyses, demonstrate that all the GI.2 strains described so far, including those that were considered non-recombinants and that in some instances co-circulated with other recombinant GI.2 strains²⁴, resulted from recombination. Several recombination events are thus associated with the evolution of this new genotype: with GI.3, GI.4, GI.1b and GI.1a, with a recombination breakpoint at the non-structural/structural boundary^{24,26}. An additional recombination event was further previously reported with a breakpoint at the p16/p23 boundary and involved GI.4 as the donor of p16, while GI.3/GI.2 (see also supplementary information) or GI.1b/GI.2 recombinants were the donors for the remaining viral genome²⁵.

Notably, the pattern described here for lagoviruses, with a recombination hotspot at the start of the major structural gene that encodes the capsid, is typical of caliciviruses^{36–38}. Indeed, the combination of low sequence divergence³⁹, presence of complex RNA secondary structures³⁹ that may promote template switching by the virus polymerase, and the existence of a subgenomic RNA⁴⁰ that may act as a secondary template when RNA replication

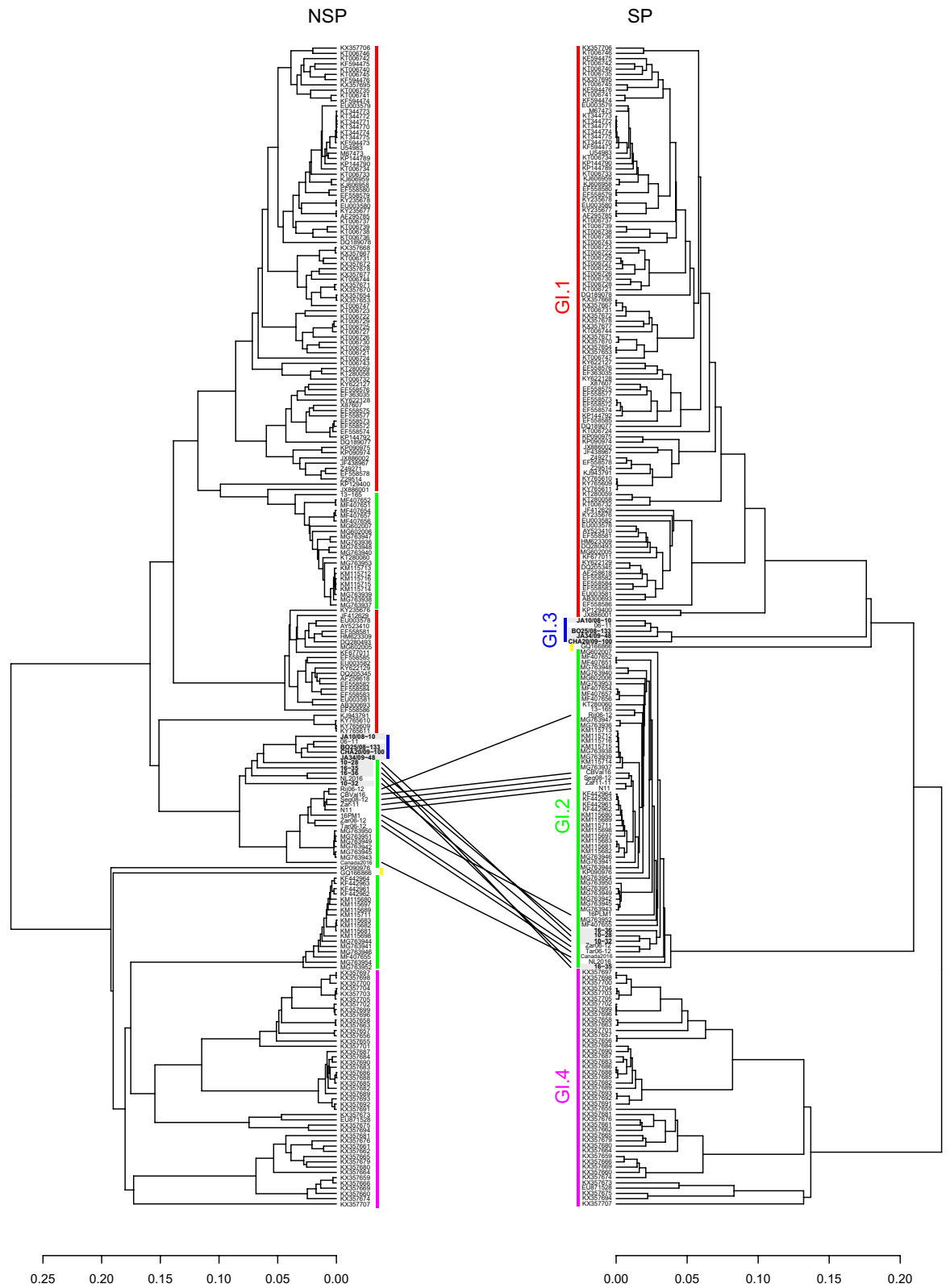


Figure 1. Tanglegram for the non-structural (NSP) and structural genes (SP) of the 221 lagovirus sequences. For the tanglegram for the non-structural genes (NSP), colouring of the sequences is according to the genotype to which the strains belong to for the structural genes (SP): GI.1 (red); GI.2 (green); GI.3 (blue); GI.4 (purple); MRCV (yellow). Lines highlight the recombinant GI.3/GI.2 strains that were identified in the RDP recombination analysis. Sequences obtained in this study appear boxed in grey.

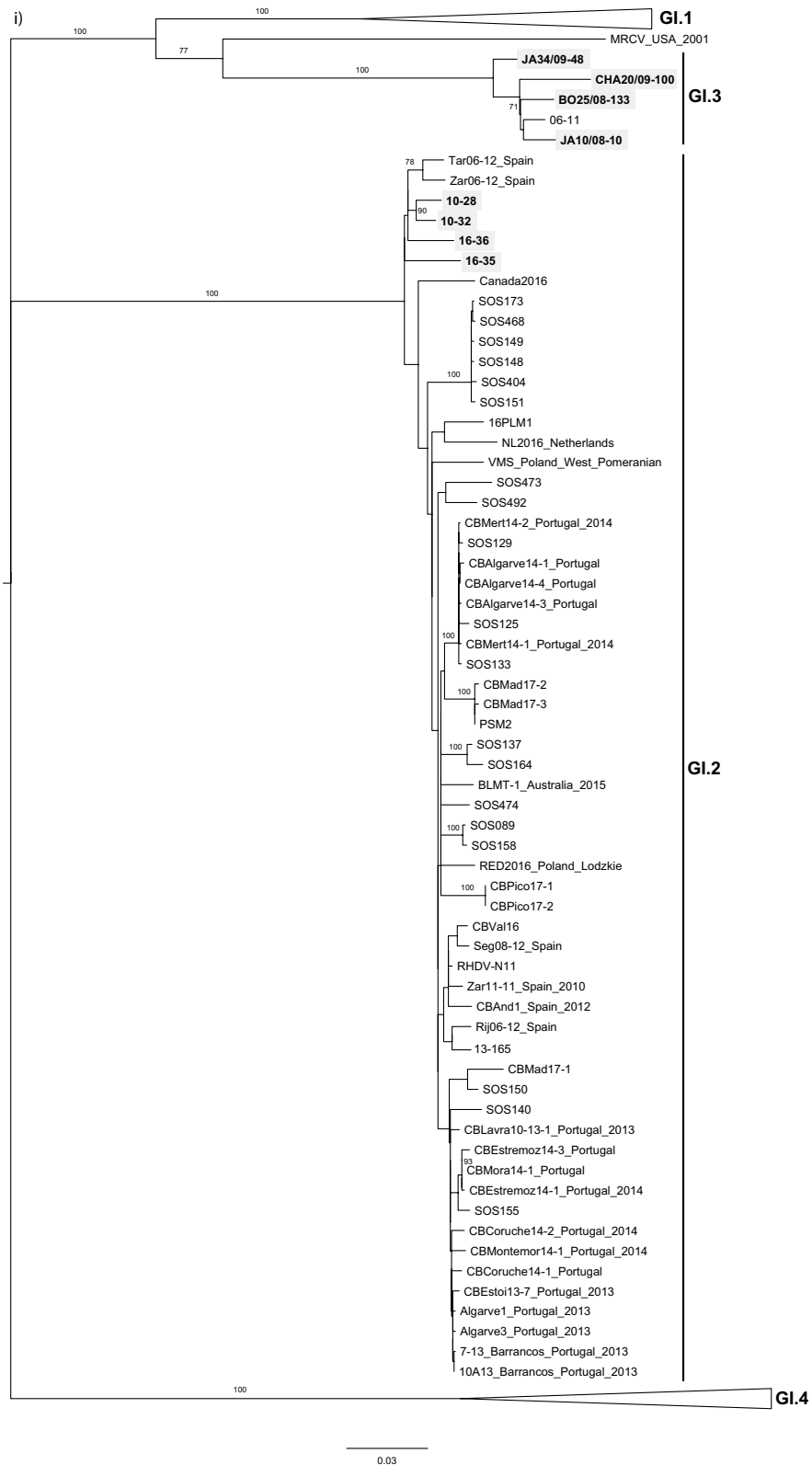


Figure 2. Maximum Likelihood (ML) phylogenetic trees for (i) the structural genes VP60 + VP10 ($n = 221$ sequences; nucleotides 5,296–7,369; nucleotide substitution model $GTR + G + \Gamma_4$), and (ii) the non-structural genes except p16 ($n = 221$ sequences; nucleotides 430–5,295; nucleotide substitution model $GTR + G + \Gamma_4$). Horizontal branch lengths are drawn to scale of nucleotide substitutions per site and the trees are mid-point rooted. The percentage of trees in which the associated taxa clustered together was determined from 1,000 bootstrap replicates and is shown next to the branches (only bootstrap values ≥ 70 are shown). Genotype or variant assignment is according to the ML tree for the structural genes. Sequences obtained in this study appear boxed in grey. * indicates strains considered previously non-recombinant GI.2, but now identified as GI.3/GI.2 recombinants. GenBank accession numbers of the sequences used are listed in supplementary information.

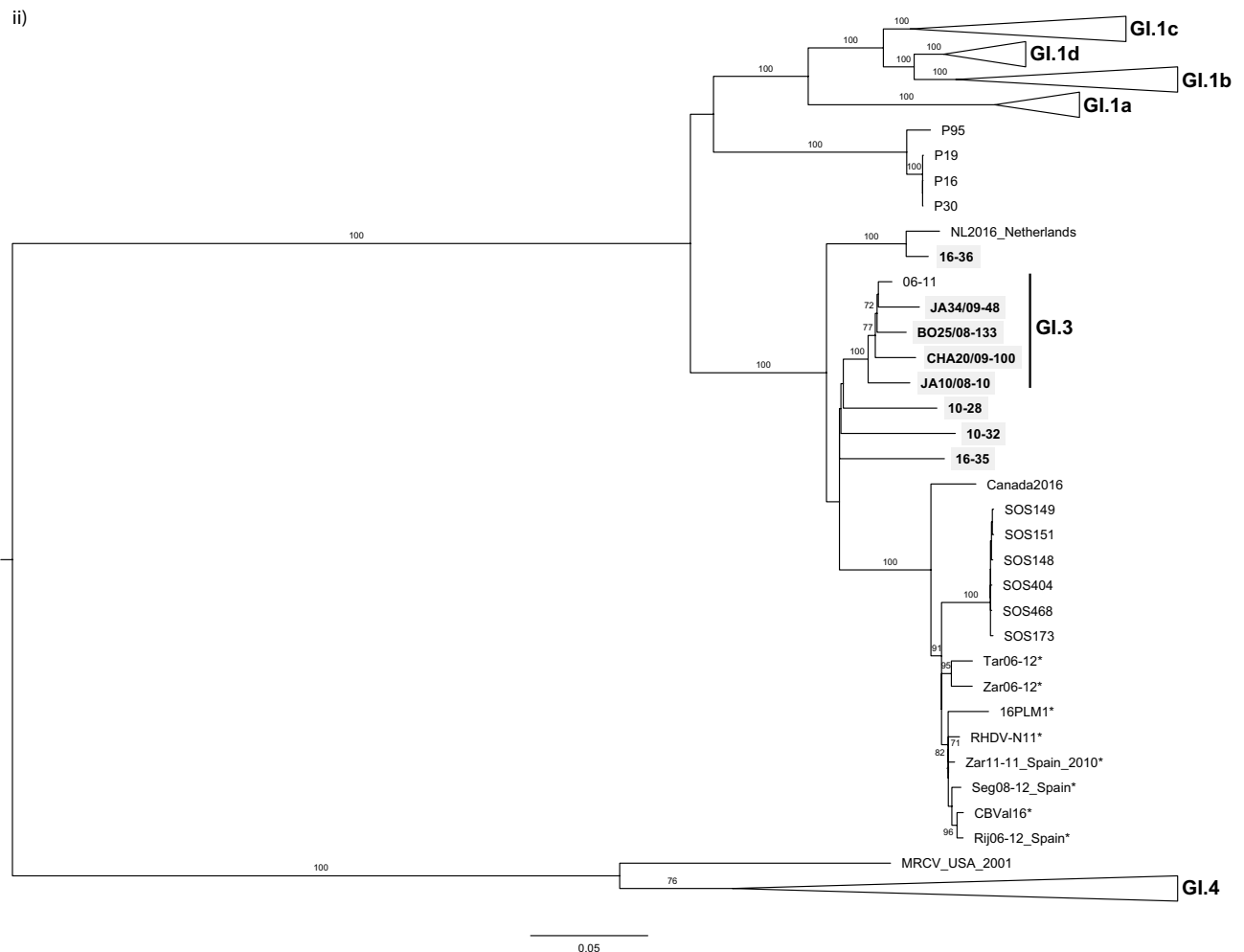


Figure 2. (continued)

is resumed upon template switching, seem to contribute for the high rates of recombination observed in this region in the genome of caliciviruses and that likely constitutes an important survival strategy in the evolution of this family⁴¹. This strategy, that resembles antigenic shift in Influenza virus⁴¹, allows caliciviruses to rapidly generate diversity by producing new genomic combinations, which might be beneficial for the adaptation to new hosts and environments, and to overcome selective pressures. Recombination is important in shaping the evolution of RNA viruses, including the closely related picornaviruses⁴² and coronaviruses⁴³ and, more importantly, is often associated with the emergence of new viruses⁴⁴ and even families, e.g. the *Hepeviridae* family⁴⁵. As for the recombination involving the breakpoint at the p16/p23 boundary, it may have originated from a non-replicative recombination mechanism where RNA strands are randomly self-ligated or joined by cellular ligases⁴⁶. Atypical recombination breakpoints such as this have been also observed for other caliciviruses and a similar mechanism proposed⁴⁷.

Estimation of the time to the most recent common ancestor (tMRCA) of GI.2 by using capsid sequences pointed to an emergence in July 2008²⁵. Although the estimated substitution rates may be inaccurate due to limited sampling, reduced sequence variation and low temporal spread, they tend to underestimate rather than overestimate the real tMRCA^{48,49}. Thus, it is possible that despite surveillance efforts, GI.2 circulated unnoticed in wildlife prior to its detection⁵⁰. The virus that recombined with GI.3 to produce this new genotype is currently unknown and undetected, even with the molecular surveys of lagoviruses performed in European leporids and our improved knowledge of the complete coding sequences of both pathogenic and benign lagoviruses. The same occurs for some older recombinant Iberian strains where the virus that originated the non-structural part has never been detected⁵¹. Both viruses, the one originating GI.2 and the one of these Iberian strains, could have either circulated harmlessly prior to their detection, thus making difficult to detect them, or have become extinct due to their lower fitness as a non-recombinant form, as suggested for noroviruses that recombined and whose partial sequences were never found⁵².

The evidence for recombination with GI.1a²⁶, GI.1b²⁴, GI.3 and GI.4^{24,25} (see also supplementary information) reveals that GI.2 successfully recombines with a great diversity of pathogenic and non-pathogenic lagoviruses. This, together with the tMRCA estimates and the hypothesis of circulating harmlessly before its emergence, supports an evolution from a non-pathogenic form that acquired pathogenicity²⁸. We suggest that evolution of

pathogenicity was not driven solely by point mutations, but was aided by recombination events. Similarly to other RNA viruses, lagoviruses genomes may function as interchangeable modules rather than as strict genomes, which leads to the appearance of mosaic-like genomes through recombination, resulting in a semi-independent evolution of structural and non-structural genes^{53,54} and with low-fitness regions being eliminated by recombination⁵⁵. This might help to explain how a non-pathogenic strain (GI.3), combined with a potentially benign strain (GI.2), gave rise to a highly lethal, pathogenic strain. Non-pathogenic forms are thus likely to be involved in the evolution of pathogenic forms and recombination might have precipitated the emergence of *Lagovirus europaeus*/GI.2. The hypothesis of a species jump may not be discarded. Indeed, reservoir species could have acted as source for the original virus that jumped to the final host, the European rabbit²⁸. Whether recombination occurred in the reservoir species or in the final host is unknown. Nonetheless, reservoir species have yet to be identified in which lagoviruses could have evolved and replicated without being detected before “jumping” into the European rabbit (Le Gall-Reculé, personal communication)^{56–59}. Genetic mechanisms such as recombination require co-infection of a host cell by two (or more) viruses. The high frequency of recombination detected in lagoviruses also implies a high frequency of co-infection, either in the host population or in the reservoir species. Leeks and colleagues established a link between co-infection and viral diversity, as more diverse populations can produce more virulent infections and better adapt to new hosts⁶⁰. This might explain the array of hare species infected by GI.2^{13–17}, which constitutes a novelty in relation to GI.1.

Our results add complexity to the diversity of lagoviruses, but help to elucidate our current understanding on the emergence of pathogenic forms. In addition, they show that full genomic sequences of lagoviruses are crucial to understand their evolutionary history and genetic relationships. Future studies should focus on unravelling the role of the structural and the non-structural parts in the emergence of pathogenicity in lagoviruses.

Methods

Virus samples and full-length genome sequencing. No animals were captured, handled or killed specifically for the purpose of this study. Duodenum samples were collected from four hunted French rabbits by the National Hunting and Wildlife Agency (ONCFS) during the hunting seasons in 2007–2008 and 2008–2009. Presence of lagoviruses had previously been detected in these samples and phylogenetic relationships of the obtained capsid VP60 gene sequences had been determined⁹; these were identified as GI.3 strains. Four liver samples were further collected in France from dead rabbits collected in three rabbitries affected by RHD, two on November 2010 (10-28 and 10-32 GI.2 strains¹¹) and one in June 2016 (strain 16-35), and in the field on June 2016 (strain 16-36). The presence of GI.2 in the 16-35 sample was diagnosed by Labovet Conseil (Les Herbiers, France). We characterised this strain as well as the 16-36 one based on their complete VP60 gene sequences, confirming that they belong to the GI.2 genotype.

For the sequencing of the complete coding genome of GI.3 strains, ~30 mg of duodenal tissue were homogenized in a mixer-mill disruptor (TissueLysed, Qiagen, Hidden, Germany). RNAs were first extracted with TRIzol[®] (TRIzol LS Reagent, Ambion, Thermo Fisher Scientific, Villebon-sur-Yvette, France) before using the RNeasy Mini kit (Qiagen, Hilden, Germany) to increase sensitivity². For GI.2 strains, RNA was extracted from 100 µL of liver exudate using the RNeasy Mini kit. Except for two GI.2 strains (16-35 and 16-36) for which the coding sequence was obtained using NGS (see below) from purified and quantified RNA using the RNeasy MinElute cleanup kit (Qiagen, Hilden, Germany) and the Qubit RNA HS Assay Kit with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Villebon-sur-Yvette, France), respectively, RNAs were reverse-transcribed using oligo-dT as a primer and Maxima Reverse Transcriptase (Thermo Fisher Scientific, Villebon-sur-Yvette, France). Protocols were performed as recommended by the manufacturers.

According to the strains, different PCR strategies were attempted for genome sequencing either by using combinations of newly designed primers (primer sequences available upon request) or primers previously published^{3,11,14}. Thus, cDNAs of JA10/08-10, BO25/08-133 and CHA20/09-100 GI.3 strains were amplified using several overlapping PCRs, including one that covered the recombinant breakpoint between the non-structural and structural encoding genes, using Expand High-fidelity PCR System (Roche, Sigma-Aldrich, Saint-Quentin-Fallavier, France). For JA34/09-48 GI.3 strain, due to lower viral loads, after two first PCRs that amplified two overlapping fragments of about 5,700 and 6,700 bp, respectively, several nested or semi-nested overlapping PCRs were performed. The complete coding sequences of the four GI.2 genomes were obtained following either one PCR amplification of about 7,300 bp using the primers 1U¹⁴ and 15L¹¹ followed by smaller overlapping PCRs for sequence confirmation (10-28 and 10-32; primer sequences available upon request), or NGS (16-35 and 16-36, see below). For these two last strains, the 5' end sequence of the CDS (530 bp) was confirmed using the PCR primers 1U and 1L as described in Le Gall-Reculé et al. (2017)¹⁴. PCRs were performed using Expand High Fidelity enzyme (Roche-Applied-Science). The different PCR products were analysed by electrophoresis, purified using the MinElute PCR Purification kit (Qiagen, Hilden, Germany) and quantified using the Qubit dsDNA HS Assay Kit with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Villebon-sur-Yvette, France). DNA sequences were determined with an ABI Prism 3100 Genetic or a 3500 Series Genetic Analysers in both directions (Applied Biosystems, Foster City, CA, USA), using the PCR and several inner primers and Big Dye Terminator v3.1 (Life Technologies) as recommended by the manufacturer. The 5' and 3' ends of the 10-28 and 10-32 strains were obtained using the rapid amplification of cDNA ends (RACE) method following the protocol developed in Lemaitre et al. (2018)³⁰. PCR products were purified and sequenced as described above. Consensus sequences were compiled using Vector NTI Advance software (Life Technologies, Thermo Fisher Scientific, Villebon-sur-Yvette, France).

For NGS, cDNA libraries were prepared using the Ion Total RNA-Seq Kit (Life Technologies, Carlsbad, CA, USA) according to a protocol adapted from the supplier's instructions (available upon request from the authors). Sequencing was performed using Ion Proton Sequencer (Life Technologies). The reads were cleaned with the

Trimmomatic⁶¹ (version 0.36) software (ILLUMINACLIP: oligos.fasta: 2:30:5:1: true; LEADING: 3; TRAILING: 3; MAXINFO: 40:0.2; MINLEN: 36). Then a Bowtie2⁶² (version 2.2.5) alignment was performed with cleaned reads on local nucleotide databank to identify lagovirus references. All cleaned reads aligned by BWA⁶³ (version 0.7.15-r1140) on "*Oryctolagus cuniculus*" (reference genome from Ensembl) were excluded from further analysis. The references with the highest number of matching reads were used for a new alignment with BWA-MEM. The aligned reads of this third alignment were collected and then down-sampled to fit a global coverage depth estimation of 80x. These cleaned reads were submitted to the SPAdes⁶⁴ (version 3.10.0) de novo assembler and their raw related reads to the Mira⁶⁵ (version 4.0.2) de novo assembler. The *de novo* contigs were then submitted to MegaBLAST⁶⁶ on a local copy of nucleotide databank. The best lagovirus reference identified with MegaBLAST was used in a last BWA alignment of cleaned reads.

The genomic sequences produced in this study have been deposited in GenBank under the accession numbers: MN746288, MN746289, MN737116, MN737117 (GI.3 strains JA10/08-10, BO25/08-133, JA34/09-48 and CHA20/09-100, respectively); MN737113, MN737114, MN738377, MN786321 (GI.2 strains 10-28, 10-32, 16-35 and 16-36, respectively).

Recombination analysis. Complete coding sequences obtained in this study were aligned with publicly available complete coding sequences of *Lagovirus europaeus* representing genotypes GI.1, GI.2, GI.3 and GI.4 in BioEdit software version 7.0.3⁶⁷. The final dataset included 221 sequences, 7,369 nucleotides in length (see supplementary information for the list of the sequences used). The dataset was screened for recombination by seven methods available in the RDP software version 4.4 (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq)³² with the following parameters: sequences were set to linear, Bonferroni correction, highest acceptable p-value of 0.05 and 500 permutations. Only recombination events detected by three or more methods were considered.

Recombinant strains were visualized by plotting a tanglegram using the "dendextend package"⁶⁸ in the RStudio software (version 3.6.1)⁶⁹.

Phylogenetic analysis. Pairwise nucleotide distance comparison (p-distance model) was computed using MEGA6⁷⁰.

Following the results obtained with RDP, phylogenetic analyses were carried out separately for the following genome partitions: nucleotide positions (i) 1–429 (p16; supplementary information), (ii) 430–5,295 (p23, p37, p29, VPg, protease and RdRp), and (iii) 5,296–7,369 (VP60 and VP10). Maximum-likelihood phylogenetic trees were inferred in MEGA6⁷⁰ using the best model of nucleotide substitution determined in the same software for the different genome partitions, according to the lowest AICc value (Akaike information criterion, corrected). Support for each cluster was obtained from 1,000 bootstrap replicates.

Received: 9 January 2020; Accepted: 10 August 2020

Published online: 02 September 2020

References

- Abrantes, J., van der Loo, W., Le Pendu, J. & Esteves, P. J. Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review. *Vet. Res.* **43**, 12. <https://doi.org/10.1186/1297-9716-43-12> (2012).
- Le Gall-Reculé, G. *et al.* Characterisation of a non-pathogenic and non-protective infectious rabbit lagovirus related to RHDV. *Virology* **410**, 395–402. <https://doi.org/10.1016/j.virol.2010.12.001> (2011).
- Strive, T., Wright, J. D. & Robinson, A. J. Identification and partial characterisation of a new Lagovirus in Australian wild rabbits. *Virology* **384**, 97–105. <https://doi.org/10.1016/j.virol.2008.11.004> (2009).
- Capucci, L., Fusi, P., Lavazza, A., Paciarini, M. L. & Rossi, C. Detection and preliminary characterization of a new rabbit calicivirus related to rabbit hemorrhagic disease virus but nonpathogenic. *J. Virol.* **70**, 8614–8623 (1996).
- Forrester, N. L., Trout, R. C. & Gould, E. A. Benign circulation of rabbit haemorrhagic disease virus on Lambay Island, Eire. *Virology* **358**, 18–22. <https://doi.org/10.1016/j.virol.2006.09.011> (2007).
- Bergin, I. L. *et al.* Novel calicivirus identified in rabbits, Michigan, USA. *Emerg. Infect. Dis.* **15**, 1955–1962. <https://doi.org/10.3201/eid1512.090839> (2009).
- Moss, S. R. *et al.* Molecular epidemiology of Rabbit haemorrhagic disease virus. *J. Gen. Virol.* **83**, 2461–2467. <https://doi.org/10.1099/0022-1317-83-10-2461> (2002).
- Le Gall-Reculé, G. *et al.* Detection of a new variant of rabbit haemorrhagic disease virus in France. *Vet. Rec.* **168**, 137–138. <https://doi.org/10.1136/vr.d697> (2011).
- Le Pendu, J. *et al.* Proposal for a unified classification system and nomenclature of lagoviruses. *J. Gen. Virol.* **98**, 1658–1666. <https://doi.org/10.1099/jgv.0.000840> (2017).
- Dalton, K. P., Nicieza, I., Abrantes, J., Esteves, P. J. & Parra, F. Spread of new variant RHDV in domestic rabbits on the Iberian Peninsula. *Vet. Microbiol.* **169**, 67–73. <https://doi.org/10.1016/j.vetmic.2013.12.015> (2014).
- Le Gall-Reculé, G. *et al.* Emergence of a new lagovirus related to rabbit haemorrhagic disease virus. *Vet. Res.* **44**, 81. <https://doi.org/10.1186/1297-9716-44-81> (2013).
- Dalton, K. P. *et al.* Variant rabbit hemorrhagic disease virus in young rabbits, Spain. *Emerg. Infect. Dis.* **18**, 2009–2012. <https://doi.org/10.3201/eid1812.120341> (2012).
- Camarda, A. *et al.* Detection of the new emerging rabbit haemorrhagic disease type 2 virus (RHDV2) in Sicily from rabbit (*Oryctolagus cuniculus*) and Italian hare (*Lepus corsicanus*). *Res. Vet. Sci.* **97**, 642–645. <https://doi.org/10.1016/j.rvsc.2014.10.008> (2014).
- Le Gall-Reculé, G. *et al.* Large-scale lagovirus disease outbreaks in European brown hares (*Lepus europaeus*) in France caused by RHDV2 strains spatially shared with rabbits (*Oryctolagus cuniculus*). *Vet. Res.* **48**, 70. <https://doi.org/10.1186/s13567-017-0473-y> (2017).
- Neimanis, A. S. *et al.* Overcoming species barriers: an outbreak of *Lagovirus europaeus* GI.2/RHDV2 in an isolated population of mountain hares (*Lepus timidus*). *BMC Vet. Res.* **14**, 367. <https://doi.org/10.1186/s12917-018-1694-7> (2018).
- Puggioni, G. *et al.* The new French 2010 Rabbit Hemorrhagic Disease Virus causes an RHD-like disease in the Sardinian Cape hare (*Lepus capensis mediterraneus*). *Vet. Res.* **44**, 96. <https://doi.org/10.1186/1297-9716-44-96> (2013).

17. Velarde, R. *et al.* 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. *Transbound. Emerg. Dis* **64**, 1750–1761. <https://doi.org/10.1111/tbed.12562> (2017).
18. Capucci, L., Cavadini, P., Schiavitto, M., Lombardi, G. & Lavazza, A. Increased pathogenicity in rabbit haemorrhagic disease virus type 2 (RHDV2). *Vet. Rec.* **180**, 426. <https://doi.org/10.1136/vr.104132> (2017).
19. Rouco, C., Aguayo-Adan, J. A., Santoro, S., Abrantes, J. & Delibes-Mateos, M. Worldwide rapid spread of the novel rabbit haemorrhagic disease virus (GI2/RHDV2/b). *Transbound. Emerg. Dis.* **66**, 1762–1764. <https://doi.org/10.1111/tbed.13189> (2019).
20. Calvete, C., Sarto, P., Calvo, A. J., Monroy, F. & Calvo, J. H. Could the new rabbit haemorrhagic disease virus variant (RHDVb) be fully replacing classical RHD strains in the Iberian Peninsula?. *World Rabbit Sci.* **22**, 91. <https://doi.org/10.4995/wrs.2014.1715> (2014).
21. Lopes, A. M. *et al.* Is the new variant RHDV replacing genogroup 1 in Portuguese wild rabbit populations?. *Viruses* **7**, 27–36. <https://doi.org/10.3390/v7010027> (2015).
22. Mahar, J. E. *et al.* Rabbit haemorrhagic disease virus 2 (GI.2) is replacing endemic strains of RHDV in the Australian landscape within 18 months of its arrival. *J. Virol.* <https://doi.org/10.1128/JVI.01374-17> (2017).
23. Rouco, C. *et al.* Epidemiology of RHDV2 (*Lagovirus europaeus*/GI.2) in free-living wild European rabbits in Portugal. *Transbound. Emerg. Dis.* **65**, e373–e382. <https://doi.org/10.1111/tbed.12767> (2018).
24. Lopes, A. M. *et al.* Full genomic analysis of new variant Rabbit Hemorrhagic Disease Virus revealed multiple recombination events. *J. Gen. Virol.* **96**, 1309–1319. <https://doi.org/10.1099/vir.0.000070> (2015).
25. Silvério, D. *et al.* Insights into the evolution of the new variant rabbit haemorrhagic disease virus (GI.2) and the identification of novel recombinant strains. *Transbound. Emerg. Dis.* **65**, 983–992. <https://doi.org/10.1111/tbed.12830> (2018).
26. Hall, R. N. *et al.* A strain-specific multiplex RT-PCR for Australian rabbit haemorrhagic disease viruses uncovers a new recombinant virus variant in rabbits and hares. *Transbound. Emerg. Dis.* **65**, e444–e456. <https://doi.org/10.1111/tbed.12779> (2018).
27. Forrester, N. L., Moss, S. R., Turner, S. L., Schirmmeier, H. & Gould, E. A. Recombination in rabbit haemorrhagic disease virus: possible impact on evolution and epidemiology. *Virology* **376**, 390–396. <https://doi.org/10.1016/j.virol.2008.03.023> (2008).
28. Esteves, P. J. *et al.* Emergence of pathogenicity in lagoviruses: evolution from pre-existing nonpathogenic strains or through a species jump?. *PLoS Pathog.* **11**, e1005087. <https://doi.org/10.1371/journal.ppat.1005087> (2015).
29. Kocher, J. F. *et al.* Bat caliciviruses and human noroviruses are antigenically similar and have overlapping histo-blood group antigen binding profiles. *MBio* <https://doi.org/10.1128/mBio.00869-18> (2018).
30. Lemaitre, E., Zwingelstein, F., Marchandeu, S. & Le Gall-Recule, G. First complete genome sequence of a European non-pathogenic rabbit calicivirus (lagovirus GI.3). *Arch. Virol.* **163**, 2921–2924. <https://doi.org/10.1007/s00705-018-3901-z> (2018).
31. Esteves, P. J. *et al.* Detection of positive selection in the major capsid protein VP60 of the rabbit haemorrhagic disease virus (RHDV). *Virus Res.* **137**, 253–256. <https://doi.org/10.1016/j.virusres.2008.07.025> (2008).
32. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003. <https://doi.org/10.1093/ve/vev003> (2015).
33. Mahar, J. E. *et al.* Benign rabbit caliciviruses exhibit similar evolutionary dynamics to their virulent relatives. *J. Virol.* **90**, 9317–9329. <https://doi.org/10.1128/JVI.01212-16> (2016).
34. Miao, Q. *et al.* Immunogenicity in rabbits of virus-like particles from a contemporary rabbit haemorrhagic disease virus type 2 (GI2/RHDV2/b) isolated in the Netherlands. *Viruses* <https://doi.org/10.3390/v11060553> (2019).
35. Lopes, A. M. *et al.* Full genome sequences are key to disclose RHDV2 emergence in the Macaronesian islands. *Virus Genes* **54**, 1–4. <https://doi.org/10.1007/s11262-017-1523-2> (2018).
36. Bull, R. A. *et al.* Norovirus recombination in ORF1/ORF2 overlap. *Emerg. Infect. Dis.* **11**, 1079–1085. <https://doi.org/10.3201/eid1107.041273> (2005).
37. Katayama, K. *et al.* Novel recombinant sapovirus. *Emerg. Infect. Dis.* **10**, 1874–1876. <https://doi.org/10.3201/eid1010.040395> (2004).
38. Coyne, K. P. *et al.* Recombination of Feline calicivirus within an endemically infected cat colony. *J. Gen. Virol.* **87**, 921–926 (2006).
39. Simmonds, P. *et al.* Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res.* **36**, 2530–2546. <https://doi.org/10.1093/nar/gkn096> (2008).
40. Meyers, G., Wirblich, C. & Thiel, H.-J. Genomic and subgenomic RNAs of rabbit hemorrhagic disease virus are both protein-linked and packaged into particles. *Virology* **184**, 677–686. [https://doi.org/10.1016/0042-6822\(91\)90437-G](https://doi.org/10.1016/0042-6822(91)90437-G) (1991).
41. Bull, R. & White, P. A. In *Caliciviruses: Molecular and Cellular Virology* (eds Hansman, G. S. *et al.*) 45–63 (Caister Academic Press, Norfolk, 2010).
42. Simmonds, P. Recombination and selection in the evolution of picornaviruses and other mammalian positive-stranded RNA viruses. *J. Virol.* **80**, 11124–11140. <https://doi.org/10.1128/JVI.01076-06> (2006).
43. Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* **84**, 3134. <https://doi.org/10.1128/JVI.01394-09> (2010).
44. Hon, C.-C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J. Virol.* **82**, 1819–1826. <https://doi.org/10.1128/JVI.01926-07> (2008).
45. Kelly, A. G., Netzler, N. E. & White, P. A. Ancient recombination events and the origins of hepatitis E virus. *BMC Evol. Biol.* **16**, 210–210. <https://doi.org/10.1186/s12862-016-0785-y> (2016).
46. Gallei, A., Pankraz, A., Thiel, H. J. & Becher, P. RNA recombination in vivo in the absence of viral replication. *J. Virol.* **78**, 6271–6281. <https://doi.org/10.1128/JVI.78.12.6271-6281.2004> (2004).
47. Ludwig-Begall, L. F., Mauroy, A. & Thiry, E. Norovirus recombinants: recurrent in the field, recalcitrant in the lab—a scoping review of recombination and recombinant types of noroviruses. *J. Gen. Virol.* **99**, 970–988. <https://doi.org/10.1099/jgv.0.001103> (2018).
48. Duchêne, S., Holmes, E. C. & Ho, S. Y. W. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B Biol. Sci.* **281**, 20140732. <https://doi.org/10.1098/rspb.2014.0732> (2014).
49. Duchêne, S., Duchêne, D., Holmes, E. C. & Ho, S. Y. W. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906. <https://doi.org/10.1093/molbev/msv056> (2015).
50. Westcott, D. G. & Choudhury, B. Rabbit haemorrhagic disease virus 2-like variant in Great Britain. *Vet. Rec.* **176**, 74. <https://doi.org/10.1136/vr.102830> (2015).
51. Lopes, A. M. *et al.* Characterization of old RHDV strains by complete genome sequencing identifies a novel genetic group. *Sci. Rep.* **7**, 13599. <https://doi.org/10.1038/s41598-017-13902-2> (2017).
52. Bull, R. A., Tanaka, M. M. & White, P. A. Norovirus recombination. *J. Gen. Virol.* **88**, 3347–3359. <https://doi.org/10.1099/vir.0.83321-0> (2007).
53. Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.* **80**, 11827–11832. <https://doi.org/10.1128/JVI.01100-06> (2006).
54. Lukashov, A. N. *et al.* Recombination in circulating human enterovirus B: independent evolution of structural and non-structural genome regions. *J. Gen. Virol.* **86**, 3281–3290 (2005).
55. Korotkova, E. *et al.* Pressure for pattern-specific intertypic recombination between sabin polioviruses: evolutionary implications. *Viruses* **9**, 353. <https://doi.org/10.3390/v9110353> (2017).

56. Gould, A. R. *et al.* The complete nucleotide sequence of rabbit haemorrhagic disease virus (Czech strain V351): use of the polymerase chain reaction to detect replication in Australian vertebrates and analysis of viral population sequence variation. *Virus Res.* **47**, 7–17 (1997).
57. Buddle, B. M. *et al.* Response of the North Island brown kiwi, *Apteryx australis mantelli* and the lesser short-tailed bat, *Mystacina tuberculata* to a measured dose of rabbit haemorrhagic disease virus. *N. Z. Vet. J.* **45**, 109–113. <https://doi.org/10.1080/00480169.1997.36004> (1997).
58. Leighton, F. A., Artois, M., Capucci, L., Gavier-Widen, D. & Morisse, J. P. Antibody response to rabbit viral hemorrhagic disease virus in red foxes (*Vulpes vulpes*) consuming livers of infected rabbits (*Oryctolagus cuniculus*). *J. Wildl. Dis.* **31**, 541–544 (1995).
59. Merchán, T. *et al.* Detection of rabbit haemorrhagic disease virus (RHDV) in nonspecific vertebrate hosts sympatric to the European wild rabbit (*Oryctolagus cuniculus*). *Infect. Genet. Evol.* **11**, 1469–1474 (2011).
60. Leeks, A., Segredo-Otero, E. A., Sanjuán, R. & West, S. A. Beneficial coinfection can promote within-host viral diversity. *Virus Evol.* **4**, vey028. <https://doi.org/10.1093/ve/vey028> (2018).
61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
64. Nurk, S. *et al.* Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737. <https://doi.org/10.1089/cmb.2013.0084> (2013).
65. Chevreaux, B., Wetter, T., & Suhai, S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99, 45–56 (1999).
66. Chen, Y., Ye, W., Zhang, Y. & Xu, Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* **43**, 7762–7768. <https://doi.org/10.1093/nar/gkv784> (2015).
67. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98. <https://doi.org/10.12691/ajmr-2-6-8> (1999).
68. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720. <https://doi.org/10.1093/bioinformatics/btv428> (2015).
69. RStudio: Integrated Development for R v. 3.6.1 (RStudio, Inc., Boston, 2015).
70. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. <https://doi.org/10.1093/molbev/mst197> (2013).

Acknowledgements

Fundação para a Ciência e Tecnologia (FCT) supported the FCT Investigator contracts of Joana Abrantes (CEECIND/00078/2017), Ana M. Lopes (CEECIND/01388/2017) and Pedro J. Esteves (IF/00376/2015). Clément Droillard's doctoral fellowship was co-funded by the Conseil Général des Côte-d'Armor and by Saint-Brieuc Armor Agglomération. This work is a result of the project LAGMED supported by FCT (PRIMA/0003/2018) and PRIMA programme, an Art.185 initiative supported and funded under Horizon 2020, the European Union's Framework Programme for Research and Innovation. This work was also supported by a grant from the French National Research Agency (ANR, Agence Nationale de la Recherche) (contracts ANR-14-ANWA-0004-01 and ANR-14-ANWA-0004-02) within the framework of the ECALEP project selected during the 2nd joint call of the ERA-Net Anihwa (Animal Health and Welfare), a Coordination Action funded under the European Commission's ERA-Net scheme within the Seven Framework Programme (contract No. 291815). We would like to thank Pascal Bihannic (FDC 29), Thierry Bouteiller, Thierry Delhorme (FDC 56), Jean-Louis Lebeau, Jacky Aubineau, Louis-Marie Pasquier, Gilles Chavas (FDC 42), Anne Pleney (IMPCF), Laurent Verrion (FDC 23), Olivier Galaup (FDC 66), Christian Bernard (FDC 66), ONCFS SD16, ONCFS SD29 and ONCFS BMI Bretagne-Pays de la Loire who contributed for collecting rabbit tissues in France. We are specially grateful to Dr. Samuel Boucher (Labovet Conseil, Les Herbiers, France) for sending the rabbit liver sample 16-35 and to Véronique Béven (GVB Unit, Ploufragan-Plouzané-Niort Laboratory, Anses, France) for constructing the 16-35 and 16-36 cDNA libraries for NGS.

Author contributions

J.A., A.M.L. and G.L.G.R. performed the analyses, discussed the data, wrote and revised the first draft of the manuscript. C.D. obtained the molecular data, performed the analyses, discussed the data and revised the first draft. E.L. obtained the molecular data, discussed the data and revised the first draft. P.L. and Y.B. analysed the NGS data. S.M. was responsible for the GI.3 strains sample collection, discussed the data and revised the first draft. P.J.E. discussed the data and revised the first draft. All authors contributed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71303-4>.

Correspondence and requests for materials should be addressed to J.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020