Supplementary Materials for

Parallel adaptation of rabbit populations to myxoma virus

Joel M. Alves[1,2,3*], Miguel Carneiro[2,4*], Jade Y. Cheng[5,6], Ana Lemos de Matos[7], Masmudur M. Rahman[7], Liisa Loog[3,8], Paula F. Campos[6,9], Nathan Wales[6,10], Anders Eriksson[11], Andrea Manica[12], Tanja Strive[13,14], Stephen C. Graham[15], Sandra Afonso[2], Diana J. Bell[16], Laura Belmont[7], Jonathan P. Day[1], Susan J. Fuller[17], Stéphane Marchandeau[18], William J. Palmer[19], Guillaume Queney[20], Alison K. Surridge[16], Filipe G. Vieira[6], Grant McFadden[7], Rasmus Nielsen[5,6], M. Thomas P. Gilbert[6,21], Pedro J. Esteves[2,22], Nuno Ferrand[2,4,23], Francis M. Jiggins[1*]

correspondence to: joel.alves@arch.ox.ac.uk, miguel.carneiro@cibio.up.pt, fmj1001@cam.ac.uk

This PDF file includes:

Materials and Methods
Tables S1 to S4
Figs. S1 to S7
References

Other Supplementary Materials for this manuscript includes the following:

Files S1 to S10

Material and Methods

Sample collection

To understand the genomic changes that occurred in the European rabbit (*Oryctolagus cuniculus*) in response to myxoma virus, we studied historical populations prior to the beginning of the epizootics, and modern populations from the same locations collected more than 50 years after. We focused on three countries where the development of genetic resistance to myxomatosis has been described: Australia, France and the United Kingdom (*6*). To mitigate confounding effects caused by existing population structure, the study area was restricted to the south of France, south of Great Britain, and southeast of Australia (Fig. 1 and Fig. S1). The map layout with the sample locations points was generated with the *R* suite (*34*) using the packages *Maps* (*35*) and *Mapdata* (*36*). For clarification purposes, the use of the term "country" refers to the country of origin regardless of being an historical or modern population. The use of the term "population" refers to an historical or modern population within a country. Sample code names were attributed based on the type of sample and location. Each sample is designated by a code of four strings of abbreviations connected by an underscore (e.g. M_AU_NSW_bh1). The first string states the sample type (M for modern and H for historical), the second refers to the country of origin (AU for Australia, FR for France and UK for the United Kingdom), the third refers to the state (Australia), department (France) or county (United Kingdom), and the fourth string represents the sample location (*e.g.* city, village) followed by the sample number (to distinguish samples collected in the same location).

The historical samples were kindly provided by the following 11 Natural History museums: from Australia, the Australian Museum, the Museum Victoria, and the Queensland Museum; from France, the Musée des Confluences and the Muséum National d'Histoire Naturelle; from the United Kingdom, the Booth Museum of Natural History and the Natural History Museum; and finally from the USA, the American Museum of Natural History, the Museum of Comparative Zoology (Harvard University), the Museum of Zoology (University of Michigan), and the National Museum of Natural History (Smithsonian Institution). We obtained samples in the form of bone, skin or residual tissue, belonging to 128 different rabbit specimens and collected between 1865 and 1956 (File S1). Six rabbits were collected shortly after the release of the virus. Excluding these from the analysis does not affect our conclusions. From these, 95 were converted into genomic libraries and sequenced. The remaining 33 samples were excluded due to extreme low DNA yields.

The modern rabbit samples from Australia and France were donated either by private hunters or collaborators in the form of purified DNA or tissue (skin clips or liver). All the rabbit samples from the United Kingdom were donated by private hunters, whose contact was facilitated in most cases by the British Association for Shooting and Conservation (BASC). In this case, liver samples were collected directly in the field and immediately stored in a falcon tube with absolute ethanol and preserved in dry ice. Tissues from all wild rabbits were taken from animals killed for recreation or pest controlling purposes by certified hunters. Therefore, no wild animal has been killed for the purpose of this study and no animal ethics permit was required. When possible, the individual GPS coordinates for each rabbit was taken and animals were sexed by visual inspection. The livers were then stored at -80ºC for future DNA extraction. Details for each sample, including sex, location and GPS coordinates

(when available) are provided in File S1. In total, 78 modern rabbit samples (26 for each population), collected between 2002 and 2013, were obtained for this study.

Rabbit samples between 1985 and 1996 from Australia and the UK were provided by collaborators in the form of purified DNA or tissue (liver or blood) (File S1).

DNA extraction from historical and modern samples

The DNA extraction of the historical samples was conducted at the ancient DNA facilities of the Centre for GeoGenetics, Natural History Museum, University of Copenhagen. These facilities integrate a set of anti-contamination procedures such as positive pressure rooms, ultraviolet light and laminar flow hoods where all sample handling is undertaken. The DNA extraction protocol was undertaken in small batches of samples, and surfaces were decontaminated with a diluted sodium hypochlorite solution and 70% ethanol before and after each batch.

Two extraction protocols were used according to the type of tissue. For bone, prior to the DNA extraction, the samples were ground manually or with a mikro-dismembrator in case of more compact bone samples. Bone powder was then mixed with 995 µl of 0.5 M EDTA and 5 µl of proteinase K solution (20 mg/ml), and left to incubate overnight at 56ºC. The digested sample was then centrifuged at 12,000 rpm for 5 min to form a pellet. The liquid portion was re-concentrated to a volume of 200-250 µl using a 30-kDa cut-off Centricon micron centrifugal filter unit (Millipore, Billerica, MA) and centrifuged at 4,000 rpm. DNA was then purified using MinElute columns (Qiagen MinElute PCR Purification Kit, Qiagen, Hilden, Germany) according to manufacturer's protocol with the following modification: during the

elution step, the spin columns with buffer EB were incubated at 37ºC for 15 minutes to increase DNA yields.

For residual dried tissue and skin samples, the hair was removed prior to the digestion to prevent blockage of the spin columns during the extraction process, and samples cut into smaller pieces to optimise the digestion step. The DNA extraction was done with Qiagen DNAeasy Blood and Tissue Kit (Qiagen, Valencia, CA), following manufacturer's protocol with the following modification: during the elution step, the spin columns with Buffer AE were incubated at 37ºC for 15 minutes to increase DNA yields.

For modern samples, the genomic DNA was extracted using the Qiagen DNAeasy Blood and Tissue Kit (Qiagen, Valencia, CA), following the manufacturer's protocol. The DNA concentration of each DNA extract was accessed with a Qubit DNA quantification system (Invitrogen, Carlsbad, CA) using Qubit high-sensitivity assay reagents.

Library Preparation

The libraries for the historical samples were constructed at the ancient DNA facilities of the Centre for GeoGenetics, Copenhagen, Denmark. From the 128 historical samples, a set of 95, for which the DNA extraction was successful, was used to make genomic libraries.

For each sample a total of 21.25 µl of DNA extract was used to construct a double stranded blunt-end Illumina library using NEBNext DNA Sample Prep Master Mix Set 2 (New England Biolabs, E6070). The protocol was followed according to

the manufacturer's instructions with the following modifications. The reaction volume was reduced by a quarter in the end-repair step and by half in the ligation and fill-in steps. The end-repair stage was performed for 30 minutes at 20ºC. The ligation reaction was performed for 25 min at 20ºC using Illumina-specific adapters specified in Meyer and Kircher (*37*). For both steps, after the incubation, the reaction was purified through MinElute spin columns and eluted with a Qiagen EB buffer volume of 15 µl in the end-repair step and 21 µl in the ligation step followed by incubation for 15 minutes at 37ºC. The final fill-in reaction was done for 20 minutes at 65ºC.

To increase library complexity, each library was amplified in two independent PCR reactions that were combined in the end. The following protocol was used for each 100 µl PCR reaction: 12 µl of unamplified library template, 2 U AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA), 1X AmpliTaq Gold buffer, 2.5 mM MgCl2, 0.2 mM dNTPs, 0.2 µM Illumina Multiplexing PCR primer inPE1.0 primer, 0.2 µM Illumina long index primer, 0.4m/mL of Bovine Serum Albumin (BSA), and water for the remaining volume. Cycling conditions were: 95.0°C for 10 min enzyme activation, 8 to 24 cycles of 95.0°C for 30 s, 60.0°C for 1 min, and 72.0°C for 40 seconds, followed by a final extension of 5 minutes at 72°C (number of cycles varied according to each sample and ranged between 8 and 24). The number of required amplification cycles was estimated based on the technique described by Nathan et al (*38*).

The long index primer contained Illumina standard indices from 1 to 24 and a set of custom indexes were used for barcoding libraries. The PCR product as purified through a single Qiagen MinElute spin column following the manufacturer protocol and eluted in 25 µl EB buffer following a 10 min incubation at 37ºC. Following

amplification, libraries tested on Agilent 2100 Bioanalyzer High Sensitivity DNA chip to evaluate the quality and insert size. Library quantification was undertaken using a qPCR KAPA Library Quantification Kit (Kapa Biosystems) according to the manufacturer's instructions.

For all modern samples, individually barcoded libraries were prepared from the purified genomic DNA at TGAC (The Genome Analysis Centre, Norwich, UK) using the KAPA LTP Library Preparation Kit for Illumina platforms (KAPA Biosystems, Boston, USA) and following the manufacturer's protocol. After PCR amplification, the libraries were quantified using qPCR KAPA Library Quantification Kit (KAPA Biosystems, Boston, USA).

Capture design, enrichment and sequencing

Exome enrichment was performed with NimbleGen solution-based captures (NimbleGen SeqCap EZ Developer Library, Roche). The technology is based on the development of millions of overlapping 50-105mer probes that cover the target region. The technique consists of hybridising the probes with the genomic libraries, followed by bead capturing of the complex of capture-oligos and genomic DNA fragments, and a final enrichment step of the captured fragments with PCR amplification.

The custom-capture design was based on the coordinates of the gene annotations of the OryCun 2.0 rabbit reference genome (Ensembl release 2.69) (*39*). In addition to the coding region, we targeted the mitochondrial genome and three regions of the Major Histocompatibility Complex region (MHC) in chromosome 12,

encompassing 1.75 Mb (20,160,00 to 21,060,000; 22,290,00 to 22,560,000; 23,000,000 to 23.580,000). The exact coordinates of targets are available as a BED file in the File S7. The targets of interest were provided to Roche Nimblegen for the final probe design. All overlapping targets were merged to create a contiguous region, and those smaller than 100 bp were padded at both sides to reach a minimum of size of 100 bp in order to increase the capture efficiency.

The capture-enrichment was undertaken in equimolar pools of individual libraries. These pools were made based on the qPCR quantifications and they were composed by either historical or modern samples. For the historical pools, samples were combined according to their insert size to avoid differential capture performance. A total of 21 pools ranging from 6 to 12-plex were made. Exome-capture was performed on each individual pool using a single capture reaction and following the manufacturer's protocol. Due to the unavailability of a rabbit specific reagent for blocking repetitive regions in the genome at the time, the universal SeqCap EZ Developer Reagent Cot-1 DNA was used following the manufacturer's recommendation. The enrichment and capture of 8 pools was conducted at TGAC (Norwich, UK), while the remaining ones at the University of Cambridge.

After capture-enrichment, each pool was independently sequenced in one lane of HiSeq 2000/2500 Illumina machine using 100bp paired-end reads at TGAC (Norwich, UK) and BGI (Beijing Genomics Institute, China). Four of the pools were sequenced twice resulting in a total of 25 lanes of HiSeq.

Sanger sequencing of candidate SNPs

To investigate if the most extreme changes in allele frequency observed in our dataset were caused by MYXV or RHDV, we genotyped SNPs in four of the strongest candidate genes/regions (*i.e.CD96, FCRL3, IFN-a21A,* and MHC region). The genotyping was made only for modern samples, before and after RHDV epizootics, using PCR followed by Sanger Sequencing using the primers described in Table S3. Genotypes are available in File S8.

Bioinformatics and processing of sequencing data

FastQC, version 0.11.2 (*40*) was used to evaluate the quality of the raw sequences and potential adaptor contamination. *Trimmomatic*, version 0.32 (*41*) was used to removed low quality bases and adaptor sequences, using the following options: *TRAILING*=15 (cut bases of the end of the read if below a threshold quality of 15), *SLIDINGWINDOW*=4:20 (perform a sliding window trimming, cutting once the average quality within the window falls below a threshold of 20), and *ILLUMINACLIP*=TruSeq3-PE.fa:2:20:10:1:true (remove adapter contamination; the values correspond in order to: input fasta file with adapter sequences to be matched, seed mismatches, palindrome clip threshold, simple clip threshold, minimum adapter length and logical value to keep both reads in case of read-through being detected in paired reads by palindrome mode). The trimming was undertaken exclusively at the 3' end to allow the identification of PCR duplicates (see below). *PEAR*, version 0.96 (*42*), was used to merge overlapping paired-end reads into single-end reads, allowing any length size of the assembled sequences.

The merged and unmerged sequence reads (fastq files) were aligned independently to the rabbit reference genome OryCun2.0 using *BWA-MEM*, version

0.7.10 (*43*) with default options with exception of the parameter –M, which marks shorter split hits as secondary a requirement for Picard compatibility. The generated SAM files were converted to their binary format (BAM) and sorted by their leftmost coordinates with *SAMtools*, version 0.1.19 (website: http://samtools.sourceforge.net/).Read Group information (RG) was added to the BAM files using the module *AddOrReplaceReadGroups* from *Picard Tools*, version 1.126 (https://broadinstitute.github.io/picard). The module *MergeSamFiles* in *Picard* was used to merge all BAM files belonging to the same sample, and the module *MarkDuplicates* was used to remove PCR and optical duplicates. Local realignments of reads around indels were performed with *GATK*, version 3.3.0 (https://www.broadinstitute.org/GATK) using with the tools *RealignerTargetCreator* (determines suspicious intervals which require realignment), and *IndelRealigner* (runs the realigner over those intervals).

In order to calculate the number of reads mapped on-target we the tool *intersectBed*, from program *BEDtools*, version 2.22.0 (*44*) was used to intersect each bam file with a BED file containing the interval coordinates of the target regions. The average percentage of reads mapped on-target was 63.64% for historical samples and 50.46% for modern samples. *SAMtools Flagstats* tool was used to obtain metrics for all generated BAM files. *GATK* module *DepthOfCoverage* was used to calculate the coverage on-target with the option "intervals" and a bed file with the capture coordinates as an input. The mean insert size was inferred with the *Picard Tools* module *CollectInsertSizeMetrics*. Sequencing metrics for all samples are available in File S9.

A specific problem is cytosine deamination in historical samples, which

results in an increase of C→T and G→A substitutions, and therefore a higher transition/transversion ratio (*45*). To mitigate this effect we followed the conventions of ancient DNA research and used a Bayesian approach to model DNA damage in each sample and reduced base quality scores in the sequence reads to account for the effect. The program *MapDamage*, version 2.06 (*46*), was used to quantify the damage patterns in these samples, followed by downscaling of the quality score of the potential post-mortem damaged bases. We used default options with exception of the number of reads per sample that was down-sampled to 100,000. Non-overlapping paired-end reads were not rescaled since the option for paired-end reads was not integrated in the program. As sites affected by damage-driven mutations would differ between samples, we also removed variants that occurred at a low frequency (minor allele frequency <0.05 across all individuals in the dataset). This resulted in a transition/transversion ratio in the historical populations that was nearly identical to modern populations at ~3.3, and similar to that observed in human exomes (*6*).

Variant calling and filtering

To avoid potential issues caused by differential mapping of historical and modern reads due to their different insert sizes, prior to variant calling we filtered out all reads with a mapping quality (MAPQ) below 40. The variant calling was then performed for each individual BAM using the *HaplotypeCaller* tool from *GATK*. Due to computational burden, the variant calling was restricted to the capture-targeted regions with a padding of 300bp around each target. We assumed that the heterozygosity (proportion of sites that are different from the reference genome) was 0.004 based on previous studies (*16*), and all the remaining parameters were kept as

default. After obtaining individual gVCF files for each sample, a joint genotyping with all historical and modern samples combined was conducted using the tool *GenotypeGVCFs* with default options.

The resulting raw variant calls (VCF file) went through a set of hard filters to remove potential false variants. The following filters were applied: *QD* < 2.0, *FS* > 60.0, *MQ* < 40.0, *MQRankSum* < -12.5, *ReadPosRankSum* < -8.0, where *QD* is the variant confidence (from the *QUAL* field) divided by the unfiltered depth of non-reference samples; *FS* is the phred-scaled p-value using Fisher's Exact Test to detect strand bias (the variation being seen on only the forward or only the reverse strand) in the reads; *MQ* is the Root Mean Square of the mapping quality of the reads across all samples; *MQRankSum* is the *U*-based *z*-approximation from the Mann-Whitney Rank Sum Test for mapping qualities (comparing reads with reference bases versus those with that have an alternate allele); and *ReadPosRankSum* is the *U*-based *z*-approximation from the Mann-Whitney Rank Sum Test for the distance from the end of the read for reads with the alternate allele (if the alternate allele is only seen near the ends of reads, this is indicative of error).

After obtaining a final VCF file, we applied a filter based on population genetics theory, where we removed all positions with a significant excess of heterozygotes assuming Hardy-Weinberg equilibrium. This calculation was made independently for modern and historical datasets and all significant variants ($P < 0.05$) inferred for the two datasets were removed. The reasoning for this filter is that variants with excess of heterozygotes are often a signature of a low mappability region where there is a combination of reads correctly and wrongly mapped. Finally, we applied a filter at the genotype level with *GATK* tool *VariantFiltration*, by keeping

only genotypes with a depth of coverage (*DP*) higher or equal to five and a genotype quality (*GQ*) higher than 20.

For all the following analyses, we only used samples with a minimum coverage on-target of 5x, resulting in final dataset of 152 samples, 75 historical and 77 modern samples (File S1). The predicted impact, the name of the gene and sequence ontology term for each variant was determined using *SnpEff*, version 4.1 (*2*, *17*, *18*). Human orthologue information, rabbit gene description, and GO Biological Process were inferred using *Better Bunny* online suite, version 2.3 ((*19*)) (*20*, *21*).

Population Genetic Analyses

To compare modern and historical populations, we undertook a set of demographic analyses. For these analyses, we generated a stricter dataset consisting of biallelic and autosomal variants that were contained within the original targets and restricted to exons. Additionally, we restricted the analysis to variants that have been called in at least 90% of the individuals across all populations. This latter filter was applied with *VCFtools*, version 0.1.12 (*47*). This resulted in a total of 220,696 SNPs.

Weir & Cockerham's F_{ST}

We quantified genetic differentiation between modern and historical populations for each country by calculating the Weir & Cockerham's F_{ST} per individual variant (*48*) using *VCFtools* (*47*). The global F_{ST} values were calculated using the Weir & Cockerham's weighted F_{ST} estimate as implemented in *VCFtools*. This analysis was restricted to positions with a minimum of 10 individuals for each population and with

a minor allele frequency > 0.05 across all individuals.

To test for parallel evolution we ranked the variants for each country by F_{ST} value and kept the top 1000 variants. The top 1000 SNPs were then intersected across the three countries to count how many of those are common between two countries and across all countries. The statistical significance of this intersection was assessed by doing 1000 random permutations of historical and modern samples within each country. We randomised historical and modern individuals within each country. We then calculated the F_{ST} between these mixed populations and took the top 1000 F_{ST} variants. We repeated this process for each country and then intersected the obtained values across the three countries. This process was then repeated 1000 times to generate a null distribution. Only variants that were polymorphic in all three countries were used in this comparison.

Changes in Allele Frequency

To estimate allele frequency for each variant in each population we took the individual genotypes of each individual and calculated the allele frequency for the reference and alternative allele for each modern and historical population across the three countries. To visualise the changes in allele frequency between historical and modern populations, we plotted the allele frequency of the reference allele of historical against modern populations. For each comparison, we removed all monomorphic variants within each country and only used variants with a minimum of 10 genotypes and a minor allele frequency of 0.05 across all individuals. The plot was generated with *R* software (*34*) using the heatscatter function in the package *LSD* (*49*).

Ohana Structure and PCA

To investigate patterns genetic structure, we started by performing a Principal Component Analysis (PCA) analysis with *Plink2*, version 1.02 (*50*). To further explore the genetic structure we used the Structure model (*51*) implemented in the *Ohana* tool suite (*52*). Briefly, *Ohana* was used to infer global ancestry and the covariance structure of allele frequencies among populations. We then modelled the joint distribution of allele frequencies across ancestry components as a multivariate Gaussian and estimated the population tree that is most compatible with the inferred covariance matrix (Fig. S2). For global ancestry and the covariance structure, we analysed *K* values ranging from 2 to 8, where *K* is the number ancestry components. For each value of *K*, we use 32 independent executions with different random seeds, and report the one that reached the highest likelihood. Only variants with a minimum of 10 genotypes and a minor allele frequency of 0.05 across all individuals were used.

Expected heterozygosity

Expected heterozygosity was inferred for each polymorphic variant in each population, and averaged across all variants. Only variants genotyped for all six populations were used. To mitigate the effect of damage-driven mutations that could increase the number of singletons due to C to T and G to A substitutions, we calculated the heterozygosity only for transversions. To obtain confidence intervals for the mean of the different chromosome arms, we resampled chromosome arms with replacement and recalculated the statistic.

Linkage Disequilibrium

To obtain an estimate of linkage disequilibrium for each population we calculated all pairwise R^2 values between pairs of SNPs less than 5Mb apart using *Plink2*, version 1.02 (*50*). Values were then averaged in non-overlapping 500bp windows for plotting. Only SNPs with a minor allele frequency of 0.05 were kept.

Selection scan comparing modern and historical samples

To detect selection, we use a new method which models allele frequency changes through time and identifies significant changes in allele frequencies using a likelihood ratio test. This method was implemented in a modified version of the software Ohana (this modified version can be obtained J.Y.C upon request). The basic model is based on the standard Structure model, as shown in Eq 1. Let g_{ij} be the genotype for individual i in marker j, coded as 0, 1, and 2, indicating homozygous for the major allele, heterozygous, and homozygous for the minor allele, respectively. Also, q_{ik} is the proportion of ancestry of the k'th ancestry component in the i'th individual, and f_{kj} is the allele frequency of the major allele in the k'th ancestry component for the j'th marker. Then, the standard log likelihood function is given by:

$$\ln(L) = \sum_i^I \sum_j^J \left(g_{ij} \cdot \ln\left(\sum_k^K (q_{ik} \cdot f_{kj})\right) + (2 - g_{ij}) \cdot \ln\left(\sum_k^K \left(q_{ik} \cdot (1 - f_{kj})\right)\right)\right) \quad \text{(Eq. 1)}$$

Here I, J, and K indicate the total number of individuals, markers, and ancestry components, respectively. We extended this model to take time-labelled data and time-dependent allele frequency changes into account. Each individual, i, has had the opportunity to change allele frequencies during t_i time units, indicating the length of time from the emergence of the infection to the time of sampling of individual i. We then allowed an individual-specific allele frequency, f_{kj}^i for each marker j and each

individual i at each ancestry component k, and redefine Equation (1) as

$$\ln(L_j) = \sum_i^I \left(g_{ij} \ln\left(\sum_k^K (q_{ik} \cdot f_{kj}^i)\right) + (2 - g_{ij}) \ln\left(\sum_k^K \left(q_{ik} \cdot (1 - f_{kj}^i)\right)\right)\right) \quad \text{(Eq. 2)}$$

where the individual specific allele frequency now is a function of the time, rescaled by the average sampling time, t_{ave}, and also depending on a parameter which measures the strength of time-dependent allele-frequency change:

$$f_{kj}^i = \min\left(1, \max\left(0, f_{kj}^{\text{est}} + \alpha \cdot \frac{t_i - t_{\text{ave}}}{t_{\text{ave}}}\right)\right) \quad \text{(Eq. 3)}$$

Here $\alpha \in \left[-\max\left(f_j^{\text{est}}\right), 1 - \min\left(f_j^{\text{est}}\right)\right]$ with $t_{ave} = {\sum_i^I t_i}/{I}$; $t_{modern} = t_{max}$ and $t_{oldest} = 0$. We used the dates of introduction of the virus as start-of-selection dates, *i.e.* 1952 in France, 1953 in the UK, and 1950 in Australia (*2*). The parameter α measures the magnitude of allele frequency changes. We use this basic model to construct three different models that can be used to test hypotheses about time dependent allele frequency change. For each lineage, *l*, in a population tree, we can associate a parameter, α, which measures the allele frequency change on that lineage. In our case, we have three lineages (lineage 1, 2, and 3) with the associated parameters α_1, α_2, and α_3.

In Model 1 (Fig. S3a and Eq. 4), for each marker j, we constructed a likelihood ratio test, where we tested the null hypothesis of $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$ against the alternative hypothesis of $H_A: \alpha_1 = \alpha_2 = \alpha_3 > 0$.

$$\frac{L(\alpha_1 = \hat{\alpha},\ \alpha_2 = \hat{\alpha},\ \alpha_3 = \hat{\alpha})}{L(\alpha_1 = 0,\ \alpha_2 = 0,\ \alpha_3 = 0)} \quad \text{(Eq. 4)}$$

In Model 2 (Fig. S3b and Eq. 5), for each marker we constructed a likelihood

ratio test, where we tested the null hypothesis of $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$ against the alternative hypothesis of $H_A: \alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$, where each scalar is estimated separately using maximum likelihood.

$$\frac{L(\alpha_1 = \widehat{a_1}, \alpha_2 = \widehat{a_2}, \alpha_3 = \widehat{a_3})}{L(\alpha_1 = 0,\ \alpha_2 = 0, \alpha_3 = 0)}$$

(Eq. 5)

In Model 3 (Fig. S3c and Eq. 6), for each marker, we constructed a likelihood ratio test, where we tested the null hypothesis of $H_0: \max\{H_0^1: \alpha_1 > 0, \alpha_2 = \alpha_3 = 0;\ H_0^2: \alpha_2 > 0, \alpha_1 = \alpha_3 = 0;\ H_0^3: \alpha_3 > 0, \alpha_1 = \alpha_2 = 0;\}$ against the alternative hypothesis of $H_A: \alpha_1 = \alpha_2 = \ \alpha_3 > 0$.

$$\frac{L(\alpha_1 = \hat{\alpha},\ \alpha_2 = \hat{\alpha}, \alpha_3 = \hat{\alpha})}{\max\{L(\alpha_1 = \widehat{a_1},\ \alpha_2 = 0, \alpha_3 = 0), L(\alpha_1 = 0,\ \alpha_2 = \widehat{a_2}, \alpha_3 = 0), L(\alpha_1 = 0,\ \alpha_2 = 0, \alpha_3 = \widehat{a_3})\}}.$$

(Eq. 6)

Manhattan plots with likelihood values from the selection analyses were plotted with the R package qqman (*53*). We permuted sample collection dates within each country 1000 times and repeated the analyses above, each time retaining the largest likelihood ratio statistic. This gave us a null distribution from which to obtain genome-wide significance.

Bayes factor analysis of parallel and population-specific selection

We examined how often a SNP is under population-specific or parallel selection by comparing our selection models using Bayes factors. First, we used the approach described above to analyse data from each of the three populations independently to

detect SNPs that were under selection (genome-wide significance <0.05 from the permutation test). Using the combined dataset of all three populations, we then calculated Bates factors (K) in support of selection acting in one (K_1), two (K_2) or three (K_3) populations, where: K2 = Pr(Data | selection in two populations)/Pr(Data | selection in one populations); K3 = Pr(Data | selection in three populations)/Pr(Data | selection in two populations) and K1 = 1/ K2. For a variant initially detected as selected in Population1:

$$\Pr(\text{Data} \mid \text{selection in one populations} = L(\alpha_1 = \widehat{\alpha_1}, \alpha_2 = 0, \alpha_3 = 0)$$

(Eq. 7)

$$\Pr(\text{Data} \mid \text{selection in two populations}) = \frac{L\left(\alpha_1 = \widehat{\alpha_{1,2}},\ \alpha_2 = \widehat{\alpha_{1,2}}, \alpha_3 = 0\right)}{2} + \frac{L\left(\alpha_1 = \widehat{\alpha_{1,3}},\ \alpha_2 = 0, \alpha_3 = \widehat{\alpha_{1,3}}\right)}{2}$$

(Eq. 8)

$$\Pr(\text{Data} \mid \text{selection in three populations}) = L(\alpha_1 = \hat{\alpha},\ \alpha_2 = \hat{\alpha}, \alpha_3 = \hat{\alpha})$$

(Eq. 9)

Bayesian allele trajectory analysis

We used a Bayesian method for inferring the timing and strength of selection from ancient DNA genotypes described in Loog et al. (*19*), where the likelihood of the data is a function of the frequency curve of the selected allele in the population and is calculated as the product of the probabilities of all observed alleles:

$$L = \prod_i f(t_i)^{x_i}\left(1 - f(t_i)\right)^{2-x_i} \qquad \text{(Eq. 10)}$$

Here L is the likelihood of the data, $f(t)$ is the derived allele frequency at the time t, t_i is the age of the sample i, and x_i is the number of copies of the derived allele observed in sample i.

We used a standard selection model with dominance h of the advantageous allele (h = 0, h=1/2 and h=1 corresponds to recessive, additive and dominant effect, respectively) and time-dependent selection strength $s(t)$ to describe the allele frequency trajectories:

$$\frac{df}{dt} = s(t)f(1-f)[h + (1-2h)f], \qquad \text{(Eq. 11)}$$

where f is the frequency of the derived allele and s is the selection coefficient. The initial value for $f(t)$ is the ancestral frequency ($f_{\text{ancestral}}$). We assume that selection starts at time t_{start} and that selection is zero before this time (and, as a consequence, that allele frequency is constant at level $f_{\text{ancestral}}$). To accommodate change in selection through time, we allowed the selection coefficient to change at the time t_{change} ($t_{\text{change}} > t_{\text{start}}$). The selection coefficient was held constant at level s_1 until time t_{change}, after which it was held constant at level s_2.

We assumed that selection began (at time t_{start}) with the first appearance of the myxoma virus (i.e. 1950 in Australia and 1953 in the UK), and then changed (at time t_{change}) when the RHDV virus was reported in the early 1990s (1995 in Australia and 1992 in the UK). For each combination of country (Australia and UK), gene (MHC, *IFN-a21A*, *FCRL3* and *CD96*), and dominance model (recessive, additive or dominant), the remaining model parameters ($f_{\text{ancestral}}$, s_1 and s_2) were inferred by making a full parameter sweep where we calculated the deterministic allele frequency trajectory for each parameter combination. We assumed a uniform prior for the

ancestral frequency ($f_{ancestral}$ in the range [0, 1] in steps of 0.01), and the two selection coefficients (s_1 and s_2). We then calculated the marginal posterior probability density distribution for each parameter by numerically integrating the likelihood of the data over the remaining parameters. In addition, to visualize the inferred allele frequency trajectories, we calculated the posterior distribution of the allele frequency trajectory through time by weighting the allele frequency trajectory of each parameter combination (sampled from their prior distributions) by the likelihood of the data given the parameters (*19*).

The analysis was implemented in the statistical environment R v. 3.2.2 (*34*). The R code is available from GitHub (https://github.com/LiisaLoog/Rabbit-Selection) and from L.L. upon request. The deterministic allele frequency curves were calculated using the *lsoda* function in the R package *deSolve* v. 1.12 (*54*).

Interferon-α21A assays

We synthesised two N-His6 tagged IFN-α21A proteins (File S10). The wild type IFN-α21A protein sequences matches the Orycun2.0 reference genome. The variant type differs in three non-synonymous mutations at coding sequence position 523 (G to A, Val175Ile), 542 (G to T, Trp181Leu) and 544 (C to A, Gln182Lys). The codon-optimised genes were cloned into the vector pUC57 and expressed in yeast (*Pichia pastoris*) by GenScript HK Limited (Piscataway, NJ, USA), and affinity purified.

Rabbit kidney cell line RK13 was purchased from Sigma, USA (Cat no. 00021715). RK13-E3 (RK13 cells stably expressing Vaccinia virus E3 protein was cultured as described before (*22*). The cell lines were cultured in DMEM

supplemented with 10% FBS, 2mM glutamine, and 100µg of penicillin-streptomycin/ml. The cells were maintained at 37^{o}C in a humidified 5% CO2 incubator. Construction of wild-type MYXV, vMyx-GFP (WT-MYXV expressing GFP under a poxvirus synthetic early/late promoter) and M029 knockout virus, vMyx-M029KO-GFP were described before (*22*). The construction of vMyx-M029KO-FLuc-tdTomato virus (M029 knockout virus expressing Firefly luciferase under a poxvirus synthetic early/late promoter), and a tandem-dimer tomato red fluorescence protein (Tr-FP) driven by poxvirus p11 late promoter) was described previously (*55*). vMyx-M029KO-GFP and vMyx-M029KO-FLuc-tdTomato viruses were grown in RK13-E3 cell lines.

RK13 cells were seeded into 48 well plates at 2x10^{4} cells per well. Next day, individual wells were treated with either IFN-α21A or varIFN-α21A. The IFNs were serially diluted 2-fold starting from 4ng/ml in the same DMEM media used for growing cells and added to the wells (200µl per well) replacing the original media. Next day, cells were infected with vMyx-GFP, vMyx-M029KO-GFP, VSV-GFP viruses at a multiplicity of infection (MOI) of 0.01 in the presence of IFNs. One hour after virus adsorption, the media was removed, washed and incubated with the respective IFNs containing media. In case of VSV virus, the incubation was with DMEM media without any IFNs. The progression of GFP-expressing virus infection was monitored using an inverted fluorescence microscope. Cells were harvested 48 hpi by replacing IFNs containing media with regular DMEM media and stored at -80^{o}C until processed. The cells were freeze-thawed at -80^{o}C and 37^{o}C for three times and sonicated for one minute in ice water to release the viruses from the infected cells. vMyx-GFP and vMyx-M029KO-GFP viruses were tittered by serial dilutions using RK13 and RK13-E3 cells respectively. The foci were counted 48 hpi. For VSV,

virus containing media was collected from each well and stored at -80^{o}C until titration was done. VSV was tittered by serial dilutions using RK13 cells and agarose overlay on the infected cells. The VSV plaques were counted after 24 hpi.

RK13 cells were seeded into 48 well plates at 2x10^{4} cells per well. Next day, the media in each well was replaced with either IFN-α21A or varIFN-α21A containing DMEM (200µl per well) after making 2-fold serial dilutions starting from 4ng/ml. After over-night incubation, cells were infected with vMyx-M029KO-FLuc-tdTomato virus at a MOI of 0.01 in the presence of IFNs. After 48 hpi cells were washed with 1x PBS and lysed using 1x cell lysis buffer (Promega, USA) for 20 mins at room temperature. Luciferase assay reagent was added to the lysate and luciferase reading was taken immediately after adding the substrate using a microplate reader.

VPS4 Assays

Cell lines and viruses: HEK293 cells (Parental) were grown in Dulbecco's Modified Eagle's medium (DMEM, VWR Life Sciences, Radnor, PA, USA) supplemented with 10% FBS (Sigma-Aldrich, St. Louis, MO, USA), 2mM glutamine, 100 U/mL streptomycin, and 100 mg/mL penicillin (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) (complete DMEM). Ecdysone-responsive stable HEK293 cell lines expressing wild-type VPS4 (Wild Type) and dominant-negative VPS4 (EQ mutant) (*30*) were cultured in complete DMEM supplemented with 400 µg/ml Zeocin (Invivogen, San Diego, CA, USA) and 800 µg/ml G418 (Caisson Labs, Smithfield, UT, USA). All the cells were maintained at 37oC in a humidified 5% CO2 incubator. Two myxoma virus constructs, previously described, were used: vMyx- tdTomato, expressing the tandem-dimer tomato red fluorescent protein driven by a synthetic

early/late poxvirus promoter (*56*) and vMyx-FLuc-tdTomato, expressing firefly luciferase (driven by a synthetic early/late poxvirus promoter) and tomato red (driven by poxvirus p11 late promoter) (*57*).

Flow cytometry analyses and fluorescence microscopy on HEK293 cells (VPS4): HEK293 Parental, Wild Type and EQ mutant cell lines were seeded into 6-well plates at $5x10^5$ cells per well. On the following day, cells were treated with and without 1 μM ponasterone A (ponA) (Santa Cruz Biotechnology, Dallas, TX, USA). After 20-24 hours, cells were mock treated or infected with vMyx- tdTomato MOI 10. Cells were harvested at 24 and 48 hours post-infection by cell dissociation with trypsin. Before harvesting cells at 48 hours post-infection, fluorescence microscopy was used to capture live images of infected cells (tdTomato+) on a Leica DMI6000 B inverted microscope (LAS X software) at 10x magnification. Sample viability was assessed using the Live/Dead® Fixable Near-IR Dead Cell Stain kit (Molecular ProbesTM, Thermo Fisher Scientific, Waltham, MA, USA). Infection and cell viability data were collected on a BD LSRFortessaTM flow cytometer with BD FACSDiva software (BD Biosciences, San Jose, CA, USA). Data were analyzed using FlowJo v10 software. A total of three independent experiments was performed, and for each experiment the corresponding time points were collected in duplicate.

Luciferase assay for HEK293 cells (VPS4): HEK293 Parental, Wild Type and EQ mutant cell lines were seeded into 12-well plates at $1.5\ x10^5$ cells per well. Next day, cells were treated with and without 1 μM ponasterone A (ponA) (Santa Cruz Biotechnology, Dallas, TX, USA). After 20-24 hours, cells were infected vMyx-FLuc-tdTomato at a multiplicity of infection (MOI) of 10. Samples were harvested at 3, 6, 12 and 24 h post-infection by vigorous pipetting, then centrifuged, washed with

1x DPBS and preserved as cell pellets at -80ºC until further processing. Before performing luciferase assay, cell pellets were lysed using 1x cell culture lysis reagent (Promega, Madison, WI, USA) for 20 mins at room temperature. Luciferase assay system (Promega, Madison, WI, USA) was used for reporter quantification following manufacturer's recommendations. Each sample was quantified in triplicate and a total of three independent experiments was performed.

MFSD1 Assays

Cell lines and viruses: RK13 (European rabbit kidney, BVDV negative) cell line (Sigma-Aldrich, St. Louis, MO, USA) was grown in Dulbecco's Modified Eagle's medium (DMEM, VWR Life Sciences, Radnor, PA, USA) supplemented with 10% FBS (Sigma-Aldrich, St. Louis, MO, USA) and 2mM glutamine (Gibco, Thermo Fisher Scientific, Waltham, MA, USA). Construction of vMyx-FLuc-tdTomato expressing firefly luciferase was described before (Zemp).

Custom siRNAs were custom designed with dTdT overhang at Dharmacon (United Kingdom. The designs are available in Table S4. siRNA transfection and luciferase assay on RK13 cells (*MFSD1*): RK13 cell line was used to screen for eight different small interfering RNA (siRNA) duplexes targeting European rabbit *MFSD1* (Dharmacon, Lafayette, CO, USA). A functional siRNA against Luciferase was used as a control, FLuc-S1 Positive Control DsiRNA (IDT, Coralville, IA, USA). Cells were plated at 30-40 % confluence per well of a 48-well plate a day before transfection with targeting siRNA, each at a 50 nM concentration, using Lipofectamine RNAiMAX (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). After 48 hours, cells were infected with vMyx-FLuc-tdTomato at an MOI 3.

After 1 hour of virus adsorption, cells were washed with 1x DPBS and fresh media was added. At 24 hours post-infection, cells were washed with 1x DPBS and collected by scrapping. Cell pellets were preserved at -80ºC until further processing. Later, pellets were lysed using 1x cell culture lysis reagent (Promega, Madison, WI, USA) for 20 mins at room temperature before performing the luciferase assay (Promega, Madison, WI, USA), following manufacturer's recommendations. Each sample was quantified in triplicate and a total of three independent experiments was performed.

siRNA transfection cytotoxicity on RK13 cells (MFSD1): To measure cell viability (cytotoxicity assay) following siRNA transfection, CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS) was purchased from Promega (Madison, WI, USA). siRNA transfection (adjusted to a 96-well plate), duration and concentration were applied as previously described. After 48 hours of siRNA delivery, the tetrazolium substrate (MTS) was added to RK13 cells and the A490 formazan product that is produced in viable cells was measured using a microplate reader after 1 hour of incubation. Each sample was quantified in triplicate and a total of three independent experiments was performed.

Fig. S1 – Map of sample locations in Australia, France and United Kingdom. Maps correspond to our study area for historical (left) and modern (right) samples. Dots and names show sample locations. Locations in bold have more than one sample.

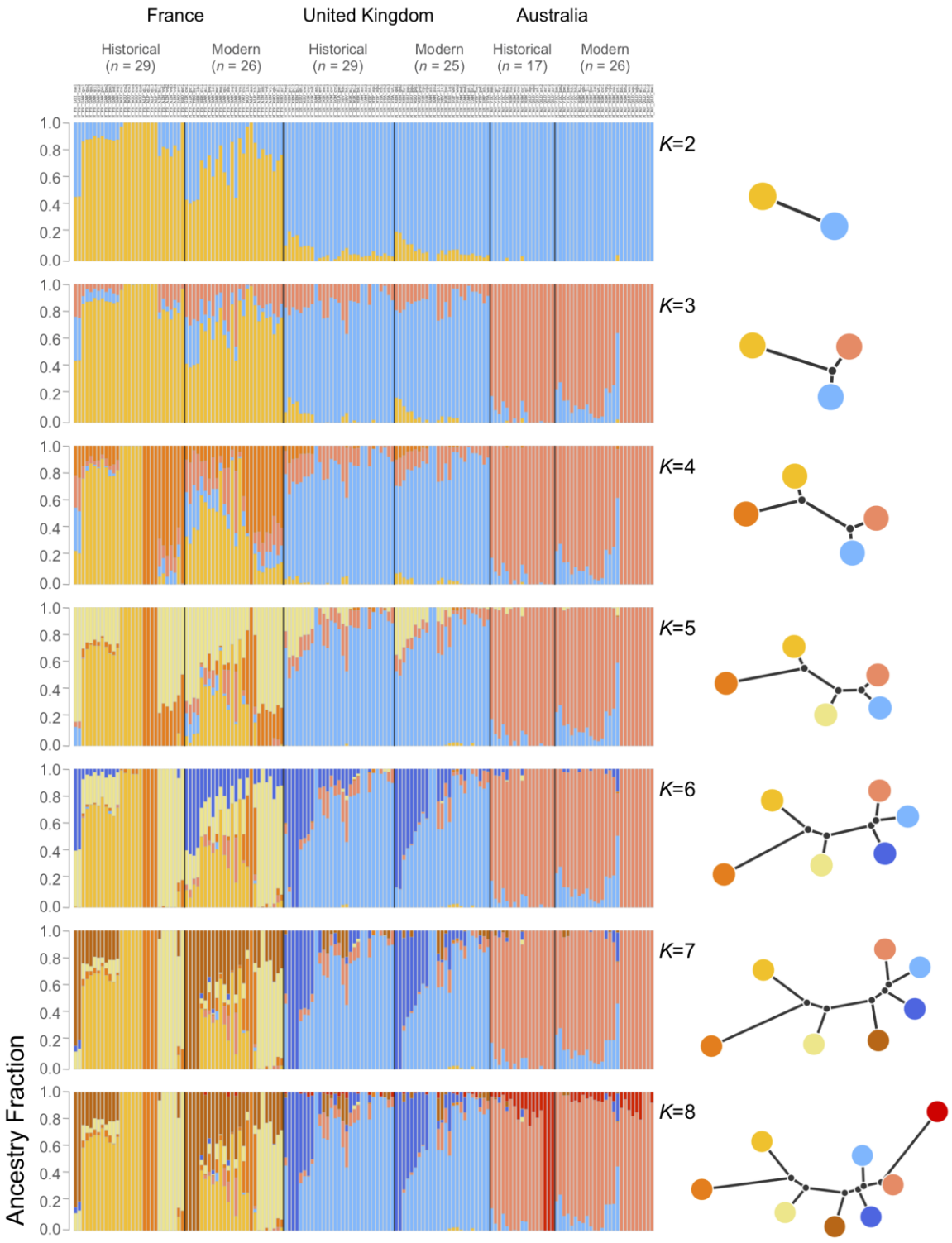

Fig. S2 – Ancestry proportions and trees estimates inferred with *Ohana* from *K*=2 until *K*=8. On the left, ancestry proportions where each bar shows the inferred ancestry fraction for an individual. The black lines between bars separate populations. Labels on the top identify country, temporal set, sample size and individuals. Individuals are ordered geographically within each population. On the right, population trees for the ancestry components that are most compatible with the inferred covariance matrix for each respective value of *K*. Branch length is an indication of drift.

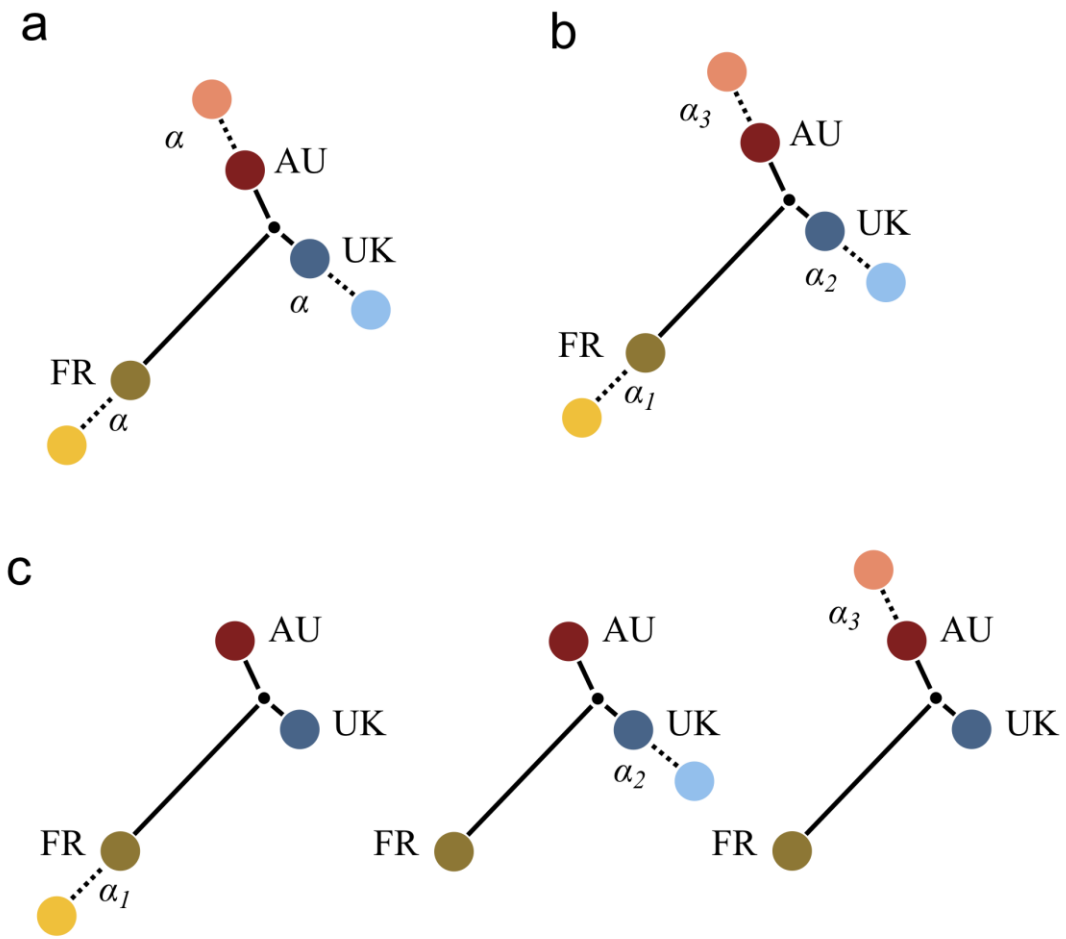

Fig. S3 – Models implemented in Ohana. (a). In Model 1, we test the null hypothesis of $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$ against the alternative hypothesis of $H_A: \alpha_1 = \alpha_2 = \alpha_3 > 0$. **(b)**. In Model 2, we test the null hypothesis of $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$ against the alternative hypothesis of $H_A: \alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$. **(c)**. In Model 3, we test the null hypothesis of $H_0: \max\{H_0^1: \alpha_1 > 0, \alpha_2 = \alpha_3 = 0;\ H_0^2: \alpha_2 > 0, \alpha_1 = \alpha_3 = 0;\ H_0^3: \alpha_3 > 0, \alpha_1 = \alpha_2 = 0;\}$ against the alternative hypothesis of $H_A: \alpha_1 = \alpha_2 = \alpha_3 > 0$.

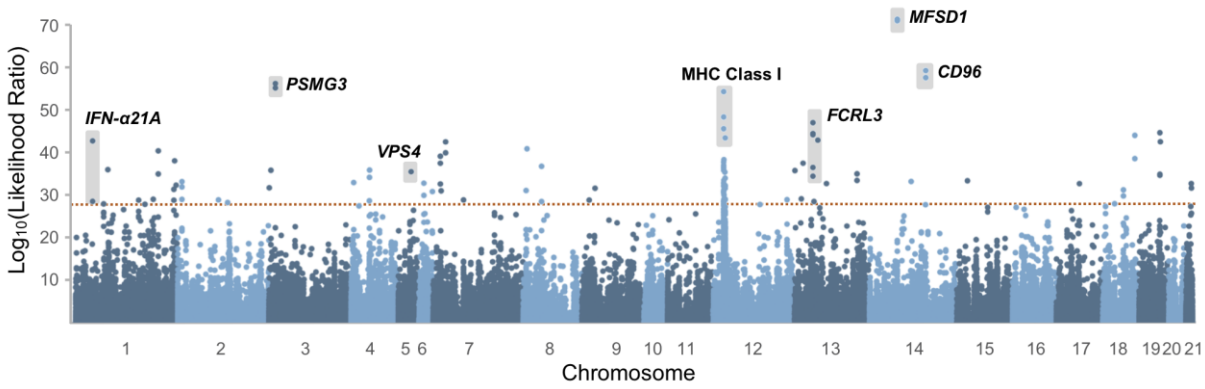

Fig. S4 – Genome-wide selection scan based on allele frequency changes after the introduction of MYXV. Selection is assumed to be the same across the three countries (Equation 4, supplementary methods) Y-axis shows likelihood ratio values for the tested model. The orange dotted line shows the genome-wide 95% significance, which was derived by threshold from permuting sample collection dates within each country. Shaded boxes show SNPs located in the highlighted genes. Different shades of blue show chromosomes.

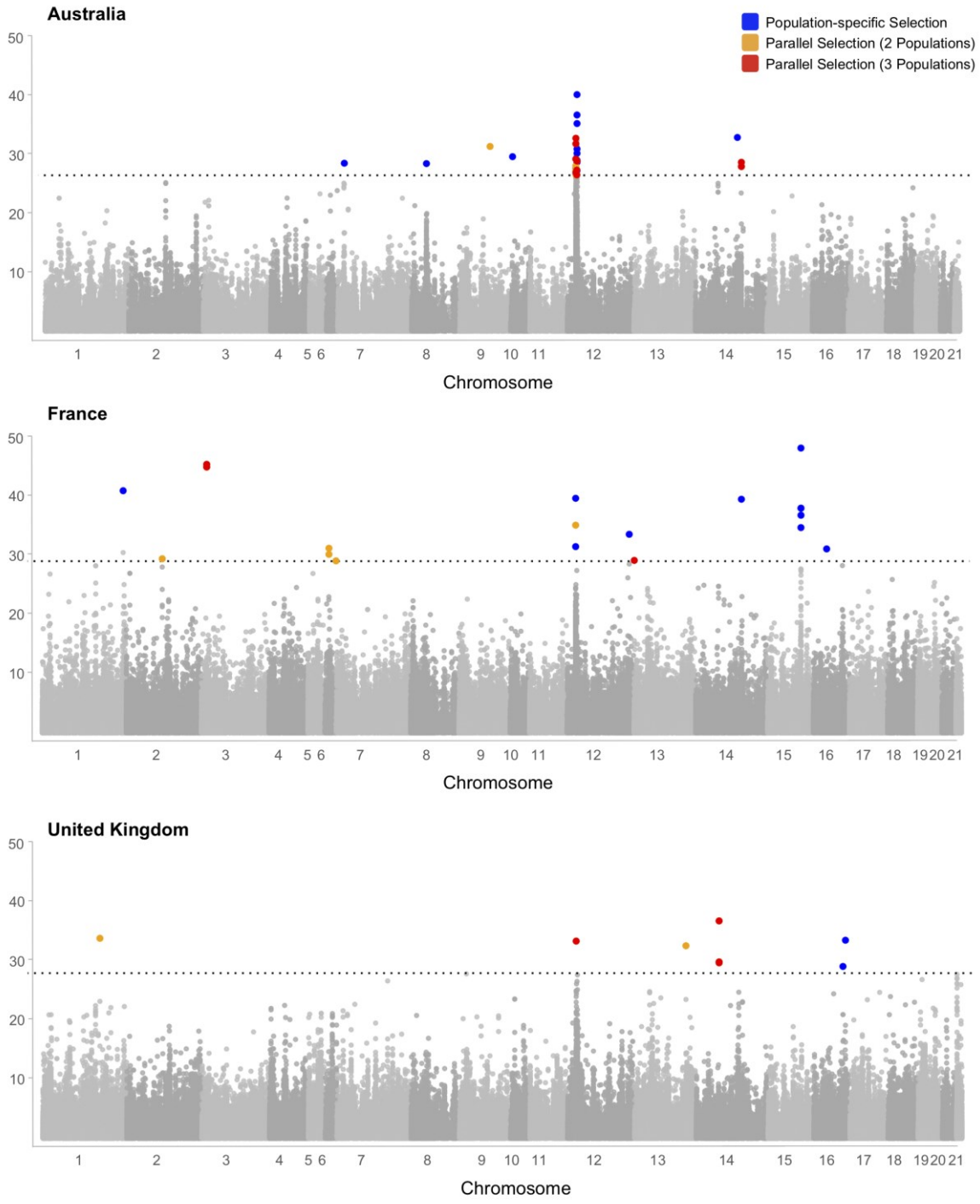

Fig. S5 - Genome-wide scan for selection in individual populations with SNPs coloured according to evidence of parallelism or population-specific selection. The horizontal dotted line shows genome-wide 95% significance threshold from permuting sample collection dates within each country. For SNPs above the significant threshold dotted line, those coloured in blue are inferred to be under selection only in that population (Bayes factor >= 3). SNPs coloured in orange and red are inferred to be under selection either in two or the three populations, respectively (Bayes factor >= 3). Different shades of grey show chromosomes. Note all SNPs are shown, not just the most significant SNP per gene.

42

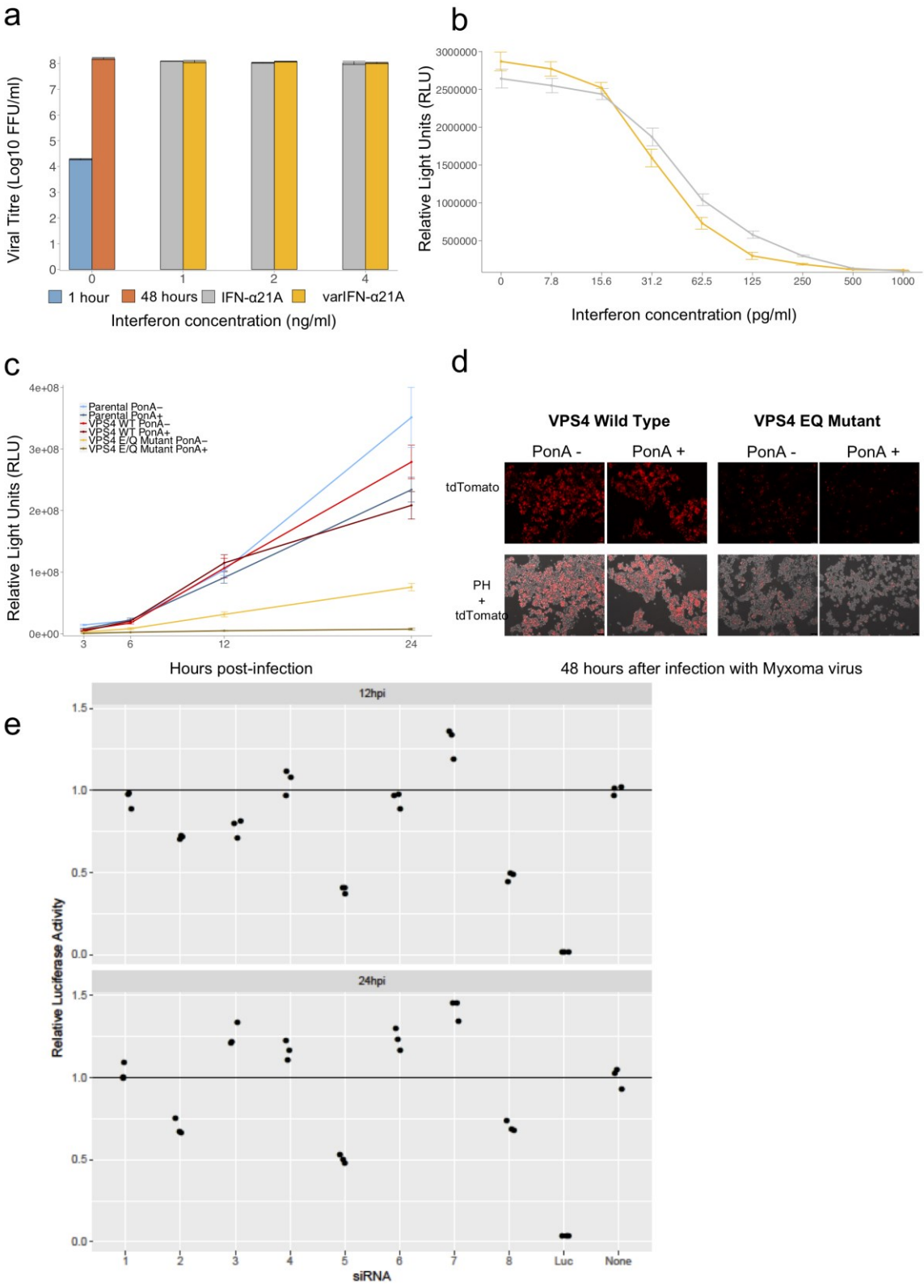

Fig. S6 – Effect of *IFN-α21A*, *VPS4* and *MFSD1* on viral titre. (a)Viral titre measured after 1 hour of infection (blue bars) and after 24/48 hours (orange bars), and after treatment with different concentrations of the wild type IFN-α21A (grey bars) and variant varIFN-α21A (yellow bars) for MYXV wild type (vMyx-GFP). Error bars

show standard error of the mean. Statistical significance between wild and variant interferon on viral titer was tested with a Student's t-test across the three replicates (* $P<0.05$; ** $P<0.01$). **(b)** Viral gene expression inferred with a luciferase essay for IFN-α21A (grey bars) and variant varIFN-α21A (yellow bars) for myxoma virus (vMyx-M029KO-FLuc-TdTomato) at different concentrations. Error bars show standard error of the mean. **(c)** Myxoma virus gene expression on infected HEK293 ecdysone-responsive stable cell lines expressing wild-type VPS4 (Wild Type), dominant-negative VPS4 (EQ Mutant), and on infected control cell line (Parental, no transfection) was evaluated by using a recombinant virus expressing firefly luciferase (vMyx-FLuc-tdTomato). Cells were either kept untreated (PonA-) or pre-treated (PonA+) with 1 μM PonA for 20-24 hours, followed by infection at multiplicity of infection (MOI) 10. At 3, 6, 12 and 24 hours post-infection cells were collected and luciferase assay was performed. Error bars show standard error of the mean. **(d)** Fluorescence microscope images of VPS4 Wild Type and EQ mutant HEK293 cells treated with and without PonA (PonA+ and PonA-, respectively), 48 hours after infection with wild-type myxoma virus expressing a red fluorescent protein (vMyx-tdTomato) at MOI 10. Red fluorescence and phase contrast images of live cells were collected 48 hours post-infection using an inverted fluorescence microscope at 10x magnification. **(e)** Effect of MFSD1 on MYXV titre using luciferase. The expression of MFSD1 was knocked down using eight different siRNAs in rabbit RK13 cells. The negative control had no siRNAs, and the positive control had an siRNA against luciferase. Cells were infected with vMyx-FLuc-tdTomato at an MOI 3, and viral titres measured by luciferase fluorescence 24 and 48 hours post infection. The results of three independent biological replicates of each treatment are shown. Viral titres are normalised to the control.

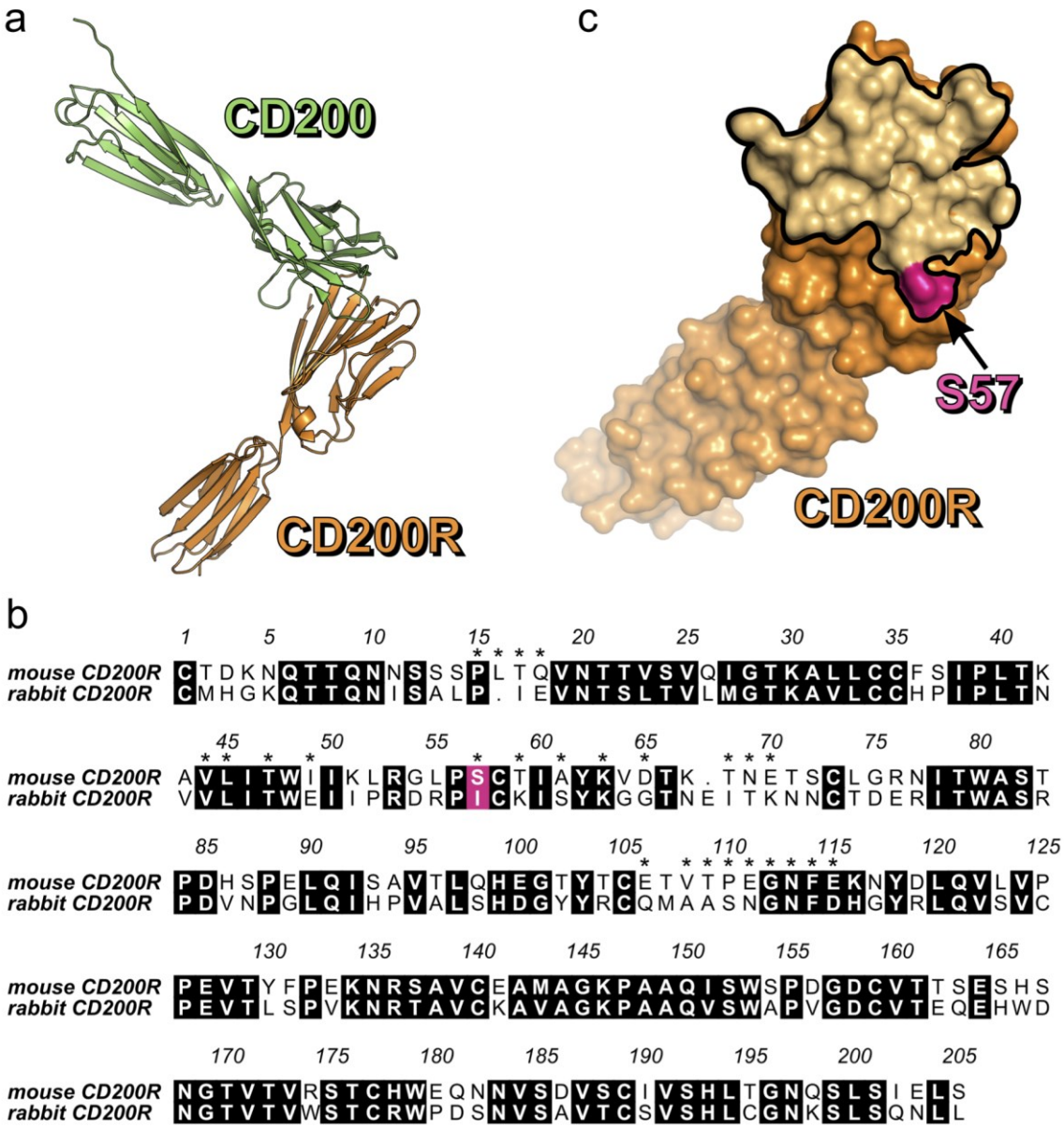

Fig. S7 – Structure of the CD200:CD200R complex. **(a)** The structure of the mouse CD200:CD200R complex (PDB ID 4BFI (*23*)) is shown as ribbons. Mouse CD200 and CD200R share high sequence identity with their rabbit homologues (78% and 53% across their structured ectodomains, respectively) and the CD200R binding interface of CD200 is strongly conserved across mammals (*23*), suggesting that the rabbit CD200:CD200R interaction will closely resemble the mouse complex. **(b)** Sequence alignment of mature mouse and rabbit CD200R. Residues of mouse CD200R that interact with CD200 are starred, and the site of the rabbit CD200R non-synonymous variant (isoleucine to threonine) that is selected in the French population is highlighted. Alignment was generated using Clustal Omega (*58*) and illustrated using *ALINE* (*59*) **(c)** Surface representation of the CD200R, with the CD200-binding interface of CD200R outlined. Mouse CD200R residue S57, equivalent to residue under selection in the French population, is highlighted. Molecular images were generated using *PyMOL* (Schrödinger LLC).

Table S1 – Global Fixation Index (F_{ST}) estimated between population pairs. Only polymorphic loci and a minimum of 10 individual genotypes were used for each two-population comparison.

Comparison	Population 1	Population 2	Variants	Mean F_{ST}
Historical versus Modern	France	France	424,922	0.020
	United Kingdom	United Kingdom	466,141	0.009
	Australia	Australia	345,335	0.014
Historical versus Historical	France	Australia	370,971	0.142
	France	United Kingdom	416,041	0.128
	United Kingdom	Australia	367,526	0.067
Modern versus Modern	France	Australia	887,287	0.095
	France	United Kingdom	968,815	0.085
	United Kingdom	Australia	850,659	0.060

Table S2 – Frequency of the selected allele of SNPs under population-specific selection.

Population selection detected in[1]	*N*	Mean frequency in Historical Population[2]		
		Australia	**France**	**UK**
Australia	9	**0.05**	0.10	0.12
France	9	0.12	**0.07**	0.09
United Kingdom	2	0.00	0.00	**0.04**

[1] Variants were identified as having experienced a significant change in frequency since the release of the myxoma virus (genome-wide $p<0.05$ when each population analysed separately) and having evidence for population specific selection (Bayes factor >3 when comparing a model of selection in 1 population with a model of selection in 2 populations). Within each gene region only the most significant SNP was retained.

[2] The mean frequency of the SNPs among historical samples. The frequency in the population where selection has acted is highlighted in bold

Table S3 – Sanger sequencing primer sequences used for genotyping variants at genes/regions *CD96*, *FCRL3*, *IFN-α21A*, MHC, *PSMG3*.

Gene/Region	Chromosome	Position	Primer Forward	Primer Reverse	Fragment Size (bp)
CD96	14	107622993	GGATCTGTAATGTACTTAAGTGACTGG	GATACTACCCTAATGCAAGGAGTC	158
FCRL3	13	35513450	CTCTGAGGCCCGGACTGT	AGGCCATTGTCAGCCTCA	167
IFNA21A	1	32518812	GCTGTGAGGAAATACTTCCAAGG	CAGGAATCATTTTATGTTGGTCCT	128
MHC	12	20266487	GAGCCTGAGACCTCATCCAG	TGTTCAGAGAGCCCCAACAT	464
PSMG3	3	-	GGACTGCGGGTCTCTTGTC	CACTGCTCCTGTCCTTCACA	429

Table S4 – siRNAs for *MFSD1* gene.

siRNA	Chromosome	Position	siRNA (Sense)	siRNA (Antisense)	Target Sequence
MFSD1	14	1278	GUUUGUAGUUCCUGAGCAUdTdT	AUGCUCAGGAACUACAAACdTdT	GTTTGTAGTTCCTGAGCAT
MFSD2	14	394	CUGUAUGCCUGGUAUUCUUdTdT	AAGAAUACCAGGCAUACAGdTdT	CTGTATGCCTGGTATTCTT
MFSD3	14	628	GUGAGUUGGUUUAAAGGCAdTdT	UGCCUUUAAACCAACUCACdTdT	GTGAGTTGGTTTAAAGGCA
MFSD4	14	955	GUCAUCUGUGUCUGCUAUUdTdT	AAUAGCAGACACAGAUGACdTdT	GTCATCTGTGTCTGCTATT
MFSD5	14	754	GGUUCUGCUGGUCAUACAAdTdT	UUGUAUGACCAGCAGAACCdTdT	GGTTCTGCTGGTCATACAA
MFSD6	14	366	GCAGGUGAAUACCACGAAAdTdT	UUUCGUGGUAUUCACCUGCdTdT	GCAGGTGAATACCACGAAA
MFSD7	14	1048	GCAAGUGCAAUUAACAGUAdTdT	UACUGUUAAUUGCACUUGCdTdT	GCAAGTGCAATTAACAGTA
MFSD8	14	694	GCAGUACAGUAAACAUGAAdTdT	UUCAUGUUUACUGUACUGCdTdT	GCAGTACAGTAAACATGAA

Additional Supplementary Material (Separate Files)

File S1 – List of all samples used in this study.

File S2 – List containing the top 1000 SNPs based on F_{ST} for each population.

File S3 – List containing the top 1000 SNPs based on *Ohana* (Parallelism).

File S4 – List containing the top 1000 SNPs based on *Ohana* (Population-specific).

File S5 – Bayes factors for most significant SNP per gene (Population-specific).

File S6 – Likelihood ratios and parameter estimation for SNP trajectories over time.

File S7 – BED file with probe coordinates for the exome capture.

File S8 – Sanger and exome-sequencing genotypes for pre-myxomatosis, pre-RHDV and post-RHDV samples for four SNPs located at *IFN-α21A*, *FCRL3*, *CD96* genes and at the MHC region. Sanger sequencing genotypes of four SNPs, one insertion, and one deletion, located in PSMG3 gene for 68 modern individuals from Australia, France, the UK and the rabbit cell line RK13.

File S9 – Sequencing metrics for exome sequencing data.

File S10 – Wild type and variant Interferon-α21A (IFN-α21A) protein sequences.

References:

1. F. Fenner, F. N. Ratcliffe, *Myxomatosis* (Cambridge University Press, Cambridge ; New York, 1965).

2. F. Fenner, B. Fantini, *Biological Control of Vertebrate Pests: The History of Myxomatosis - an Experiment in Evolution.* (CABI publishing, New York, NY, USA, 1999).

3. J. Ross, M. F. Sanders, The development of genetic resistance to myxomatosis in wild rabbits in Britain. *J Hyg (Lond)*. **92**, 255–261 (1984).

4. F. Fenner, J. Ross, in *The European rabbit: The History and Biology of a Successful Colonizer*, H. V. Thompson, C. M. King, Eds. (Oxford University Press, Oxford ; New York, 1994), pp. 205–239.

5. I. D. Marshall, F. Fenner, Studies in the epidemiology of infectious myxomatosis of rabbits. V. Changes in the innate resistance of Australian wild rabbits exposed to myxomatosis. *J Hyg (Lond)*. **56**, 288–302 (1958).

6. P. J. Kerr, Myxomatosis in Australia and Europe: a model for emerging infectious diseases. *Antiviral Res.* **93**, 387–415 (2012).

7. P. J. Kerr *et al.*, Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proc Natl Acad Sci USA*. **114**, 9397–9402 (2017).

8. E. M. Veale, The Rabbit in England. *Agric Hist Rev*. **5**, 85–90 (1957).

9. K. Myers, I. Parer, P. Wood, B. D. Cooke, in *The European rabbit: The History and Biology of a Successful Colonizer*, H. V. Thompson, C. M. King, Eds. (Oxford University Press, Oxford ; New York, 1994), pp. 108–157.

10. M. Carneiro, N. Ferrand, M. W. Nachman, Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). *Genetics*. **181**, 593–606 (2009).

11. L. Kruglyak, Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*. **22**, 139–144 (1999).

12. A. V. S. Hill, Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Phil. Trans. R. Soc. B*. **367**, 840–849 (2012).

13. M. M. Magwire *et al.*, Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in Drosophila melanogaster. *PLoS Genet*. **8**, e1003057 (2012).

14. J. C. Stephens *et al.*, Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).

15. P. J. Kerr *et al.*, Evolutionary history and attenuation of myxoma virus on two continents. *PLoS Pathog.* **8**, e1002950 (2012).

16. J. Abrantes, W. van der Loo, J. Le Pendu, P. J. Esteves, Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): a review. *Vet Res*. **43**, 12 (2012).

17. H. E. Fuller, D. Chasey, M. H. Lucas, J. C. Gibbens, Rabbit haemorrhagic disease in the United Kingdom. *Vet. Rec.* **133**, 611–613 (1993).

18. J. P. Morisse, G. Le Gall, E. Boilletot, Hepatitis of viral origin in Leporidae: introduction and aetiological hypotheses. *Rev Sci Tech*. **10**, 269–310 (1991).

19. L. Loog *et al.*, Inferring Allele Frequency Trajectories from Ancient DNA Indicates That Selection on a Chicken Gene Coincided with Changes in Medieval Husbandry Practices. *Mol Biol Evol*. **34**, 1981–1990 (2017).

20. G. L. Smith, J. A. Symons, A. Alcamí, Poxviruses: interfering with interferon. *Semin Immunol*. **8**, 409–418 (1998).

21. C. Upton, K. Mossman, G. McFadden, Encoding of a homolog of the IFN-gamma receptor by myxoma virus. *Science*. **258**, 1369–1372 (1992).

22. M. M. Rahman, J. Liu, W. M. Chan, S. Rothenburg, G. McFadden, Myxoma virus protein M029 is a dual function immunomodulator that inhibits PKR and also conscripts RHA/DHX9 to promote expanded host tropism and viral replication. *PLoS Pathog.* **9**, e1003465 (2013).

23. D. Hatherley, S. M. Lea, S. Johnson, A. N. Barclay, Structures of CD200/CD200 receptor family and implications for topology, regulation, and evolution. *Structure*. **21**, 820–832 (2013).

24. D. Hatherley, A. N. Barclay, The CD200 and CD200 receptor cell surface proteins interact through their N-terminal immunoglobulin-like domains. *Eur. J. Immunol.* **34**, 1688–1694 (2004).

25. A. L. Hughes, M. Yeager, Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**, 415–435 (1998).

26. Y. Kochi *et al.*, FCRL3, an autoimmune susceptibility gene, has inhibitory potential on B-cell receptor-mediated signaling. *J. Immunol.* **183**, 5502–5510 (2009).

27. L. Martinet, M. J. Smyth, Balancing natural killer cell activation through paired receptors. *Nat. Rev. Immunol.* **15**, 243–254 (2015).

28. A. Fuchs, M. Colonna, The role of NK cell recognition of nectin and nectin-like proteins in tumor immunosurveillance. *Semin. Cancer Biol.* **16**, 359–366 (2006).

29. G. Sivan *et al.*, Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis. *Proc Natl Acad Sci USA*. **110**, 3519–3524 (2013).

30. C. M. Crump, C. Yates, T. Minson, Herpes simplex virus type 1 cytoplasmic envelopment requires functional Vps4. *J Virol*. **81**, 7380–7387 (2007).

31. F. I. Schmidt *et al.*, Vaccinia virus entry is followed by core activation and proteasome-mediated release of the immunomodulatory effector VH1 from lateral bodies. *Cell Rep*. **4**, 464–476 (2013).

32. S. Gandon, Y. Michalakis, Evolution of parasite virulence against qualitative or quantitative host resistance. *Proc. Biol. Sci.* **267**, 985–990 (2000).

33. G. Stack *et al.*, CD200 receptor restriction of myeloid cell responses antagonizes antiviral immunity and facilitates cytomegalovirus persistence within mucosal tissue. *PLoS Pathog.* **11**, e1004641 (2015).

34. R Core Team, "R: A Language and Environment for Statistical Computing" (R Foundation for Statistical Computing., Vienna, Austria, 2015), (available at http://www.R-project.org/).

35. R. A. Becker, A. R. Wilks, R. Brownrigg, T. P. Minka, "Maps: draw geographical maps. R package version 2.3-9" (2010).

36. R. Brownrigg, "Mapdata: Extra Map Databases, R Package Version 2.2-5" (2013).

37. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*. **2010**, pdb.prot5448 (2010).

38. N. Wales *et al.*, New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. *BioTechniques*. **59**, 368–371 (2015).

39. M. Carneiro *et al.*, Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*. **345**, 1074–1079 (2014).

40. S. Andrews, FastQC: a quality control tool for high throughput sequence data (2010), (available at http://www.bioinformatics.bbsrc. ac.uk/projects/fastqc).

41. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

42. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. **30**, 614–620 (2014).

43. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

44. I. M. H. Aaron R Quinlan, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

45. A. W. Briggs *et al.*, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).

46. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–1684 (2013).

47. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics*. **27**, 2156–2158 (2011).

48. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution*. **38**, 1358 (1984).

49. B. Schwalb *et al.*, Package "LSD" (2015).

50. C. C. Chang *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. **4**, 7 (2015).

51. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. **155**, 945–959 (2000).

52. J. Y. Cheng, T. Mailund, R. Nielsen, Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics*. **33**, 2148–2155 (2017).

53. S. D. Turner, qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots, 1–2 (2014).

54. K. Soetaert, T. Petzoldt, R. W. Setzer, Solving differential equations in R: package deSolve. *Journal of Statistical Software* (2010).

55. M. M. Rahman, G. McFadden, Myxoma Virus dsRNA Binding Protein M029 Inhibits the Type I IFN-Induced Antiviral State in a Highly Species-Specific Fashion. *Viruses*. **9** (2017), doi:10.3390/v9020027.

56. J. Liu *et al.*, Myxoma virus expressing interleukin-15 fails to cause lethal myxomatosis in European rabbits. *J Virol*. **83**, 5933–5938 (2009).

57. F. J. Zemp *et al.*, Treating brain tumor-initiating cells using a combination of myxoma virus and rapamycin. *Neuro-oncology*. **15**, 904–920 (2013).

58. F. Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2011).

59. C. S. Bond, A. W. Schüttelkopf, ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 510–512 (2009).