

Pathogen evolution and disease emergence in carnivores

Alex J. McCarthy*, Marie-Anne Shaw and Simon J. Goodman*

Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

Emerging infectious diseases constitute some of the most pressing problems for both human and domestic animal health, and biodiversity conservation. Currently it is not clear whether the removal of past constraints on geographical distribution and transmission possibilities for pathogens alone are sufficient to give rise to novel host–pathogen combinations, or whether pathogen evolution is also generally required for establishment in novel hosts. Canine distemper virus (CDV) is a morbillivirus that is prevalent in the world dog population and poses an important conservation threat to a diverse range of carnivores. We performed an extensive phylogenetic and molecular evolution analysis on complete sequences of all CDV genes to assess the role of selection and recombination in shaping viral genetic diversity and driving the emergence of CDV in non-dog hosts. We tested the specific hypothesis that molecular adaptation at known receptor-binding sites of the haemagglutinin gene is associated with independent instances of the spread of CDV to novel non-dog hosts in the wild. This hypothesis was upheld, providing compelling evidence that repeated evolution at known functional sites (in this case residues 530 and 549 of the haemagglutinin molecule) is associated with multiple independent occurrences of disease emergence in a range of novel host species.

Keywords: canine distemper virus; molecular adaptation; zoonoses; pathogen evolution; haemagglutinin; SLAM receptor

1. INTRODUCTION

Emerging infectious diseases (EIDs) constitute some of the most pressing problems facing human and domestic animal health, and biodiversity conservation (Daszak *et al.* 2000). Such diseases have obvious implications for human welfare, and also cause enormous economic impacts, and in the context of conservation, can drive rapid population declines, or even extinctions (Daszak *et al.* 2000; Woolhouse & Gowtage-Sequeria 2005). Recent examples, which have received high levels of attention, include epidemics of SARS disease (Stadler *et al.* 2003), the invasion of West Nile virus into the Americas (Nash *et al.* 2001), the potential for the emergence of new influenza pandemic strains (Monto 2005), and the role of the fungal disease chytridiomycosis in global amphibian declines (Daszak *et al.* 1999). The drivers of disease emergence are still poorly understood. Globalization of trade and travel, along with human impingement into previously undisturbed habitats, can remove past constraints on geographical distributions and transmission possibilities. However, whether such factors alone give rise to novel host–pathogen combinations, or whether pathogen evolution is generally required for establishment in novel hosts or for changes to virulence, remains a major unanswered question (Altizer *et al.* 2003). Resolving this question has important implications for surveillance, predicting impacts and planning responses to novel EIDs.

The powerful statistical methods now available for the analysis of molecular evolution make it possible to detect

signatures of selection in DNA/RNA sequences from pathogen populations, but linking such signals or specific genetic polymorphisms to the process of disease emergence is more problematic (Yang *et al.* 2000; Haydon *et al.* 2001; Woelk *et al.* 2001; Anishchenko *et al.* 2006). Rather than simply identifying sites that might be under selection, this requires a hypothesis-testing approach where variation at sites known to be functionally significant in the infection or virulence process, for example at residues involved in interactions with specific host receptors (and therefore cell types or species), can be tested for association with host switches or modulation of virulence. To date this has been achieved for only a very small number of pathogens, generally with clear implications for human health, and most represent only single instances of host switches. For example, variation at residues 226 and 228 of the haemagglutinin protein (HA) in influenza A H2 and H3 serotypes is a determinant of receptor specificity, while variation at residues 182 and 192 of HA in influenza H5N1 is likely to be involved in the adaptation of H5N1 to human hosts (Connor *et al.* 1994; Yamada *et al.* 2006); and Anishchenko *et al.* (2006) have identified a mutation that allows an equine avirulent strain of Venezuelan equine encephalitis virus (VEEV, a zoonotic human pathogen) circulating in rodents to infect and amplify in horses. There are few robust examples of such evolution in emerging wildlife diseases or of evolution associated with multiple independent switches to a range of different hosts by the same pathogen (Altizer *et al.* 2003). The best characterized examples include the sudden emergence of canine parvovirus (CPV), which infects domesticated dogs, wolves and coyotes, in 1978 and its subsequent development into a global pandemic. CPV evolved from feline parvovirus (FPV; Hueffer & Parrish 2003) and the jump into a canine host

* Authors for correspondence (bgy2ajm@leeds.ac.uk; s.j.goodman@leeds.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.0884> or via <http://www.journals.royalsoc.ac.uk>.

range was facilitated by substitutions in the FPV capsid protein (e.g. VP2 residues K93N and D323N) which allowed CPV to use the canine transferrin receptor to infect canine host cells (Chang *et al.* 1992; Hueffer *et al.* 2003). Recently, a T249P amino acid substitution in the NS3 helicase of American isolates of West Nile Virus (WNV) has been found to modulate increased virulence of WNV in corvids (Brault *et al.* 2007). Diversification of lentiviruses in primates probably constitutes the best example of adaptation associated with multiple independent emergence events in different host species (Hahn *et al.* 2000). Other examples of positive selection and host adaptation do exist, e.g. for rabies (Holmes *et al.* 2002), but for most cases the specific functional roles of the residues involved have not yet been elucidated. In order to understand the broader role of pathogen evolution in disease emergence, more studies are required on pathogens that infect a diverse host range, and have well-characterized molecular and cell biology.

Canine distemper virus (CDV), a morbillivirus in the paramyxovirus family, has high prevalence in the world dog population, causing a range of symptoms including nasal and conjunctival discharge, respiratory congestion, fever, immunosuppression and neurological damage. CDV poses an important conservation threat to many carnivore species. Spillover resulting from interactions between domestic or feral dogs and wild species has led to mass mortalities in species ranging from wild canids (African wild dog, *Lycaon pictus*, bat-eared fox, *Otocyon megalotis*; Carpenter *et al.* 1998), felids (lion, *Panthera leo*; Harder *et al.* 1996), hyaenids (spotted hyaena, *Crocuta crocuta*; Haas *et al.* 1996), phocids (Caspian seal, *Phoca caspica*; Baikal seal, *Phoca sibirica*; Mamaev *et al.* 1995; Kennedy *et al.* 2000), mustelids (black-footed ferret, *Mustela nigripes*; Williams *et al.* 1988), viverrids (palm civet, *Parguma larvata*; Machida *et al.* 1992), ailurids (red panda, *Ailurus fulgens*) to procyonids (raccoon, *Procyon lotor*; Lemberger *et al.* 2005). In many cases CDV is a direct threat to the continued persistence of small populations of conservation concern, having extirpated the last remnant wild population of the black-footed ferret in 1985 (Williams *et al.* 1988), and causing recurrent mortality among African wild dogs (Alexander & Appel 1994). In addition, CDV may have contributed to the extinction of the Tasmanian tiger (*Thylacinus cynocephalus*) at the beginning of the twentieth century (Guiler 1961). CDV therefore meets the criterion of having a diverse host range in which it is possible to test repeated independent emergence and to evaluate the role of evolutionary changes in such events.

Morbilliviruses also have the benefit of having well-characterized cell and molecular biology, mostly based on the work on measles virus (MV) and CDV. The morbilliviruses have a non-segmented single-stranded negative sense RNA genome encoding six proteins. Haemagglutinin (H) and fusion (F) glycoproteins form the virus envelope and are crucial to cellular infection via attachment to signalling lymphocyte activation molecule (SLAM) and CD46 cellular receptors (Tatsuo *et al.* 2001). The H protein binds to one or more receptors resulting in cellular attachment and activation of the F protein by tissue-specific proteases leading to cellular infection, therefore these proteins have key roles in determining host range and tropism (Lamb & Kolakofsky 2001; Seki *et al.* 2003). These molecules are

well characterized for MV and CDV, with predicted tertiary structures available, including the identification of key functional residues. In particular, sites 530 and 548 of the haemagglutinin protein have been shown experimentally to determine host cell tropism *in vitro* (Vongpunsawad *et al.* 2004; von Messling *et al.* 2005). The genome is packed in a ribonucleoprotein (RNP) structure composed of the nucleocapsid protein (N), phosphoprotein (P) and RNA-dependent RNA-polymerase (L). The matrix protein (M) links the RNP to the virus envelope.

Previous evolutionary studies of CDV have largely focused on phylogenetic relationships among different strains, but lack consensus, using a variety of complete and partial sequences from different genes, and a range of phylogenetic methodologies (Hashimoto *et al.* 2001; Martella *et al.* 2006, 2007). Studies of complete *H* gene sequences have identified seven distinct geographically separated clusters of CDV isolates, but some relationships among these lineages remain ambiguous. No studies have attempted to investigate patterns of selection or recombination or how these relate to the many instances of emergence in non-dog hosts.

Here we perform extensive phylogenetic and molecular evolution analyses using full-length sequences for all CDV genes in order to determine the role of selection and recombination in shaping viral genetic diversity. We test the specific hypothesis that molecular adaptations at the receptor-binding sites 530 and 548 of the haemagglutinin gene, which have a functional role in determining host cell tropism, are associated with independent incidents of the spread of CDV to non-dog hosts. This hypothesis was upheld for residue 530, and we additionally identify residue 549 in the SLAM-binding region as another strong candidate, providing compelling evidence that repeated pathogen evolution at functional sites is associated with multiple, independent incidents of disease emergence in a range of novel hosts.

2. MATERIAL AND METHODS

(a) Datasets

Sequences were extracted from the GenBank database. Datasets for the analysis of selection and recombination include CDV sequences acquired from dogs and non-dog hosts. Datasets were constructed for *F* ($n=13$, 1782 bp), *H* ($n=73$, 1812 bp), *L* ($n=9$, 6552 bp), *M* ($n=11$, 1005 bp), *N* ($n=12$, 1569 bp) and *P* ($n=9$, 1521 bp) genes (Accession numbers available in table S1 in the electronic supplementary material and figure 1) and comprise all distinct full-length sequences publicly available at the time of writing. Sequences were aligned using CLUSTAL X (Thompson *et al.* 1997).

(b) Selection analysis

To investigate selection pressures on gene sequences, the maximum likelihood method of Yang *et al.* (2000) was implemented. The method calculates d_N/d_S ratios (ω value) taking into account the phylogenetic relationship between taxa, and codon usage and transition/transversion biases for a number of models (Yang 1997). The M0 model averages the ω value between 0 and 1. The M1a model of neutral evolution calculates the proportion of sites that are conserved ($\omega=0$, p_0) and neutrally evolving ($\omega=1$, p_1). The M2a model extends the M1a model by incorporating a third class of sites

(p_2) that can estimate $\omega > 1$, therefore taking into account positive selection. The M3 model of positive selection estimates the ω ratio for three classes of sites (p_0 , p_1 and p_2). A second model of neutral evolution, the M7 model, calculates values of ω between 0 and 1 for 10 categories using a discrete β distribution (controlled by parameters p and q). The M8 model extends the M7 model of neutral evolution by the addition of two parameters that estimate $\omega > 1$ for an additional class of sites, and therefore takes into account positive selection. Comparison of likelihood values for nested models by likelihood ratio tests (LRTs) determines if models of positive selection (M2a, M3 and M8) are significantly more likely than models of neutral evolution (M1a and M7). Bayesian methods are used to locate specific sites that have $\omega > 1$ with high posterior probabilities. The analysis was implemented using the CODEML program of the PAML package (Yang 1997).

(c) Recombination analysis

The LDHAT program was used to estimate the population recombination rate (ρ) and population mutation rate (Watson's θ) using a composite-likelihood method from a set of aligned sequences (McVean *et al.* 2002). Estimations are based on biallelic sites where the frequency of alleles was higher than 0.1. The significance of recombination rates are tested against a null hypothesis of no recombination ($\rho = 0$) via a likelihood permutation-based approach. Relative estimates of the population recombination rate are robust to violations in key assumptions of the model, such as the presence of selection (Richman *et al.* 2003; Smith & Fearnhead 2005). Alternative measures of recombination r^2 and $|D'|$ (Awadalla *et al.* 1999) were also assessed. Tests for these measures are based on the correlation of linkage disequilibrium (LD) with distance between pairs of polymorphic sites, under the expectation that a significant decline in LD with distance signifies that recombination has occurred within a population (McVean *et al.* 2002). The program GENECONV, v. 1.81 (Sawyer 1989, 1999), was used to identify intragenic recombination/gene conversion events.

(d) Phylogenetic analysis

The most appropriate models of molecular evolution for use in the phylogenetic analyses were identified using the MODELGENERATOR program (Keane *et al.* 2006). Phylogenetic trees for the *H* gene were constructed using Bayesian and Maximum likelihood approaches. Bayesian trees were obtained using MRBAYES v. 3.1.1. (Huelsbeck & Ronquist 2001). Three replicate analyses were run for a minimum of 1 000 000 generations to ensure convergence of the MCMC chains, with an approximate 20% burn-in period assessed by the standard deviation between chains falling below 0.05. Maximum likelihood phylogenies with support derived from 1000 bootstrap replicates were constructed in TREEFINDER (Jobb *et al.* 2004).

3. RESULTS

(a) Detection of selection

For the CDV *H* gene, 73 CDV sequences isolated from dogs and non-dog hosts (see figure 1) were analysed, excluding vaccine sequences AF259552, AF378705, AB212966 and Z35493. Likelihood scores were highest in models of positive selection (M2a, M3 and M8; Table S2 in the electronic supplementary material). 0.09% of sites had

a ω value of 2.51, and 2.3% of sites had a ω value of 1.886 for the M3 and M8 models, respectively. This suggests that a small proportion of residues are under the influence of positive selection. LRTs were performed between nested models to identify the most probable model. The M7 model was rejected in favour of the M8 model with high probability ($\chi^2 = 12.01$, $p = 0.0025$), demonstrating that the observed ω ratio is significantly greater than 1 (table 1). Under the M8 model of positive selection, 10 individual residues were identified as being under the influence of positive selection ($\omega > 1$) by Bayesian analysis (Sites 29 (Pr $\omega > 1 = 0.803$), 178 (0.541), 180 (0.929), 225 (0.608), 386 (0.552), 412 (0.669), 475 (0.784), 530 (0.973), 549 (0.865) and 603 (0.868)). The Bayesian posterior support for some sites is relatively low when taken alone, but support for the key 530 site and 549 (which is also in the SLAM-binding region) is among the highest identified. Site 548 was not identified as being under selection. We report all sites with $\omega > 1$ as the phylogenetic distribution of these sites among CDV strains is informative. Bayesian posterior probability values for sites with Pr ≥ 0.75 are shown in italics.

Analysis of full-length gene sequences for CDV *F*, *L*, *M*, *N* and *P* genes revealed no ω ratios that were significantly greater than 1 based on LRTs, and therefore positive selection cannot be detected in these genes. Within the *F* gene, a minority of sites (approx. 24%) were confirmed as evolving neutrally (where $\omega = 1$; table S2 in the electronic supplementary material), an observation supported by the rejection of the M0 model of purifying selection by LRT (where $\omega < 1$) in favour of models incorporating neutral evolution (table 1).

For the *M* gene, the highest likelihood scores were observed in the models of positive selection (M2a, M3 and M8; table S2 in the electronic supplementary material). However, none of these models could be accepted in favour of nested models of neutral evolution (table 1). A high proportion of sites in this gene are highly conserved (approx. 98%) and likely to be constrained by purifying selection (table S2 in the electronic supplementary material). LRTs suggest that the *L* gene is highly conserved (table 1), with a small proportion (4.5–8.2%) of sites evolving neutrally (table S2 in the electronic supplementary material). Positive selection was not significant in the analysis of the *N* gene dataset, again the M0 model was rejected in favour of models of neutral evolution (table 1), and this suggests that the majority of sites are conserved (approx. 92%; table S2 in the electronic supplementary material). Analysis of the *P* gene shows no significant signal of positive selection (table 1). The M0 model was rejected in favour of neutral models. A high proportion of sites are conserved (approx. 72%) and a proportion of sites could be under weak positive selection ($\omega = 1.18341$; table S2 in the electronic supplementary material), but this cannot be confirmed as the M8 model could not be accepted over the M7 model.

(b) Detection of recombination

Estimates of the population recombination rate (ρ) for CDV genes ranged from 1.202 to 23.835 (table 2). The strongest signal of recombination was identified in the *F* gene (table 2), returning values of $\rho = 23.835$ and $|D'| = -0.0516$, both of which were significantly different from zero ($p < 0.01$). Permutation tests of r^2 yielded borderline significant results ($p = 0.049$) for the CDV *L*

Table 1. Likelihood ratio tests (LRTs) between models for selection analysis on CDV. (LRTs are calculated by taking twice the difference between the two models of codon evolution and comparing with a χ^2 distribution. Significant *p* values (in italics) indicate that the null hypothesis (first model) can be rejected in favour of the alternative hypothesis (second model), e.g. for M7 versus M8, M7 is rejected in favour of M8.)

| gene | LRT | χ^2 score | <i>p</i> value | gene | LRT | χ^2 score | <i>p</i> value |
|----------|----------|----------------|----------------|---------------|---------------|----------------|----------------|
| <i>F</i> | M0vM1 | 70.790504 | <i>0.0001</i> | <i>H</i> | M0vM1 | 135.373318 | <i>0.0001</i> |
| | M1vM2 | 0.000676 | 0.9997 | | M1vM2 | 0 | 0.9999 |
| | M0vM3 | 70.790516 | <i>0.0001</i> | | M0vM3 | 140.154214 | <i>0.0001</i> |
| | M1vM3 | 0.000688 | 1.0000 | | M1vM3 | 4.780896 | 0.3105 |
| | M2vM3 | 0.000012 | 1.0000 | | M2vM3 | 4.780896 | 0.0916 |
| | M7vM8 | 0.827042 | 0.6613 | | M7vM8 | 12.012246 | <i>0.0025</i> |
| | <i>L</i> | M0vM1 | 51.520978 | | <i>0.0001</i> | <i>M</i> | M0vM1 |
| M1vM2 | | 0 | 1.0000 | M1vM2 | 1.746534 | | 0.4176 |
| M0vM3 | | 51.958598 | <i>0.0001</i> | M0vM3 | 14.716234 | | <i>0.0053</i> |
| M2vM3 | | 0.43762 | 0.9793 | M2vM3 | 3.138318 | | 0.5350 |
| M1vM3 | | 0.43762 | 0.8035 | M1vM3 | 1.391784 | | 0.4986 |
| M7vM8 | | 0.207754 | 0.9013 | M7vM8 | 3.932918 | | 0.1400 |
| <i>N</i> | | M0vM1 | 14.062176 | <i>0.0009</i> | <i>P</i> | | M0vM1 |
| | M1vM2 | 0.439956 | 0.8025 | M1vM2 | | 0.279998 | 0.8694 |
| | M0vM3 | 14.062176 | <i>0.0071</i> | M0vM3 | | 15.371228 | <i>0.0040</i> |
| | M1vM3 | 0.439956 | 0.9791 | M1vM3 | | 0.279998 | 0.9911 |
| | M2vM3 | 0 | 1.0000 | M2vM3 | | 0 | 1.0000 |
| | M7vM8 | 0.476686 | 0.7879 | M7vM8 | | 0.29092 | 0.8646 |

Table 2. Recombination analysis in CDV. (Calculation of $4N_e r$ (ρ), correlation scores for r^2 and $|D'|$ against distance, and permutation tests to evaluate if each statistic was significantly different from zero were performed in the LDHAT package. Significant *p* values from permutations tests, indicating observed test statistics are significantly greater than zero, are shown in italics.)

| gene | $\rho(4N_e r)$ | r^2 | $ D' $ | permutation tests | | |
|----------|----------------|----------|----------|-------------------|---------------|---------------|
| | | | | P_{ik} | P_{r^2} | $P_{ D' }$ |
| <i>F</i> | 23.835 | -0.01458 | -0.05126 | <i>0.0001</i> | 0.1210 | <i>0.0070</i> |
| <i>H</i> | 6.000 | -0.03171 | -0.03335 | 0.9090 | 0.9470 | 0.9230 |
| <i>L</i> | 3.653 | -0.01171 | -0.01131 | 0.1290 | <i>0.0490</i> | 0.1160 |
| <i>M</i> | 0.000 | 0.00527 | -0.00717 | 0.0720 | 0.5100 | 0.2560 |
| <i>N</i> | 1.202 | -0.02523 | -0.3226 | 0.8370 | 0.1870 | <i>0.0440</i> |
| <i>P</i> | 15.158 | 0.01471 | 0.03176 | 0.4550 | 0.7880 | 0.9460 |

gene, and a similar borderline result was returned for the *N* gene with the $|D'|$ statistic ($p=0.044$). No test statistics were significantly different from zero for the *H*, *M* and *P* genes. GENECONV identified two pairs of recombinant sequences in the *F* gene from dog isolates (AY964108 : AY964114 and AY964112 : AY964114), and none for the *H* gene. In the *F* gene, the recombinant sections lay between nucleotides 907 and 1361 (amino acids 302–453) of the coding domain in both pairs of recombinant sequences.

(c) Phylogenetic analysis of CDV *H* gene sequences and distribution of sites under positive selection

In order to map the phylogenetic distribution of sites identified as being under positive selection in the *H* gene, we reconstructed the phylogenetic relationships between CDV strains based on the nucleotide alignment of complete *H* gene sequences (figure 1). Bayesian posterior probabilities and maximum likelihood bootstrap values within continental clades are not shown for clarity of the figure, but were consistently high.

Our analysis identifies the same geographical clades for CDV isolates that have been previously reported (Hashimoto *et al.* 2001; Martella *et al.* 2006, 2007).

However, the topology derived in our analysis, rooted with phocine distemper virus (PDV-1) strains, has consistently higher support values than previous studies and places the Arctic-like clade basal to the Asia-2 clade.

Mapping the distribution of substitutions at sites identified to be under the influence of positive selection on to the phylogeny showed that the majority of amino acid substitutions were concentrated in non-dog host isolates (figure 1). In particular, substitutions at sites 530 and 549 predominantly occur in CDV isolates obtained from novel host species, indicating that the spread of CDV into to these non-dog hosts is associated with evolution at these sites. Six different residues (D, E, G, N, R and S) are observed at residue 530 of the CDV *H* protein (figure 1). The majority of CDV isolates from dogs (46/52) have either 530G or 530E, where 530E is distributed in all Asia-2 isolates. 530G is observed in Asia-1 and European isolates, as well as the entire America-2 cluster excluding two raccoon isolates (figure 1). Out of 21 non-dog hosts, 9 (host species: raccoon; ferret; mink, *Mustela vison*; red fox, *Vulpes vulpes*; red panda and Baikal seal) have substitutions of E/G530 to R, D or N residues. This provides further support along with experimental evidence that site 530 could be important in determining host tropism (Seki *et al.* 2003).

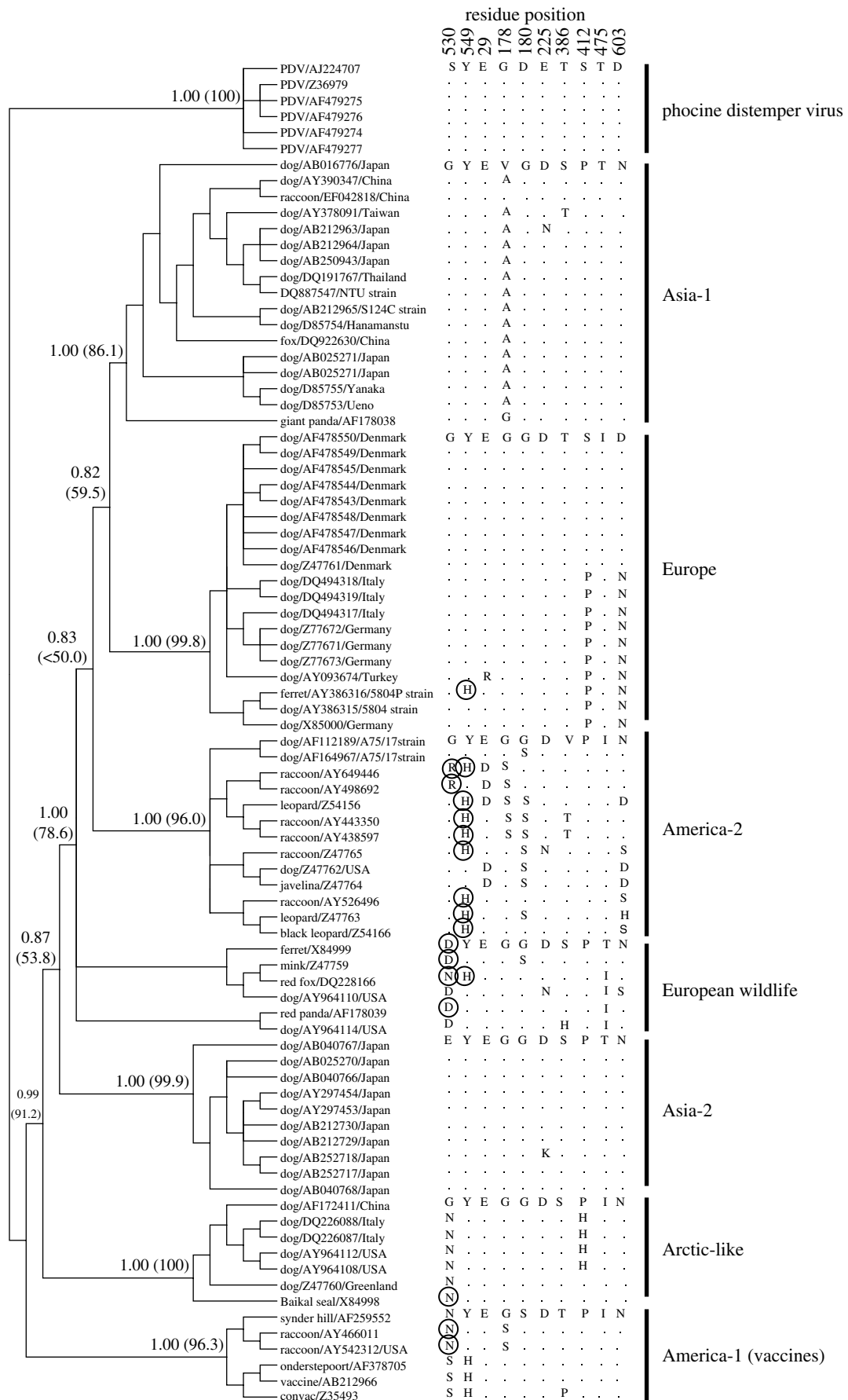


Figure 1. Phylogenetic relationship between CDV strains based on H gene sequences and the distribution of H protein residues at sites under positive selection across lineages. PDV-1 strains were used as out-group and only the tree topology is shown for clarity. Bayesian posterior probabilities and maximum likelihood bootstrap support values (in parentheses) for the major lineages are given at the nodes. Host species, accession number and isolate location are also shown. CDV strains form seven distinct clusters: America-1 (vaccines), America-2, Arctic-Like, Asia-1, Asia-2, Europe and European wildlife. Residues 530 and 549 are located in the regions of SLAM receptor-binding importance. Circled residues are substitutions at residues 530 and 549 of isolates from non-dog hosts.

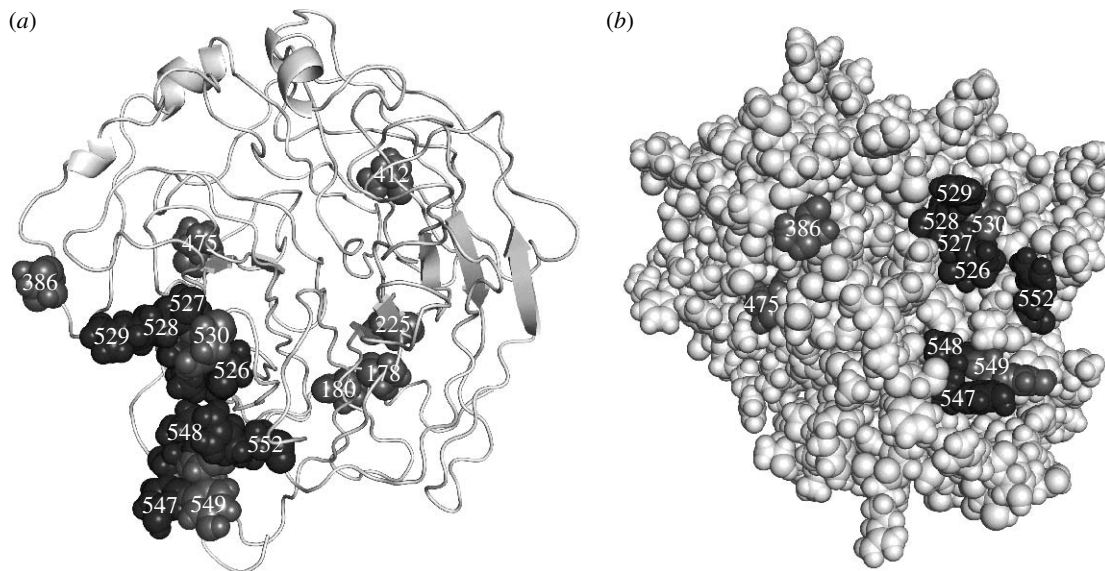


Figure 2. CDV H protein predicted tertiary structure (von Messling *et al.* 2005). Residues 527, 528, 529 and 552 (required for SLAM-dependent fusion and conserved between all morbilliviruses), as well as residues 526, 547 and 548 (also important in SLAM-dependent fusion), are shown in black. Residues under the influence of positive selection are shown in grey. (a) Top view. (b) Side view. The illustrations of the models were prepared with Pymol (DeLano 2002).

Twelve non-dog hosts do not have an E/G530 substitution. However, 7 of these 12 isolates (host species: black leopard and leopard *Panthera pardus*, ferret and raccoon) have substitution Y549H, which was identified to be under the influence of positive selection and maps to a receptor-binding domain (figure 1). Elsewhere, the Y549H substitution is only observed in three vaccine sequences. In addition, a raccoon isolate and the red fox isolate have substitutions at both 530 and 549 sites (figure 1). Substitutions at residue positions 29, 178 and 603 could also be associated with host switches, though the strength of evidence of their involvement is weaker than residues 530 and 549 (figure 1). E29D is observed in four non-dog isolates and one dog host isolate. 178S is only observed in non-dog isolates. Finally, H/S603 substitutions are observed in one dog isolate, two raccoon isolates, one black leopard and one leopard isolate. Changes at residues 180, 225, 386, 412 and 475 do not show any association to host specificity.

4. DISCUSSION

We assessed the role of evolution in driving host switches and the emergence of CDV, one of the most important carnivore pathogens, in non-dog host species. CDV is an ideal candidate for studying the role of pathogen evolution in host switches, not only because it infects a diverse host range and is an important threat to wild carnivores of conservation concern (Mamaev *et al.* 1995; Harder *et al.* 1996; Kennedy *et al.* 2000; Cleaveland *et al.* 2006), but also because its molecular and cell biology are well characterized with tertiary structures predicted for key virus molecules. This allows a hypothesis-testing approach to be used where evidence for evolution at specific virus residues with a known function can be assessed. Using phylogenetic and molecular evolution approaches, we tested the specific hypothesis that molecular adaptation at residues 530 and 548 of the CDV haemagglutinin protein, which are involved in receptor tropism, is associated with host switches and the spread of CDV to non-dog hosts. We found that site 530

is indeed under strong positive selection, along with site 549, another residue in the SLAM receptor-binding region, but not site 548. The great majority of amino acid substitutions at these sites map to CDV isolates derived from non-dog hosts, indicating that adaptation at these receptor-binding sites is associated with the spread of CDV from dogs to non-dog hosts. The study provides compelling evidence for pathogen evolution driving multiple independent emergences of disease in novel hosts, and suggests that pathogen evolution is likely to be an important general driver of the spread of diseases to novel host ranges alongside other human mediated drivers.

Our analysis reveals signals of both selection and recombination in different genes for the global CDV population. Analysis of the CDV *H* gene reveals both that the overall ω ratio is significantly greater than 1 under the M8 model ($p=0.0025$), and that a small number of specific residues are important in the adaptive evolution of CDV through positive selection, owing to their interactions with cellular receptors or with the host immune system.

Experimental evidence from *in vitro* receptor-binding studies shows that residues 527, 528, 529 and 552, conserved between all morbilliviruses, are crucial for CDV SLAM-dependent fusion (von Messling *et al.* 2005). Residues 526, 547 and 548 are also important in efficient SLAM-dependent fusion. These clusters are located in β -sheet 5 of the predicted H protein tertiary structure, and are predicted to be accessible for receptor interactions (figure 2; von Messling *et al.* 2005). It has previously been determined that mutation of site 530, in addition to substitution at site 548, of the H protein was critical in the adaptation of CDV from vero cells to marmoset B cells in culture, signifying adaptation from CD46-dependent fusion to SLAM-dependent fusion (Seki *et al.* 2003).

Residue 530 ($\text{Pr } \omega > 1 = 0.969$) was identified by our analysis as being under the influence of positive selection. No signal of positive selection was detected at site 548, suggesting that this residue is not an important determinant of tropism in the wild; however, positive selection was

detected at site 549. Residues 530 and 549 both fall into receptor-binding domains located on propeller β -sheet 5 of the haemagglutinin molecule (figure 2). This evidence indicates that molecular adaptation of residues 530 and 549 may change the affinity of interaction between CDV H protein and SLAM in different host species, and supports our hypothesis that molecular adaptation in receptor-binding domains is associated with independent incidents of spread of CDV to non-dog hosts. MV H residues 534 and 553 (corresponding to CDV sites 530 and 549) are not under the influence of positive selection (Woelk *et al.* 2002). In addition to receptor interactions, B-cell epitopes (BCEs) and T-cell epitopes (TCEs) are also reported to localize to the CDV H protein, and the antigenic region of the H protein has been reported to have changed in recent CDV field isolates (Orvell *et al.* 1985; Iwatsuki *et al.* 2000). However, conclusions regarding CDV H residues under the influence of immune selection cannot be drawn at this time as predictive methodologies currently only describe human and mouse epitopes.

We determined the phylogenetic distribution of substitutions at all sites identified to be under the influence of positive selection (figure 1). Strikingly, of the 21 non-dog host isolates 9 have an E/G530D/N/R substitution, 8 have an Y549H substitution and 2 have substitutions at both residue positions (figure 1). In addition, partial *H* gene sequences from CDV isolates from the year 2000 epidemic in Caspian seals contain both E/G530N and Y549H substitutions (data not shown—Barrett & Banyard 2007, unpublished data). Phylogenetic analysis (not shown) indicates that these sequences form a single cluster that diverges after the America-1 clade, but basal to the Arctic-like clade. As these were not full-length sequences, we did not include them in the full analysis.

Substitutions at residues 530 and 549 are associated with CDV isolated from novel host species, and in the light of our analysis, such substitutions can also be observed directly in experimental infections. In a previous study by von Messling *et al.* (2003), CDV strain 5804 was passaged experimentally through ferrets yielding the 5804P ferret-adapted strain (AY386316) which differed from the parental 5804 dog strain (AY396315) at residues 106 and 549 of the haemagglutinin protein. This provides strong experimental evidence for a functional role of residue 549 in host switches. Further analysis of the role of residues 530 and 549 by manipulation of viral sequences and experimental infections would provide more insight into the mechanisms which determine host tropism for this virus. In general analyses which combine phylogenetic, molecular evolution and experimental approaches are likely to be productive routes for investigating the evolutionary basis of host switches in many viral pathogens.

Statistically significant evidence for positive selection was not detected in analyses of the *F*, *L*, *M*, *N* and *P* genes. These analyses suggest that a proportion of sites in these genes are conserved, and that a small proportion of sites are neutrally evolving. Notably, analysis of the *L* and *M* genes suggest that a high proportion of sites are conserved, and likely to be subject to purifying selection, as previously identified in measles virus *L* and *M* genes (Woelk *et al.* 2002). The high sequence conservation identified in the *L* and *M* proteins could be due to crucial roles in viral transcription and replication, and lytic formation, respectively (Lamb & Kolakofsky 2001). The lack of selection

signal from the CDV *F* gene is surprising when previous studies identified immune selection in the MV *F* gene and mapped selected residues to BCEs and TCEs, but could be due to the small number of full-length gene sequences currently available for analysis (Muller *et al.* 1993, 1996; van Binnendijk *et al.* 1993; Woelk *et al.* 2002). Recombination could also be a confounding factor but this is more likely to generate a false signal of selection (Anisimova *et al.* 2003). Our analyses suggest that the CDV *N* and *P* genes are not under the influence of positive selection, despite their known interactions with the host immune response (Tipold *et al.* 1999; Yoshida *et al.* 1999; Gotoh *et al.* 2001; Palosaari *et al.* 2003; Shaffer *et al.* 2003; Kerdiles *et al.* 2006). The availability of CDV gene sequences isolated from a wider range of host species would increase the power of analysis and allow a more detailed assessment of the selection pressures acting on the CDV *F*, *N* and *P* genes. The present work assessed all publicly available full-length *H* gene sequences at the time of writing. However, there are CDV isolates from epizootics in non-dog hosts for which no sequence covering the SLAM-binding region, or the full length of the *F* gene is available, which would be very informative to assess in the future. Of special interest would be strains from Serengeti lions and Caspian seals (Roelke-Parker *et al.* 1996; Carpenter *et al.* 1998; Kennedy *et al.* 2000).

Previous studies have revealed the importance of recombination in shaping RNA virus diversity and identified that recombination acts at lower levels in single-stranded negative-sense RNA viruses compared with other RNA viruses (Worobey & Holmes 1999; McVean *et al.* 2002; Chare *et al.* 2003; Schierup *et al.* 2005). In our study, although there is evidence of recombination in the *F* gene, in which two alternative permutation tests yielded recombination estimates that were significantly different from 0 and two pairs of recombinant sequences were identified in dog isolates, results for other genes should be treated with caution. For the *L* and *N* genes, the marginally significant results may be type I errors. Recombination was not detected in an informative site test of CDV *F* gene sequences in a previous study (Chare *et al.* 2003). Theoretically, variation in the *F* gene provided by recombination could change interactions between host tissue-specific proteases and the *F* protein, thus influencing tropism (Lamb & Kolakofsky 2001). However, the recombinant segments identified by the GENECONV analysis did not map to regions with a known role in host specificity. Alternatively, recombination could be driven by immune selection, but again without knowledge of epitopes generated from this gene this must remain as speculation at this time. Our *H* gene ρ estimate (6.00) is slightly lower than previous estimates in the MV *H* gene (7.2–15.8; Chare *et al.* 2003; Schierup *et al.* 2005). The underlying reason behind the apparent low levels of recombination in paramyxoviruses is not currently understood. The paramyxovirus template for replication abides to the ‘the rule of six’ which dictates that each nucleocapsid monomer is tightly associated with six nucleotides by hydrophobic bonds. Though this prevents the presence of naked viral RNA, it also raises the question of how template recognition occurs in order to allow recombination to take place, possibly offering an explanation for the apparent low levels of recombination in negative sense RNA viruses (Egelman *et al.* 1989). Our

analysis detected higher levels of recombination in glycoprotein genes; this could increase variation in proteins that are determinants of cellular tropism and subjected to neutralizing antibodies, an obvious benefit to a virus population (Orvell *et al.* 1985; Lamb & Kolakofsky 2001; Hiramama *et al.* 2002).

Few studies have detected substitutions at specific functional sites in pathogen proteins mediating the molecular adaptation of a pathogen to different host species. In order for avian influenza A to jump hosts to humans, it must adapt from using the avian form (SA α 2,3Gal) to the human form (SA α 2,6Gal) of the sialic acid (SA) receptor. Variation at HA residues 226 and 228 in H2 and H3 serotypes have long been known to be associated with host receptor specificity (Connor *et al.* 1994). More recently, substitutions N182L and Q192R of H5N1 influenza HA molecule were identified to convert HAs to recognize SA α 2,6Gal instead of SA α 2,3Gal (Yamada *et al.* 2006). Though mutation at these sites may not alone be sufficient to cause a full pandemic, they may be selected in the early phases of human infection and serve as good markers for the assessment of the ability of avian field H5N1 isolates to replicate in humans. The potential of H5N1 to adapt to humans and emerge as a pandemic strain can be gauged by surveillance of sites 182 and 192 in populations of avian field isolates (Yamada *et al.* 2006). In the case of CDV, if experimental evidence confirmed the influential role of sites 530 and 549 of CDV H protein in the molecular adaptation of CDV into novel hosts, then surveillance of these sites would be an informative technique to assess the potential of CDV to spread to novel wildlife populations. This could prove to be a valuable tool for combating an infectious disease that has already been associated with declines in wild carnivore populations, and the impact of which is likely to increase in the future.

Our analysis shows that pathogen evolution plays a crucial role in the establishment of CDV in novel carnivore host species. Our findings constitute one of only a small number of cases where repeated independent emergence of a pathogen in novel hosts is associated with evolution of the pathogen at key functional residues that determine host specificity. Probably the best other example comes from lentiviruses (which are also RNA viruses and include the simian (SIV) and human (HIV) immunodeficiency viruses) in African primates. Here, multiple species jumps by different SIV viruses have been documented, with evidence for infection in at least 26 species, associated with polymorphisms in the V2 and V3 loop regions of the *env* gene, which are involved in interactions with host receptors and determine cell tropism (Hahn *et al.* 2000). Adaptive variation has also been implicated in species jumps for feline lentiviruses (Poss *et al.* 2006). Together with other examples such as avian influenza and canine parvovirus, these present a growing body of evidence for the general importance of pathogen evolution in disease emergence, and in the case of host switches by viruses, that subtle adaptations in receptor-binding regions which determine host cell tropism may be an important step for establishment in new hosts. Of course other functional adaptation may also be important for, or arise after establishment in new hosts (Webby *et al.* 2004), as evidenced by the substitutions which modulate virulence in the VEEV and WNV examples. The relative rarity of these examples so far likely reflects lack of sequence data

for viral isolates from multiple hosts species and detailed information about the cell biology and structure of viral proteins which preclude the hypothesis-testing approach taken here and in the other highlighted studies. The future challenges are to increase our understanding of how these evolutionary processes interact with new selection pressures on pathogens arising from human drivers such as globalization, agricultural intensification and impingement into new habitats, to promote disease emergence. Unifying identification of functional adaptation with molecular estimates of the time of disease emergence events and linking these to known changes in human drivers will likely offer a productive way forward.

CDV belongs to the genus *Morbillivirus*, other members of which have also caused mass mortalities among wildlife and livestock. The introduction of rinderpest virus to East Africa in the 1880s caused mortality levels of up to 90% among wild bovid populations such as buffalo *Syncerus caffer*, lesser kudu *Tragelaphus imberbis* and eland *Taurotragus oryx* (Plowright 1982; Kock *et al.* 1999), and PDV has emerged in harbour seals, *Phoca vitulina* and grey seals, *Halichoerus grypus* of Europe causing mass mortalities (Harkonen *et al.* 2006). Assessing the role of orthologous H gene residues of 530 and 549, in other morbilliviruses, in molecular adaptation to novel hosts would indicate if the same evolutionary mechanisms operate in the host switches of members of the genus *Morbillivirus*. These sites and similar sites in other RNA viruses involved in receptor binding may be potential targets for antiviral therapies. This research shows how combining molecular evolutionary and phylogenetic analysis with findings on the functional and structural properties of pathogen proteins can yield important insights into the process of disease emergence, and could prove crucial in accurately predicting the impact of and planning response to disease emergence.

We thank Roberto Cattaneo for providing the predicted tertiary structure of the CDV H protein, and Tom Barrett and Ashley Banyard of the Institute for Animal Health, Pirbright, UK, for giving us access to their unpublished sequence for the H gene of the Caspian seal CDV isolate. Two anonymous referees provided valuable comments which helped improve the original manuscript. A.J.M. was supported by a BBSRC PhD studentship.

REFERENCES

- Alexander, K. & Appel, M. 1994 African wild dogs (*Lycaon pictus*) endangered by a canine distemper epizootic among domestic dogs near the Masai Mara National Reserve, Kenya. *J. Wildl. Dis.* **30**, 481–485.
- Altizer, S., Harvell, D. & Friedle, E. 2003 Rapid evolutionary dynamics and disease threats to biodiversity. *Trends Ecol. Evol.* **18**, 589–596. (doi:10.1016/j.tree.2003.08.013)
- Anishchenko, M., Bowen, R., Paessler, S., Austgen, L., Greene, I. & Weaver, S. 2006 Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proc. Natl Acad. Sci. USA* **103**, 4994–4999. (doi:10.1073/pnas.0509961103)
- Anisimova, M., Nielsen, R. & Yang, Z. 2003 Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**, 1229–1236.
- Awadalla, P., Eyre-Walker, A. & Smith, J. M. 1999 Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**, 2524–2525. (doi:10.1126/science.286.5449.2524)

- Brault, A. C. *et al.* 2007 A single positively selected West Nile viral mutation confers increased virulence in American crows. *Nat. Genet.* **39**, 1162–1166. (doi:10.1038/ng2097)
- Carpenter, M., Appel, M., Roelke-Parker, M., Munson, L., Hofer, H., East, M. & O'Brien, S. 1998 Genetic characterization of canine distemper virus in Serengeti carnivores. *Vet. Immunol. Immunopathol.* **65**, 259–266. (doi:10.1016/S0165-2427(98)00159-7)
- Chang, S. F., Sgro, J. Y. & Parrish, C. R. 1992 Multiple amino acids in the capsid structure of canine parvovirus coordinately determine the canine host range and specific antigenic and hemagglutination properties. *J. Virol.* **66**, 6858–6867.
- Chare, E. R., Gould, E. A. & Holmes, E. C. 2003 Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen. Virol.* **84**, 2691–2703. (doi:10.1099/vir.0.19277-0)
- Cleaveland, S., Meslin, F. X. & Breiman, R. 2006 Dogs can play useful role as sentinel hosts for disease. *Nature* **440**, 605. (doi:10.1038/440605b)
- Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. 1994 Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17–23. (doi:10.1006/viro.1994.1615)
- Daszak, P., Berger, L., Cunningham, A. A., Hyatt, A. D., Green, D. E. & Speare, R. 1999 Emerging infectious diseases and amphibian population declines. *Emerg. Infect. Dis.* **5**, 735–748.
- Daszak, P., Cunningham, A. A. & Hyatt, A. D. 2000 Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* **287**, 443–449. (doi:10.1126/science.287.5452.443)
- DeLano, W.L. 2002 *The PyMol molecular graphic system*. San Carlos, CA: DeLano Scientific. See <http://www.pymol.org>.
- Egelman, E. H., Wu, S. S., Amrein, M., Portner, A. & Murti, G. 1989 The Sendai virus nucleocapsid exists in at least four different helical states. *J. Virol.* **63**, 2233–2243.
- Gotoh, B., Komatsu, T., Takeuchi, K. & Yokoo, J. 2001 Paramyxovirus accessory proteins as interferon antagonists. *Microbiol. Immunol.* **45**, 787–800.
- Guiler, E. R. 1961 The former distribution and decline of the thylacine. *Aust. J. Sci.* **23**, 207–210.
- Haas, L., Hoffer, H., East, M., Wohlsein, P., Liess, B. & Barrett, T. 1996 Canine distemper virus infection in Serengeti spotted hyaenas. *Vet. Microbiol.* **49**, 147–152. (doi:10.1016/0378-1135(95)00180-8)
- Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. 2000 AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614. (doi:10.1126/science.287.5453.607)
- Harkonen, T. *et al.* 2006 The 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Dis. Aquat. Organ.* **68**, 115–130. (doi:10.3354/dao068115)
- Harder, T. C. *et al.* 1996 Canine distemper virus from diseased large felids: biological properties and phylogenetic relationships. *J. Gen. Virol.* **77**, 397–405.
- Hashimoto, M., Une, Y. & Mochizuki, M. 2001 Hemagglutinin genotype profiles of canine distemper virus from domestic dogs in Japan. *Arch. Virol.* **146**, 149155. (doi:10.1007/s007050170198)
- Haydon, D., Bastos, A., Knowles, N. & Samuel, A. 2001 Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* **157**, 7–15.
- Hirama, K., Togashi, K., Wakasa, C., Yoneda, M., Nishi, T., Endo, Y., Miura, R., Tsukiyama-Kohara, K. & Kai, C. 2002 Cytotoxic T-lymphocyte activity specific for hemagglutinin (H) protein of canine distemper virus in dogs. *J. Vet. Med. Sci.* **605**, 109–112.
- Holmes, E. C., Woelk, C. H., Kassis, R. & Bourhy, H. 2002 Genetic constraints and the adaptive evolution of rabies virus in nature. *Virology* **292**, 247–257. (doi:10.1006/viro.2001.1271)
- Hueffer, K., Parker, J. S. L., Weichert, W. S., Geisel, R. E., Sgro, J. Y. & Parrish, C. R. 2003 The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. *J. Virol.* **77**, 1718–1726. (doi:10.1128/JVI.77.3.1718-1726.2003)
- Hueffer, K. & Parrish, C. R. 2003 Parvovirus host range, cell tropism and evolution. *Curr. Opin. Microbiol.* **6**, 392–398. (doi:10.1016/S1369-5274(03)00083-3)
- Hulsenbeck, J. & Ronquist, F. 2001 MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755. (doi:10.1093/bioinformatics/17.8.754)
- Iwatsuki, K., Tokiyoshi, S., Hirayama, N., Nakamura, K., Ohashi, K., Wakasa, C., Mikami, T. & Kai, C. 2000 Antigenic differences in the H proteins of canine distemper viruses. *Vet. Microbiol.* **71**, 281–286. (doi:10.1016/S0378-1135(99)00172-8)
- Jobb, G., von Haeseler, A. & Strimmer, K. 2004 TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* **4**, 18. (doi:10.1186/1471-2148-4-18)
- Keane, T., Creevey, C., Pentony, M., Naughton, N. & McInerney, J. 2006 Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29. (doi:10.1186/1471-2148-6-29)
- Kennedy, S. *et al.* 2000 Mass die-Off of Caspian seals caused by canine distemper virus. *Emerg. Infect. Dis.* **6**, 637–639.
- Kerdiles, Y. M. 2006 Immunomodulatory properties of morbillivirus nucleoproteins. *Viral Immunol.* **19**, 324–334. (doi:10.1089/vim.2006.19.324)
- Kock, R. A., Wambua, J. M., Mwanzia, J., Wamwayi, H., Ndungu, E. K., Barrett, T., Kock, N. D. & Rossiter, P. B. 1999 Rinderpest epidemic in wild ruminants in Kenya 1993–97. *Vet. Rec.* **145**, 275–283.
- Lamb, R. & Kolakofsky, D. 2001 Paramyxoviridae: the viruses and their replication. In *Fields Virology*, vol. 1 (eds D. M. Knipe & P. M. Howley).
- Lemberger, K. Y., Gondim, L. F., Pessier, A. P., McAllister, M. M. & Kinsel, M. J. 2005 *Neospora caninum* infection in a free-ranging raccoon (*Procyon lotor*) with concurrent canine distemper virus infection. *J. Parasitol.* **91**, 960–961. (doi:10.1645/GE-407R.1)
- Machida, N., Izumisawa, N., Nakamura, T. & Kiryu, K. 1992 Canine distemper virus infection in a masked palm civet (*Paguma larvata*). *J. Comp. Pathol.* **107**, 439–443. (doi:10.1016/0021-9975(92)90017-0)
- Mamaev, L. *et al.* 1995 Characterisation of morbilliviruses isolated from Lake Baikal seals (*Phoca sibirica*). *Vet. Microbiol.* **44**, 251–259. (doi:10.1016/0378-1135(95)00018-6)
- Martella, V. *et al.* 2006 Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. *Vet. Microbiol.* **116**, 301–309. (doi:10.1016/j.vetmic.2006.04.019)
- Martella, V. *et al.* 2007 Genotyping canine distemper virus (CDV) by a hemi-nested multiplex PCR provides a rapid approach for investigation of CDV outbreaks. *Vet. Microbiol.* **122**, 32–42. (doi:10.1016/j.vetmic.2007.01.005)
- McVean, G., Awadalla, P. & Fearnhead, P. 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241.
- Monto, A. S. 2005 The threat of an avian influenza pandemic. *New Engl. J. Med.* **352**, 323–325. (doi:10.1056/NEJMp048343)
- Muller, C. P., Ammerlaan, W., Fleckenstein, B., Krauss, S., Kalbacher, H., Schneider, F., Jung, G. & Wiesmuller, K. H.

- 1996 Activation of T cells by the ragged tail of MHC class II-presented peptides of the measles virus fusion protein. *Int. Immunol.* **8**, 445–456. (doi:10.1093/intimm/8.4.445)
- Muller, C. P., Schroeder, T., Tu, R., Brons, N. H., Jung, G., Schneider, F. & Wiesmuller, K. H. 1993 Analysis of the neutralizing antibody response to the measles virus using synthetic peptides of the haemagglutinin protein. *Scand. J. Immunol.* **38**, 463–471. (doi:10.1111/j.1365-3083.1993.tb02589.x)
- Nash, D. *et al.* 2001 The outbreak of West Nile virus infection in the New York City area in 1999. *New Engl. J. Med.* **344**, 1807–1814. (doi:10.1056/NEJM200106143442401)
- Orvell, C., Sheshberadaran, H. & Norrby, E. 1985 Preparation and characterization of monoclonal antibodies directed against four structural components of canine distemper virus. *J. Gen. Virol.* **66**(Pt 3), 443–456.
- Palosaari, H., Parisien, J. P., Rodriguez, J. J., Ulane, C. M. & Horvath, C. M. 2003 STAT protein interference and suppression of cytokine signal transduction by measles virus V protein. *J. Virol.* **77**, 7635–7644. (doi:10.1128/JVI.77.13.7635-7644.2003)
- Plowright, W. 1982 The effects of rinderpest and rinderpest control on wildlife in Africa. *Symp. Zool. Soc. Lond.* **50**, 1–28.
- Poss, M., Ross, H. A., Painter, S. L., Holley, D. C., Terwee, J. A., VandeWoude, S. & Rodrigo, A. 2006 Feline lentivirus evolution in cross-species infection reveals extensive G-to-A mutation and selection on key residues in the viral polymerase. *J. Virol.* **80**, 2728–2737. (doi:10.1128/JVI.80.6.2728-2737.2006)
- Richman, A. D., Herrera, L. G., Nash, D. & Schierup, M. H. 2003 Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in *Peromyscus maniculatus*. *Genet. Res.* **82**, 89–99. (doi:10.1017/S0016672303006347)
- Roelke-Parker, M. *et al.* 1996 A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* **379**, 441–445. (doi:10.1038/379441a0)
- Sawyer, S. A. 1989 Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526–538.
- Sawyer, S. A. 1999 GENECONV: a computer package for statistical detection of gene conversion. See <http://www.math.wustl.edu/~sawyer/mbprogs/>.
- Schierup, M. H., Mordhorst, C. H., Muller, C. P. & Christensen, L. S. 2005 Evidence of recombination among early-vaccination era measles virus strains. *BMC Evol. Biol.* **5**, 52. (doi:10.1186/1471-2148-5-52)
- Seki, F., Ono, N., Yamaguchi, R. & Yanagi, Y. 2003 Efficient isolation of wild strains of canine distemper virus in vero cells expressing canine SLAM (CD150) and their adaptability to marmoset B95a cells. *J. Virol.* **77**, 9943–9950. (doi:10.1128/JVI.77.18.9943-9950.2003)
- Shaffer, J. A., Bellini, W. J. & Rota, P. A. 2003 The C protein of measles virus inhibits the type I interferon response. *Virology* **315**, 389–397. (doi:10.1016/S0042-6822(03)00537-3)
- Smith, N. G. & Fearnhead, P. 2005 A comparison of three estimators of the population-scaled recombination rate: accuracy and robustness. *Genetics* **171**, 2051–2062. (doi:10.1534/genetics.104.036293)
- Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D. & Rappuoli, R. 2003 SARS—beginning to understand a new virus. *Nat. Rev. Microbiol.* **1**, 209–218. (doi:10.1038/nrmicro775)
- Tatsuo, H., Ono, N. & Yanagi, Y. 2001 Morbilliviruses use signaling lymphocyte activation molecules (CD150) as cellular receptors. *J. Virol.* **75**, 5842–5850. (doi:10.1128/JVI.75.13.5842-5850.2001)
- Tipold, A., Moore, P., Zurbriggen, A., Burgener, I., Barben, G. & Vandevelde, M. 1999 Early T cell response in the central nervous system in canine distemper virus infection. *Acta Neuropathol. (Berl.)* **97**, 45–56. (doi:10.1007/s004010050954)
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997 The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876–4882. (doi:10.1093/nar/25.24.4876)
- van Binnendijk, R. S., Versteeg-van Oosten, J. P., Poelen, M. C., Brugghe, H. F., Hoogerhout, P., Osterhaus, A. D. & Uytendaele, F. G. 1993 Human HLA class I- and HLA class II-restricted cloned cytotoxic T lymphocytes identify a cluster of epitopes on the measles virus fusion protein. *J. Virol.* **67**, 2276–2284.
- von Messling, V., Springfield, C., Devaux, P. & Cattaneo, R. 2003 A ferret model of canine distemper virus virulence and immunosuppression. *J. Virol.* **77**, 12 579–12 591. (doi:10.1128/JVI.77.23.12579-12591.2003)
- von Messling, V., Oezguen, N., Zheng, Q., Vongpunsawad, S., Braun, W. & Cattaneo, R. 2005 Nearby clusters of hemagglutinin residues sustain SLAM-dependent canine distemper virus entry in peripheral blood mononuclear cells. *J. Virol.* **79**, 5857–5862. (doi:10.1128/JVI.79.9.5857-5862.2005)
- Vongpunsawad, S., Oezgun, N., Braun, W. & Cattaneo, R. 2004 Selectively receptor-blind measles viruses: identification of residues necessary for SLAM- or CD46-induced fusion and their localization on a new hemagglutinin structural model. *J. Virol.* **78**, 302–313. (doi:10.1128/JVI.78.1.302-313.2004)
- Webby, R., Hoffmann, E. & Webster, R. 2004 Molecular constraints to interspecies transmission of viral pathogens. *Nat. Med.* **10**, S77–S81. (doi:10.1038/nm1151)
- Williams, E. S., Thorne, E. T., Appel, M. J. & Belitsky, D. W. 1988 Canine distemper in black-footed ferrets (*Mustela nigripes*) from Wyoming. *J. Wildl. Dis.* **24**, 385–398.
- Woelk, C. H., Jin, L., Holmes, E. C. & Brown, D. W. 2001 Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *J. Gen. Virol.* **82**, 2463–2474.
- Woelk, C. H., Pybus, O. G., Jin, L., Brown, D. W. & Holmes, E. C. 2002 Increased positive selection pressure in persistent (SSPE) versus acute measles virus infections. *J. Gen. Virol.* **83**, 1419–1430.
- Woolhouse, M. & Gowtage-Sequeria, S. 2005 Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842–1847.
- Worobey, M. & Holmes, E. C. 1999 Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80**, 2535–2543.
- Yamada, S. *et al.* 2006 Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382. (doi:10.1038/nature05264)
- Yang, Z. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.
- Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.
- Yoshida, E., Shin, Y. S., Iwatsuki, K., Gemma, T., Miyashita, N., Tomonaga, K., Hirayama, N., Mikami, T. & Kai, C. 1999 Epitopes and nuclear localization analyses of canine distemper virus nucleocapsid protein by expression of its deletion mutants. *Vet. Microbiol.* **66**, 313–320. (doi:10.1016/S0378-1135(99)00022-X)