Tracking Viral Evolution During a Disease Outbreak: The Rapid and Complete Selective Sweep of a Circovirus in the Endangered Echo parakeet

JVI
Journals.ASM.org

Tracking Viral Evolution during a Disease Outbreak: the Rapid and Complete Selective Sweep of a Circovirus in the Endangered Echo Parakeet

Samit Kundu,[a] Christopher G. Faulkes,[b] Andrew G. Greenwood,[c] Carl G. Jones,[d] Pete Kaiser,[e] Owen D. Lyne,[f] Simon A. Black,[a] Aurelie Chowrimootoo,[d] and Jim J. Groombridge[a]

Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, United Kingdom[a]; School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom[b]; International Zoo Veterinary Group, Station House, Keighley, West Yorkshire, United Kingdom[c]; Durrell Wildlife Conservation Trust, Les Augrés Manor, Trinity, Jersey, Channel Islands, United Kingdom, and Mauritian Wildlife Foundation, Vacoas, Mauritius[d]; The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom[e]; and School of Mathematics, Statistics, and Actuarial Science, University of Kent, Canterbury, United Kingdom[f]

Circoviruses are among the smallest and simplest of all viruses, but they are relatively poorly characterized. Here, we intensively sampled two sympatric parrot populations from Mauritius over a period of 11 years and screened for the circovirus *Beak and feather disease virus* (BFDV). During the sampling period, a severe outbreak of psittacine beak and feather disease, which is caused by BFDV, occurred in Echo parakeets. Consequently, this data set presents an ideal system for studying the evolution of a pathogen in a natural population and to understand the adaptive changes that cause outbreaks. Unexpectedly, we discovered that the outbreak was most likely caused by changes in functionally important regions of the normally conserved *replication-associated protein* gene and not the immunogenic *capsid*. Moreover, these mutations were completely fixed in the Echo parakeet host population very shortly after the outbreak. Several *capsid* alleles were linked to the *replication-associated protein* outbreak allele, suggesting that whereas the key changes occurred in the latter, the scope of the outbreak and the selective sweep may have been influenced by positive selection in the capsid. We found evidence for viral transmission between the two host populations though evidence for the invasive species as the source of the outbreak was equivocal. Finally, the high evolutionary rate that we estimated shows how rapidly new variation can arise in BFDV and is consistent with recent results from other small single-stranded DNA viruses.

Managing the effects of virally mediated infectious disease in host populations requires an understanding of how these pathogens evolve. Rapid evolution and selection can enable a virus to escape immune recognition and to reinfect individuals that have been previously exposed. However, logistical difficulties in surveying wild populations may undermine our ability to understand evolution in pathogens that do not infect humans or domestic livestock. Endangered species, which are often monitored and regularly sampled over long periods of time, offer a resolution. Using a data set of samples collected over 11 years from two populations of parrot, the endangered Echo parakeet (*Psittacula echo*) and the rose-ringed parakeet (*Psittacula krameri*), and from both symptomatic and asymptomatic birds, we tracked the evolution of *Beak and feather disease virus* (BFDV) before, during, and after a significant disease outbreak.

BFDV (family *Circoviridae*) is one of the most common infections of parrots (57). It is a small and simple virus with a circular, ambisense ssDNA genome of around 2,000 nucleotides encoding two products of known function (52). Open reading frame 1 (ORF1) encodes the replication-associated protein Rep, which is responsible for viral replication (16) and has been shown to be highly conserved (23, 30, 34). ORF2 encodes a capsid protein responsible for encapsidation of the virus and its entry into the cell via cell surface receptors (37, 38, 52). Another protein and other ORFs have been putatively identified, but they are not consistently present (3, 4, 9, 51, 52). The virus has been shown to be environmentally stable (61) and may be transmitted both horizontally and vertically (18, 48, 63). Infection by BFDV leads to psittacine beak and feather disease (PBFD), which is characterized by immunosuppression, feather dystrophy, and, in some species, beak deformity (33, 41, 51, 52).

The Echo parakeet is an endangered parrot species endemic to the Indian Ocean island of Mauritius. By the early 1980s, this species had declined to such a degree (around 20 individuals) that it was regarded as the world's rarest parrot (28). Intervention by a conservation program has resulted in a period of population growth over the last decade, with the number of birds currently estimated at around 500. Between 2005 and 2006 the recovery was interrupted by the presence of a pathogen; a large proportion of the population showed clinical signs consistent with PBFD. One suspected source of infection in the Echo parakeets is a population of feral rose-ringed parakeets. Since being founded in about 1886 (6), this population has grown rapidly to number over 10,000 birds and is widespread throughout Mauritius. In the context of infectious disease, invasive species represent an especially pernicious problem. Besides the potential for introducing novel pathogens and pathogen strains for which the endemic species may have no immunity, in such large and growing host populations natural

Received 6 October 2011 **Accepted** 7 February 2012
Published ahead of print 15 February 2012
Address correspondence to Samit Kundu, samit.kundu@ucl.ac.uk.

doi:10.1128/JVI.06504-11

selection will tend to favor pathogen virulence (5). The presence of such a large pathogen reservoir can also result in a persistence of disease outbreaks in small endemic host populations through a pattern of constant reinfection.

Since 1993, on an almost annual basis, a conservation program has collected blood samples from the Echo parakeet population. By chance, the sampling periods have encompassed the PBFD outbreak, and consequently this archive presents a novel opportunity to monitor the evolution of a pathogen in a natural host population before, during, and after an outbreak. By reconstructing the phylogenetic history of BFDV in the Echo parakeet population, we aimed to identify the mutation(s) that may have led to the outbreak. Since the circovirus capsid is immunogenic (17), we expected the outbreak to be defined by mutations in the *capsid* gene and for codons in this gene to show evidence for positive selection (as opposed to the *rep*, which should be conserved due to its critical role in replication). We reconstruct the BFDV phylogeny using isolates from both parrot populations to identify any evidence for transmission and in particular for any indication that the outbreak was caused by the invasive species. Finally, we take advantage of our extensive time-stamped archive of samples to estimate the evolutionary rate of BFDV. Recent evidence from studies of other single-stranded DNA (ssDNA) viruses suggests that these pathogens are characterized by high rates of evolution, on a par with single-stranded RNA (ssRNA) viruses (14, 55, 56).

MATERIALS AND METHODS

Sampling and DNA extraction. Blood samples were taken during each breeding season (September to May) from Echo parakeets and feral rose-ringed parakeets between 1993 and 1998 and between 2003 and 2009 (there was no sampling between 1999 and 2002 due to no obvious presence of disease). The field teams involved in sampling the birds followed strict anticontamination protocols, including soaking all field equipment in Virkon for 48 h between visits to different host nest sites. Since 2004 this sampling was carried out regardless of apparent disease status and represented between 22 and 41% of the known population of Echo parakeets each season. In total, 592 Echo parakeets and 125 feral rose-ringed parakeets were sampled. DNA was extracted from 50 to 100 μl of whole blood using a standard phenol-chloroform extraction protocol (54). An extraction blank was included to check that there was no contamination.

PCR and sequencing. We screened for BFDV using a PCR assay targeting a 717-bp region of ORF1 (*rep* gene) (65). The reaction volume comprised 1 μl of DNA extracted from parrot blood, 2.5 μl of NH_4 buffer (10×), 0.75 μl of $MgCl_2$ (50 mM), 0.5 μl of each primer at 25 pmol/μl (primers 2 and 4 [65]), and 0.2 μl of *Taq* polymerase (5 U/μl) made up to 25 μl with DNA-grade water. Reactions were run as follows: 92°C for 3 min, followed by 30 cycles of 92°C for 30 s, 57°C for 30 s and 72°C for 45 s, with a final 10 min at 72°C. We included the extraction blank and a negative control in all of the PCRs to check for contamination. Of the 163 Echo parakeet samples that tested positive, 74 were sequenced at ORF1 by Macrogen (Korea) using BigDye terminator reactions. Of the 46 rose-ringed parakeet samples that tested positive, 31 were sequenced at ORF1. We targeted a 765-bp region comprising most of ORF2 (*capsid* gene) (23) in samples that tested positive for the presence of BFDV. The reaction volumes and cycling conditions were identical to the ORF1 PCR except for the annealing temperature, which was reduced to 54°C. We sequenced 47 Echo parakeet and 31 rose-ringed parakeet BFDV isolates at ORF2.

Data Analyses. Sequences were aligned using the program ClustalX (32); no insertions or deletions were inferred from the alignments. Bayesian phylogenetic trees, the evolutionary rate, and the time to most recent common ancestor (TMRCA) were inferred using the program Beast, v1.6.1 (12). Two independent Monte Carlo-Markov chains were run for 200,000,000 iterations for each gene under a number of different nucleotide substitution models and tree priors with a thinning of 20,000. The Bayesian skyline coalescent demographic prior was used since it allows for temporal changes in population size (13). Runs were checked for convergence and that a reasonable effective sample size (ESS > 200) had been reached for all parameters. Runs were combined using LogCombiner (v1.6.1) and Bayes' factors (29) and parameter values obtained using Tracer (v1.5). TreeAnnotator, version 1.6.1, was used to obtain the tree with the highest clade credibility and the posterior probabilities for each node. For both genes Bayes' factor analysis indicated a general-time-reversible model of nucleotide substitution with a gamma-shaped rate variation and a proportion of invariable sites (GTR+I+Γ; identification of the best model was confirmed using the program jModelTest [46]). The evolutionary rate was inferred under a relaxed (uncorrelated exponential and uncorrelated lognormal) and strict clock. We also concatenated the two BFDV genes and repeated the Beast analyses (GTR+I+Γ was identified as the best model of evolution). The program Path-O-Gen (http://tree.bio.ed.ac.uk/software/pathogen/), which regresses the root-to-tip distance against sampling date, was used to assess the "clock-likeness" of the *capsid* and *rep* genes, using neighbor-joined trees inferred by PAUP 4.0 (61). Phylogenetic trees were visualized and drawn using TreeGraph 2 (60).

We tested for evidence of recombination using the programs SBP and GARD (44) on the Datamonkey webserver (10), and the Bootscan, GENECONV, RDP, and SiScan methods contained in the RDP3 program (35). The SBP and GARD analyses were run under a range of models of nucleotide substitution and rate variation. The RDP3 program analyses were carried out using the default parameters.

To look for evidence of positive selection (where the selection parameter ω, which represents the ratio of the nonsynonymous and synonymous substitution rates, is greater than 1), we used a number of methods. We used the likelihood ratio method of Goldman and Yang (19) implemented in PAML 4.4 (64). We compared model M7, which only allows the ω ratio to take values between zero and unity, to the selection model M8, which has an extra site class, a free ω ratio. We also used the methods implemented in the programs FEL and SLAC (43) on the Datamonkey webserver to look for positive selection (having already screened the data for recombination). Finally, to look for episodic positive selection we used the MEME method (45), which identifies sites where a proportion of branches have evolved under positive selection.

To investigate the possible effect of synonymous substitutions on DNA folding, we generated secondary structure predictions for both BFDV genes using the web-based RNAfold webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) with the DNA parameters (21).

RESULTS

Changes in BFDV in Echo parakeet host population. Analyses using microsatellite DNA, cytochrome *b*, and cytochrome *c* oxidase subunit I (COI) loci amplified from the host (data not shown) indicated no obvious evidence for cross-contamination between host individuals or species. All sequences derived from the present study have been submitted to GenBank under the following accession numbers: HQ641457 to HQ641563 and HQ662336 to HQ662409. To identify the mutations that may have elicited the 2005/2006 season (2005/06) PBFD outbreak, we constructed phylogenies using the Echo parakeet BFDV isolates. We found no significant evidence for recombination using any method and, consequently, the *capsid* and *rep* genes were concatenated to produce the dated phylogeny of Echo parakeet isolates shown in Fig. 1 (inferred using a relaxed lognormal molecular clock; the frequency histogram of the mean coefficient of variation [mean CoV = 1.09] does not sit against zero, indicating rate heterogeneity). Of the isolates sampled during the PBFD outbreak, the majority (37 of 42) group together within one large clade (highlighted by the gray box). All of the isolates collected after this

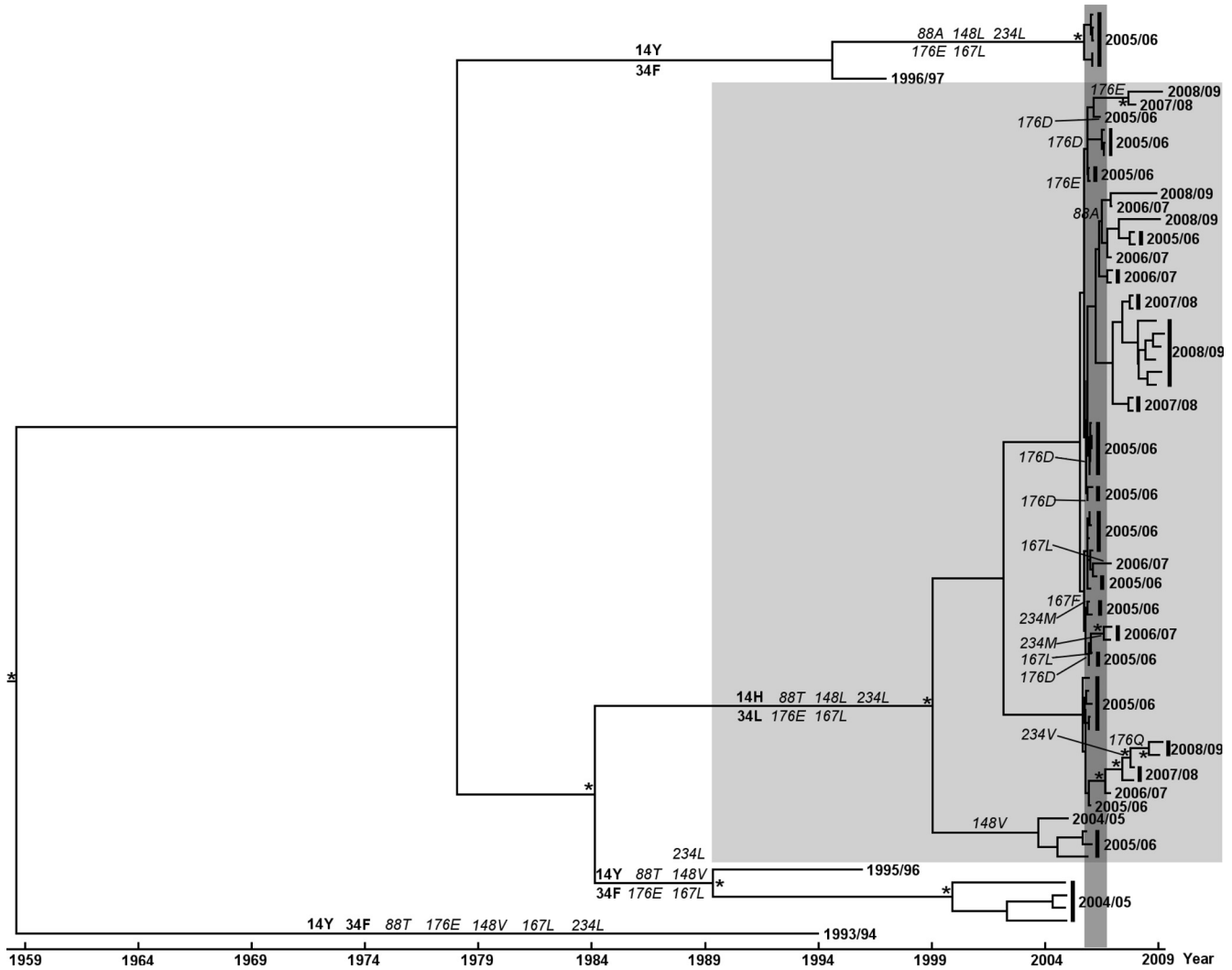

FIG 1 Maximum clade credibility tree showing the inferred phylogenetic relationships between BFDV isolates collected from the Echo parakeet population (from Bayesian analysis of the *capsid* and *rep* genes using Beast). The tips are labeled with the Echo parakeet breeding season from which the sample was collected. The tree was automatically rooted by using a relaxed clock model in Beast. Nodes with posterior probability of ≥0.9 are indicated with asterisks and with a double-dagger for $P \geq 0.7$. The dark grey vertical bar highlights the 2005/06 PBFD outbreak season. Inferred nonsynonymous substitutions at codons in the *replicase* (in boldface) and *capsid* (in italics) are indicated at the appropriate lineages (where the majority of isolates in a clade possess a particular substitution then the ancestral node has been labeled). The clade highlighted by the gray box represents all of the isolates with the 14H and 34L *replicase* gene substitutions identified by the present study as closely corresponding to the PBFD outbreak. The estimated mean TMRCA for all of the isolates is 1959 (95% HPD 1920 and 1988).

period are also found in this clade. In contrast to the exception of an isolate collected in 2004/05, all of the isolates that predated the outbreak were found outside of the highlighted clade. The clade support values indicate good evidence for this differentiation of BFDV isolates sampled prior to the PBFD outbreak from those sampled during and afterwards, although they also indicate that there is no clear pattern of genetic differentiation within the highlighted clade. Of the 10 mutations that appear to differentiate the highlighted clade from the remaining isolates eight occur in the *rep* gene, two of which are nonsynonymous substitutions. The first is a mutation within a potential pyrophosphatase domain (38), giving rise to a histidine residue (from tyrosine) at codon 14. The second occurs at codon 34, which is in a rolling-circle replication motif (26), and was a substitution from a phenylalanine residue to a leucine (both changes have been highlighted in Fig. 1). The *capsid* gene had just two substitutions, both synonymous, that clearly differentiated the two phylogenetic groups. However, we observed eight polymorphic sites with nonsynonymous changes involving more than one isolate of which codons 88, 148, 167, 176, and 234 appeared to be particularly variable (these are mapped onto Fig. 1). The pattern of changes indicates that prior to 2005 the "14Y34F" *rep* and "88T148V176E167L234M" *capsid* alleles were present in the majority of Echo parakeet isolates. From early 2005 a new *rep* allele, "14H34L," appears in the population and represents almost all of the isolates sampled during and after the PBFD outbreak; after February 2006, the "14Y34F" *rep* allele is no longer present. Multiple *capsid* alleles are evident both during and after the outbreak, including the "88T148V176E167L234M" allele that was present in the population before 2005.

Epidemic history. A Bayesian skyline plot depicting the inferred effective number of BFDV infections in the Echo parakeet population over time is shown in Fig. 2a (from the Beast analysis

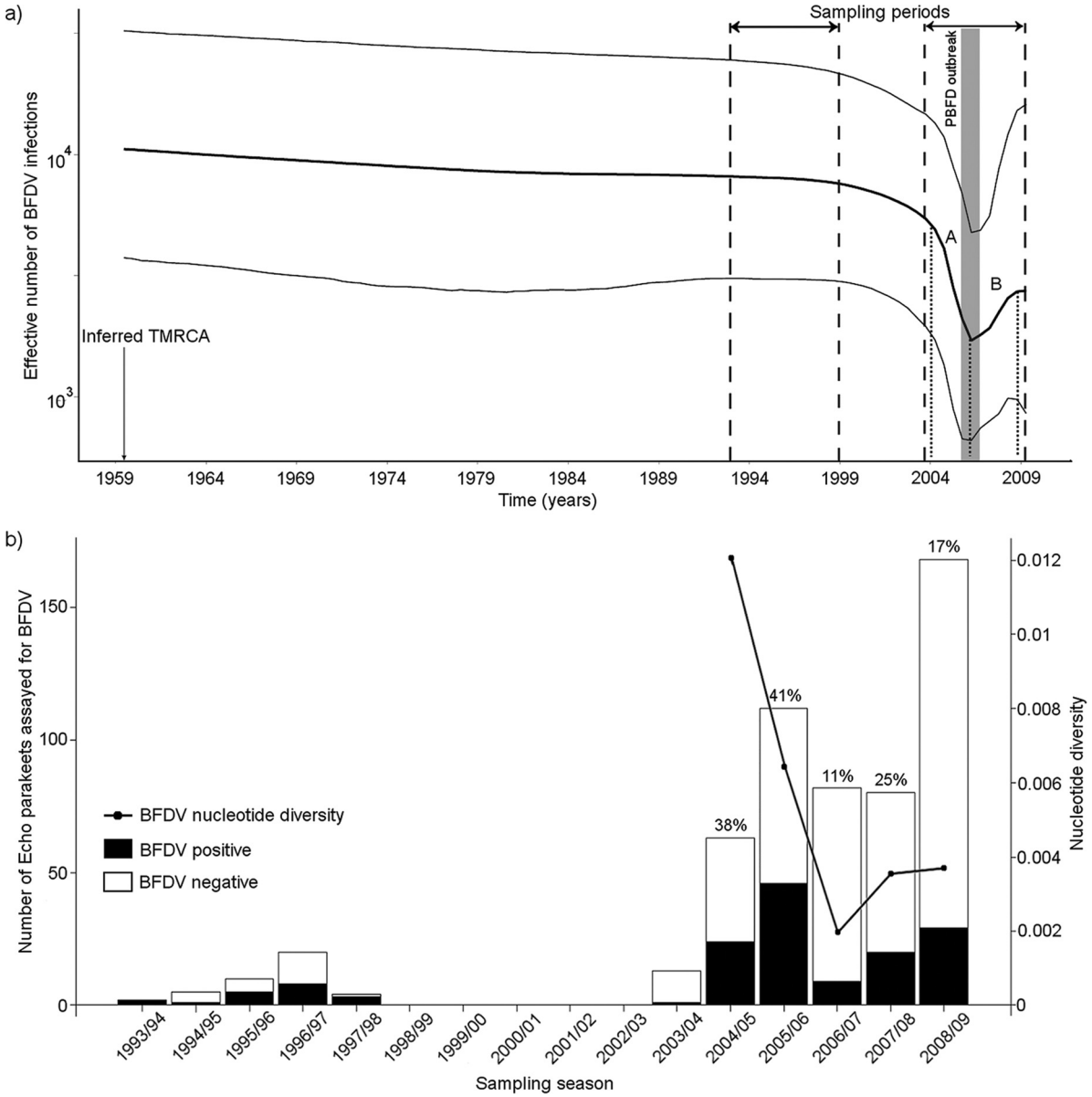

FIG 2 Prevalence and nucleotide diversity (average number of nucleotide differences per site) of BFDV in the Echo parakeets. (a) Bayesian skyline plot showing the effective number of BFDV infections in the Echo parakeet population since the estimated mean TMRCA (in 1959). The plot shows a period of exponential decline in viral diversity (highlighted as "A") coinciding with the 2005/06 PBFD outbreak (highlighted by gray shading), followed by a subsequent increase ("B"). (b) Number of infected and uninfected Echo parakeets for the breeding seasons between 1993 and 1998 and between 2003 and 2009 (the estimated prevalence of BFDV per season from 2004 is also indicated). The estimated nucleotide diversity of BFDV is also shown from the 2004/05 season and indicates a reduction in genetic diversity during the PBFD outbreak and the following season (2006/07).

used to infer the phylogenetic relationships in Fig. 1). The skyline plot, which gives an indication of change in viral diversity over time (13), indicates no evidence for any change from the estimated TMRCA (1959 [95% HPD = 1920 to 1987]) until 2004 when there is a period of exponential decline (labeled as "A"). This decline continues until approximately the middle of the PBFD outbreak, after which there is a period of increase (labeled "B"). The skyline plots inferred for each gene separately show a similar pattern to that in Fig. 2a (data not shown). The estimated nucleotide diversity (π) of BFDV in each Echo parakeet breeding season from 2004/05 is shown in Fig. 2b (alongside the prevalence); we did not calculate π for previous seasons since there were too few isolates to obtain a reliable estimate. The pattern is consistent with that indicated by the skyline plot and shows a decline in genetic diversity during the outbreak (π = 0.0064, standard deviation [SD] = 0.0013) and in season 2006/07 (π = 0.0019, SD = 0) compared to that estimated from 2004/05 (π = 0.012, SD = 0.005).

Transmission between the Echo parakeet and feral rose-ringed parakeet population. Figure 3 shows the phylogenetic relationships among BFDV isolates from both parrot populations in Mauritius (also inferred using the relaxed molecular clock since

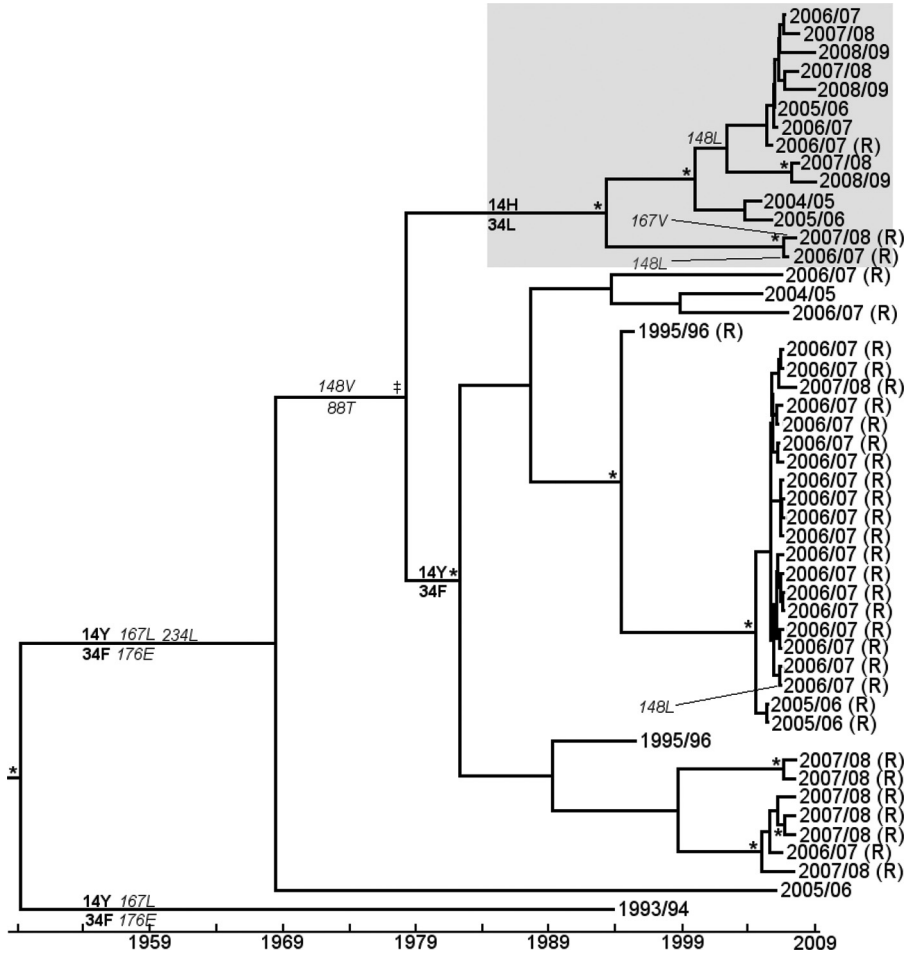

FIG 3 Phylogenetic relationships between BFDV isolates from the Echo parakeet and rose-ringed parakeet inferred from Bayesian analysis using Beast (the tree was automatically rooted using a relaxed clock model). Nodes with a posterior probability of ≥0.9 are indicated with an asterisk and with a double dagger for $P \geq 0.7$. Inferred nonsynonymous substitutions at codons in the *replicase* (in boldface) and *capsid* (in italics) genes are indicated. To ease visualization, we include only 15 isolates from the Echo parakeet population representing all of the major genotypes identified in Fig. 1. Isolates from rose-ringed parakeets are indicated by the suffix "(R)". As in Fig. 1, the gray-shaded box represents all of the isolates with the 14H and 34L *replicase* gene substitutions identified as corresponding to the outbreak. The estimated mean TMRCA for all sequences is 1949 (95% HPD = 1899 and 1983).

the frequency histogram for the coefficient of variation indicated evidence against the strict clock); for ease of visualization, we do not include all Echo parakeet isolates but instead retain 15 isolates representative of the polymorphisms identified in Fig. 1. The rose-ringed parakeet isolates, which in Fig. 3 are indicated by "(R)," form two clades: the majority are closely related to the Echo parakeet viral isolates carrying the 14Y34F *rep* allele, i.e., those that were identified from Fig. 1 to not correspond to the PBFD outbreak. This group includes rose-ringed parakeets sampled at the same time as the PBFD outbreak in the Echo parakeet population. The second group (highlighted by the gray box), comprising just three isolates, have the 14H34L *rep* allele that appears to closely correspond to the outbreak. This second group also shares some of the synonymous substitutions that correspond to the PBFD outbreak; one isolate shares three of six of the *rep* gene substitutions and two others each carry one of the *capsid* gene substitutions. The five sites in the *capsid* gene that we observed to be particularly polymorphic in the Echo parakeet isolates (codons 88, 148, 167, 176, and 234) appeared to be much less variable in the rose-ringed parakeet isolates. Only codons 148 and 167 showed nonsynonymous polymorphism.

Positive selection in the BFDV genes. Table 1 shows the positively selected sites inferred by the FEL, PAML, and SLAC methods. These analyses failed to find any significant evidence for positive selection in either the *capsid* gene of rose-ringed parakeet isolates or the *rep* gene of Echo parakeet isolates. Four of the five *capsid* gene polymorphisms highlighted in Fig. 1 have been identified as positively selected, but inconsistently across the three methods: codons 167 and 176 by PAML (significant evidence [1%] in favor of the selection model, M8, over M7 $2\Delta L = 10.714$, $v = 2$, where $2\Delta L$ is the test statistic and v is the number of degrees of freedom), codon 148 by FEL and codon 234 by SLAC. The polymorphic sites in the *replicase* gene that we identified as most closely associated with the PBFD outbreak, 14 and 34, have not been identified by any method. The substitution counts determined by SLAC showed that except for sites 88, 167, and 176, there is no evidence for synonymous substitutions at any of the sites in Fig. 1. The phylogenetic pattern of these substitutions (Fig. 1)

TABLE 1 Positively selected sites in the BFDV genes of the Echo parakeet and feral rose-ringed parakeet isolates inferred by a number of methods[a]

Species	Gene	No. of sequences	Codon(s)			
			PAML, model M8 (ω; P)	FEL (ω; P)	SLAC (ω; P)	MEME (P)
Echo parakeet	*capsid*	50	167 (3.8; 0.99), 176 (10.3; 0.99)	126 (∞; 0.03), 148 (∞; 0.05)	234 (7.28; 0.04)	36 (0.09), 121 (0.005), 126 (0.001), 148 (0.02), 167 (0.07), 176 (0.07), 234 (0.03)
	rep	73	No sites	No sites	No sites	8 (0.06)
Rose-ringed parakeet	*capsid*	23	No sites	No sites	No sites	60 (0.03), 246 (0.01)
	rep	34	227 (2.26; 0.95 [NS])	180 (∞, 0.08)	No sites	159 (0.08), 180 (0.05)

[a] The number of sequences in the analysis is specified. ω is the selection parameter, and P the posterior probability of $\omega > 1$. NS, not significant. The PAML, FEL, SLAC, and MEME methods are described in the text.

suggests that tests of episodic selection may be more appropriate than the site-prediction methods in FEL, PAML, and SLAC. The results of the MEME analyses for episodic selection (Table 1) indicate that seven codons in the *capsid* (including sites 148, 167, 176, and 234) and one in the *rep* gene showed evidence of periods of positive selection in the evolutionary history of BFDV in the Echo parakeets. Codons 14 and 34 in the *rep* have not been identified.

Secondary structure in BFDV genes. Previous studies have highlighted the selection potential of synonymous substitutions in viruses with highly compact genomes primarily as a consequence of their effect on the formation of secondary structure. Indeed, analyses of the intergenic region between ORF1 and ORF2 in BFDV and the related porcine circovirus have identified a putative stem-loop structure that may be involved in replication (3, 59). Eight of the substitutions that closely correspond to the PBFD outbreak were synonymous, and we investigated their effects upon the predicted secondary structure (data not shown). Of the two observed in the *capsid*, only the substitution at codon 174 has an effect, resulting in the formation of a number of additional stem-loop structures. We observed the synonymous substitution at codon 121 in the *rep* gene to have a minimal and localized effect altering a predicted loop structure. Substitutions at codons 54 and 84 have even smaller effects, resulting in minor changes to the loop structures that these sites occur in. The remaining substitutions appear to have no effect whatsoever.

Evolutionary rate of BFDV genes. Across all of the clock models, the mean evolutionary rate estimates varied only slightly (data not shown), falling within a narrow range between 10^{-4} and 10^{-3} substitutions per site per year, which is high but consistent with rates estimated from other ssDNA and ssRNA viruses (15). The coefficient of variation histograms indicated that for the Echo parakeet isolates there is not sufficient evidence to reject the strict clock, but there is for the rose-ringed parakeet isolates. Table 2 shows the rate estimated from the strict clock for the Echo parakeet isolates and the relaxed clock (uncorrelated lognormal) for the rose-ringed isolates. The inferred rates do not differ substantially between the two genes or between the two host populations, although the 95% HPD intervals for the *capsid* gene analyses from rose-ringed isolates are very wide, suggesting that there may not be very much temporal structure. Given the phylogenetic evidence for allele-sharing between the two host populations, we also estimated the rate based on the combined data where there was sufficient evidence to reject the strict molecular clock for the *rep* gene but not for the *capsid* gene.

To assess the level of clock-like evolution in the data, we regressed the root-to-tip genetic distances, inferred from neighbor-joined trees, against sampling time by using Path-O-Gen. The results indicate that there is good evidence for temporal signal in the c*apsid* gene but not in the *replicase*: (i) Echo parakeet isolate *capsid* gene, $r^2 = 0.625$, residual mean square (rms) $= 3.23 \times 10^{-6}$, and *replicase* gene, $r^2 = 0.122$, rms $= 1.55 \times 10^{-6}$, and (ii) rose-ringed parakeet isolate *capsid* gene, $r^2 = 0.461$, rms $= 4.32 \times 10^{-5}$, and *replicase* gene, $r^2 = 0.073$, rms $= 7.21 \times 10^{-7}$. Following Duffy and Holmes (14), to assess the temporal structure of the evolutionary rate, analyses were repeated for each data set, but with the sampling dates randomly shuffled among the tips. In both genes of the Echo parakeet isolates the actual estimated mean

TABLE 2 Mean evolutionary rate and the 95% HPD intervals estimated for the *capsid* and *replicase* genes of BFDV isolates from Echo and feral rose-ringed parakeets[a]

Host species	Gene	Clock model	Mean evolutionary rate (per site/yr)	95% HPD interval	
				Lower	Upper
Echo parakeet	*capsid*	Strict	1.01×10^{-3}	5.63×10^{-4}	1.44×10^{-3}
	rep	Strict	3.44×10^{-4}	1.39×10^{-4}	5.67×10^{-4}
Rose-ringed parakeet	*capsid*	Relaxed (ULN)	5.50×10^{-3}	1.99×10^{-6}	9.63×10^{-3}
	rep	Relaxed (ULN)	2.75×10^{-3}	7.31×10^{-4}	4.44×10^{-3}
Combined data set	*capsid*	Relaxed (ULN)	1.08×10^{-3}	6.15×10^{-4}	1.63×10^{-3}
	rep	Relaxed (ULN)	4.74×10^{-4}	1.65×10^{-4}	8.31×10^{-4}

[a] Estimates obtained by combining the data from the two host populations are also shown. For each data set, we only included results from the significant clock model, i.e., the strict clock was rejected for the rose-ringed and combined datasets (in favor of the relaxed uncorrelated lognormal clock [ULN], which had a higher marginal likelihood than the relaxed uncorrelated exponential clock).

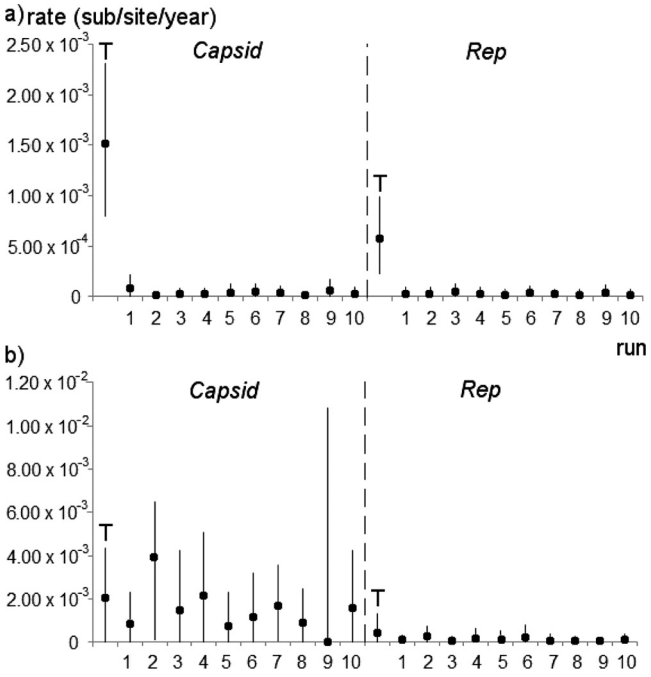

FIG 4 Evolutionary rate of the *capsid* and *rep* BFDV genes (mean and 95% HPD) estimated from the true dates of sample isolation (labeled as the "T" data point) and from the shuffled dates of isolation. Analyses using the shuffled dates were performed using a GTR+Γ+I model of evolution and a relaxed (uncorrelated lognormal) molecular clock. (a) Results from the Echo parakeet data set indicate no overlap between the estimates from the true dates and from the shuffled dates, indicating good temporal signal. (b) Results from the rose-ringed parakeet isolates indicate no clear temporal signal.

evolutionary rates do not coincide with the 95% HPD intervals for any of the randomized runs, indicating that there is good temporal structure (Fig. 4). In contrast, there is little difference between the estimated evolutionary rates from the rose-ringed parakeet isolates and their randomized results, suggesting that there is no temporal structure. Taken together, these results suggest that the high evolutionary rate estimated from the BFDV *capsid* gene from Echo parakeet isolates is well supported. There is marginal support for the rate estimated from the *rep* gene in the Echo parakeet isolates but no support for the results from the rose-ringed parakeet BFDV data.

DISCUSSION

The pattern of changes that we observed in the *rep* gene suggests that there was a major selective sweep of BFDV mutations that coincided with the outbreak of PBFD in the Echo parakeet. The speed and extent of this sweep is such that the *rep* allele occurring prior to the outbreak appears to be completely absent in the season immediately afterward. A number of factors may have contributed to the scope and rapidity of the sweep, including the environmental stability of circoviruses (62), both vertical and horizontal modes of transmission (18, 48, 63), and host population dynamics (infection in a small population of a highly gregarious species).

Of course, the pattern of changes observed in *rep* may represent stochastic events rather than the effects of natural selection. In addition, we observed a significant decline in viral diversity at both BFDV genes and events that reduce variation may be demographic, e.g., bottlenecks, which increase the rate of drift, or selective, sweeps are characterized by a rapid reduction in variability at linked loci (7, 36). Since the fixation of *rep* alleles occurred during an outbreak, we suggest that strong directional selection for adaptive changes was responsible rather than drift. Moreover, no significant declines have been observed in the Echo parakeet population since 1993 and analyses of the host genetic diversity using microsatellite DNA loci has not identified any significant evidence for a secondary bottleneck (49). Although changes in viral coat proteins are most commonly associated with disease outbreaks, mutations in viral replication genes have been shown to result in increased replication activity and virulence (20, 53). It is notable that the changes that we observed occur in potentially important regions (a pyrophosphatase domain and a rolling-circle replication motif). It is still possible that the *rep* gene changes that we observed are nonadaptive and have simply been "carried" through a sweep by genetic hitchhiking to other mutations. The regularity of our sampling suggests that it is unlikely that we missed any other key mutations, and the phylogenetic pattern of nonsynonymous changes in the *capsid* gene does not correspond as well to the chronology of the outbreak. However, we did observe a number of synonymous substitutions in both BFDV genes that also correspond to the appearance of the PBFD outbreak. Previous studies in RNA viruses, particularly those carrying a positive-stranded genome, have indicated that synonymous substitutions may be under selection, often as a consequence of their impact upon secondary structure, which itself can affect translation and gene expression (39, 58, 8). Small ssDNA viruses may behave in a similar way, and our analyses indicate that two of the synonymous substitutions that we observed, one in the *capsid* gene and the other in the *rep* gene, may indeed affect the secondary structure. However, a recent experimental analysis of synonymous substitutions in viruses found that while they may have significant fitness effects in ssRNA viruses, the same may not be true for ssDNA viruses, where the effect appears to be considerably diminished (8). These observations, taken together, suggest that the two nonsynonymous substitutions in the *rep* gene at codons 14 and 34 remain the most plausible explanation for the PBFD outbreak in the Echo parakeets, although ultimately some degree of experimental validation to ascertain their functional relevance will be needed. Currently, it is difficult to conceive of how this may be achieved since our study system includes a critically endangered host species and a virus for which there is currently no cell culture system (24).

While adaptive changes in the *rep* gene appear to be the likeliest explanation for the PBFD outbreak, the presence of variation in the capsid may also have been important. The analyses showed the presence of multiple *capsid* alleles linked to the outbreak, and almost 10% of the gene showed some level of nonsynonymous diversity. Mutations in capsid proteins can affect a virus' ability to avoid the immune system and cause infection. Consequently, it may be that while the outbreak was elicited by changes in the *rep* gene, variation in the *capsid* gene was an additional factor in the dissemination of the virus.

Our analyses indicated little consistent evidence for positive selection despite the presence of potentially adaptive mutations. Although current site prediction methods arguably represent the most rigorous tools for analyzing selection in DNA sequences (2), it has been suggested that the statistical methods that underpin them have some limitations (40). Specifically, many of these methods require considerable sequence divergence (1) and have difficulty in identifying rare mutational events that are adaptive (40). The phylogenetic analyses suggest that the nonsynonymous substitutions in the *rep* gene that we have identified as most closely

linked to the outbreak arose only once. These sites also had no other nonsynonymous variation, i.e., they are characterized by a pattern of selective neutrality or purifying selection punctuated by a single potentially adaptive event. The results thus seemingly corroborate previous studies suggesting that site prediction methods may be inappropriate for detecting rare adaptive events, including those that have gone through a major selective sweep. Given the role of the capsid, adaptive changes are expected to be comparatively frequent and thus more detectable. A previous study of BFDV did indeed identify evidence of positive selection in this gene, albeit using only a single method of analysis (23). However, even here our results were not unequivocal across three of the site-prediction methods that we used (FEL, PAML, and SLAC). The fourth method for detecting selection, MEME, attempts to identify sites that have periodically experienced positive selection over their evolution (45). The majority of the polymorphic sites that we document in Fig. 1 were identified by this method, suggesting that the evolution of the *capsid* gene is characterized by episodic positive selection. This is consistent with a pattern of frequency-dependent selection, which is often observed in the dynamic between pathogens and the host immune system (11, 31, 42, 47).

Considering the proximity of the two parrot populations in Mauritius and the recent common ancestry of the host species (22), we expected to find good evidence for viral transmission. Indeed, the phylogenetic analyses indicated the presence of BFDV allele sharing between the two populations. Although the literature (28) and field reports indicate behavioral and ecological separation between the parrot populations in Mauritius, the environmental stability of BFDV (62) means that the virus still can be indirectly transmitted. The most recent common ancestor for all Echo parakeet isolates dates to around 1959 (or 1949 when the rose-ringed isolates are included in the analysis), which approximately coincides with a period when the rose-ringed parakeet started expanding into native forest, coming into direct competition with the Echo parakeet for nest sites (6). Interestingly, the oldest rose-ringed parakeet isolate was taken from a bird housed in aviaries that were central to the initial recovery of the Echo parakeets. Thus, the apparent transmission of BFDV actually may in some cases be attributable to human intervention. Although the evidence for viral transmission between the two parrot populations appears to be clear, the evidence for the rose-ringed parakeets being the source of the PBFD outbreak is more equivocal. We only identified three rose-ringed parakeet isolates that possessed the 14H34L *rep* allele, and all of them were collected at least one season after the outbreak. One possibility is that the mutations responsible for the outbreak originated in the Echo parakeet population with subsequent transmission to the rose-ringed population. Indeed, one of the three rose-ringed isolates was collected from an individual that most likely became infected after indirect contact with the endemic population (it was found in an Echo parakeet nest in the 2006/07 season). Nevertheless, given the considerable size and distribution of the invasive population, it undoubtedly remains the likeliest source of the PBFD outbreak in the Echo parakeets. It may be that our sampling of this population prior to 2005/06 was not sufficient to identify the 14H34L mutations.

Documented evolutionary rates for ssDNA viruses are very comparable to those of RNA viruses (15, 50, 55, 56). It has been suggested that ssDNA viruses may tolerate such high rates since their generally small genome size may be less likely to accumulate deleterious mutations (25). The only previous study to have estimated an evolutionary rate for the *Circoviridae* (27) suggested that this group of viruses have an evolutionary rate only about three times higher than that of the vertebrate host. Clearly, this rate is lower than that implied by our study and appears contrary to the relationship between viral genome size and evolutionary rate; circoviruses have the smallest genomes of all known DNA viruses. However, Holmes (25) points to a number of flaws in the Johne et al. study (27), including the limited number of host species and the presence of within host genetic diversity. A caveat to our own findings is the potential effect of the background mutation rate to inflate our estimates for the evolutionary rate. It is possible that some of the variation identified here may be from deleterious mutations that have yet to reach fixation. The randomization analyses of the rose-ringed parakeet isolates suggested that there is insufficient support for the evolutionary rate inferred from this data set. However, our analyses suggest that the rate estimated from the BFDV *capsid* gene in the Echo parakeet isolates is well supported. This result indicates that as with other ssDNA viruses, BFDV has a high evolutionary rate.

Concluding remarks. The growing impact of emerging infectious disease in humans, livestock, and endangered species has increased the importance of understanding the changes that can elicit an outbreak and how a pathogen subsequently evolves during such a disease. Although infectious disease in endangered populations represents a major problem for biodiversity, such events also present an excellent opportunity to resolve evolutionary patterns in pathogens. Populations such as the Echo parakeet, which have been sampled in a frequent and unbiased (in terms of disease) fashion, enable studies to access a level of detail not normally possible in a natural population.

ACKNOWLEDGMENTS

We thank Paul-Michael Agapow, Amanda Church, Richard Nichols, Andy Overall, Claire Raisin, and three anonymous reviewers for their helpful comments. Samples were collected by the Mauritius Wildlife Foundation.

This study was supported by a grant awarded to J.J.G. by the Leverhulme Trust, United Kingdom (F/00 236/T).

REFERENCES

1. **Anisimova M, Bielawski JP, Yang Z.** 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. **19**:950–958.
2. **Anisimova M, Kosiol C.** 2009. Investigating protein-coding sequence evolution with probabilistic codon substitution models. Mol. Biol. Evol. **26**:255–271.
3. **Bassami MR, Berryman D, Wilcox GE, Raidal SR.** 1998. Psittacine beak and feather disease virus nucleotide sequence analysis and its relationship to porcine circovirus, plant circoviruses, and chicken anaemia virus. Virology **249**:453–459.
4. **Bassami MR, Ypelaar I, Berryman D, Wilcox GE, Raidal SR.** 2001. Genetic diversity of beak and feather disease virus detected in psittacine species in Australia. Virology **279**:392–400.
5. **Bergstrom CT, McElhany P, Real LA.** 1999. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proc. Natl. Acad. Sci. U. S. A. **96**:5095–5100.
6. **Cheke A, Hume JP.** 2008. Lost land of the dodo: the ecological history of Mauritius, Reunion, and Rodrigues. T&AD Poyser, London, England.
7. **Cobey S, Koelle K.** 2008. Capturing escape in infectious disease dynamics. Trends Ecol. Evol. **23**:572–577.
8. **Cuevas JM, Domingo-Calap P, Sanjuán R.** 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. Mol. Biol. Evol. **29**:17–20.
9. **de Kloet E, de Kloet SR.** 2004. Analysis of the beak and feather disease viral genome indicates the existence of several genotypes which have a complex psittacine host specificity. Arch. Virol. **149**:2393–2412.

10. **Delport W, Poon AF, Frost SDW, Pond SLK.** 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics **26**:2455–2457.
11. **Dionne M, Miller KM, Dodson JJ, Bernatchez L.** 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos. Trans. R. Soc. B **364**:1555–1565.
12. **Drummond AJ, Rambaut A.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7**:1–8.
13. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG.** 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. **22**:1185–1192.
14. **Duffy S, Holmes EC.** 2009. Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. J. Gen. Virol. **90**:1539–1547.
15. **Duffy S, Shackelton LA, Holmes EC.** 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. **9**:267–276.
16. **Faurez F, Dory D, Grasland B, Jestin A.** 2009. Replication of porcine circoviruses. Virology **6**:60.
17. **Fenaux M, Opriessnig T, Halbur PG, Elvinger F, Meng XJ.** 2004. Two amino acid mutations in the capsid protein of type 2 porcine circovirus (PCV2) enhanced PCV2 replication in vitro and attenuated the virus in vivo. J. Virol. **78**:13440–13446.
18. **Gerlach H.** 1994. *Circoviridae*-psittacine beak and feather disease virus, p 894–903. *In* Ritchie BW, Harrison GT, Harrison LR (ed), Avian medicine: principles and practice. Wingers Publishing, Inc, Lake Worth, FL.
19. **Goldman N, Yang Z.** 1994. Codon-based model of nucleotide-based substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.
20. **Goodman LB, et al.** 2007. A point mutation in a Herpesvirus polymerase determines neuropathogenicity. PLoS Pathog. **3**:1583–1592.
21. **Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL.** 2008. The Vienna RNA websuite. Nucleic Acids Res. **36**:W70–W74.
22. **Groombridge JJ, Jones CG, Nichols RA, Carlton M, Bruford MW.** 2004. Molecular phylogeny and morphological change in the *Psittacula* parakeets. Mol. Phylogenet. Evol. **31**:96–108.
23. **Heath LD, et al.** 2004. Evidence of unique genotypes of beak and feather disease virus in southern Africa. J. Virol. **78**:9277–9284.
24. **Heath LD, Williamson A-L, Rybicki EP.** 2006. The capsid protein of beak and feather disease virus binds to the viral DNA and is responsible for transporting the replication-associated protein into the nucleus. J. Virol. **80**:7219–7225.
25. **Holmes EC.** 2009. The evolution and emergence of RNA viruses. Oxford University Press, Oxford, United Kingdom.
26. **Ilyina TV, Koonin EV.** 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eukaryotes, and archaebacteria. Nucleic Acids Res. **20**:3279–3285.
27. **Johne R, Fernandez-de-Luco D, Hofle U, Muller H.** 2006. Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. J. Gen. Virol. **87**:1189–1195.
28. **Jones CG.** 1987. The larger land-birds of Mauritius, p 208–301. *In* Diamond AW (ed), Studies of Mascarene island birds. Cambridge University Press, Cambridge, United Kingdom.
29. **Kass RE, Raftery AE.** 1995. Bayes factors. J. Am. Stat. Assoc. **90**:773–795.
30. **Kondiah K, Albertyn J, Bragg RR.** 2006. Genetic diversity of the *rep* gene of beak and feather disease virus in South Africa. Arch. Virol. **151**:2539–2545.
31. **Koskella B, Lively CM.** 2009. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. Evolution **63**:2213–2221.
32. **Larkin MA, et al.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947–2948.
33. **Latimer KS, et al.** 1990. Extracutaneous viral inclusions in psittacine beak and feather disease. J. Vet. Diagn. Invest. **2**:204–207.
34. **Mankertz A, Mankertz J, Wolf K, Buhk HJ.** 1998. Identification of a protein essential for replication of porcine circovirus. J. Gen. Virol. **79**:381–384.
35. **Martin DP, Lemay P, Lott M, Moulton V, Posada D, LeFeuvre P.** 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics **26**:2462–2463.
36. **Maynard-Smith J, Haigh J.** 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.
37. **Misinzo G, Delputte PL, Meerts P, Lefebvre DJ, Nauwynck HJ.** 2006. Porcine circovirus 2 uses heparan sulfate and chondroitin sulfate B glycosaminoglycans as receptors for its attachment to host cells. J. Virol. **80**:3487–3494.
38. **Niagro FD, et al.** 1998. Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. Arch. Virol. **143**:1723–1744.
39. **Novella IS, Zárate S, Metzger D, Ebendick-Corpus BE.** 2004. Positive selection of synonymous mutations in vesicular stomatitis virus. J. Mol. Biol. **342**:1415–1421.
40. **Nozawa M, Suzuki Y, Nei M.** 2009. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. Proc. Natl. Acad. Sci. U. S. A. **106**:6700–6705.
41. **Pass DA, Perry RA.** 1984. The pathology of psittacine beak and feather disease. Aust. Vet. J. **61**:69–74.
42. **Plotkin JB, Dushoff J.** 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. Proc. Natl. Acad. Sci. U. S. A. **100**:7152–7157.
43. **Pond SLK, Frost SDW.** 2005. Not so different after all: a comparison of methods for detecting amino acids under selection. Mol. Biol. Evol. **22**:1208–1222.
44. **Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23**:1891–1901.
45. **Pond SLK, et al.** 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. **28**:3033–3043.
46. **Posada D.** 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. **25**:1253–1256.
47. **Potts WK, Slev PR.** 1995. Pathogen-based models favoring MHC genetic diversity. Immunol. Rev. **143**:181–197.
48. **Rahaus M, et al.** 2008. Detection of beak and feather disease virus DNA in embryonated eggs of psittacine birds. Vet. Med. **1**:53–58.
49. **Raisin C, et al.** Genetic consequences of intensive conservation management for the Mauritius parakeet. Conserv. Genet., in press.
50. **Ramsden C, Melo FL, Figueiredo LM, Holmes EC, Zanotto PMA.** 2008. High rates of molecular evolution in hantaviruses. Mol. Biol. Evol. **25**:1488–1492.
51. **Ritchie BW, et al.** 1990. Ultrastructural, protein composition, and antigenic comparison of psittacine beak and feather disease virus purified from four genera of psittacine birds. J. Wildl. Dis. **26**:196–203.
52. **Ritchie BW, Niagro FD, Lukert PD, Steffens WL, Latimer KS.** 1989. Characterisation of a new virus from cockatoos with psittacine beak and feather disease. Virology **171**:83–88.
53. **Rolling T, et al.** 2009. Adaptive mutations resulting in enhanced polymerase activity contribute to high virulence of influenza A virus in mice. J. Virol. **83**:6673–6680.
54. **Sambrook J, Fritsch EF, Maniatis T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
55. **Shackelton LA, Holmes EC.** 2006. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. J. Virol. **80**:3666–3669.
56. **Shackelton LA, Parrish CR, Truyen U, Holmes EC.** 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. U. S. A. **102**:379–384.
57. **Shearer P, Bonne N, Clark P, Sharp M, Raidal SR.** 2008. Beak and feather disease virus infection in cockatiels (*Nymphicus hollandicus*). Avian Pathol. **37**:75–81.
58. **Simmonds P, et al.** 2008. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. Nucleic Acids Res. **36**:2530–2546.
59. **Steinfeldt T, Finsterbusch T, Mankertz A.** 2006. Demonstration of nicking/joining activity at the origin of DNA replication associated with the Rep and Rep′ proteins of porcine circovirus type 1. J. Virol. **80**:6225–6234.
60. **Stover B, Muller K.** 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinform. **11**:7.
61. **Swofford DL.** 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.
62. **Todd D.** 2000. Circoviruses: immunosuppressive threats to avian species: a review. Avian Pathol. **29**:373–394.
63. **Todd D.** 2004. Avian circovirus diseases: lessons for the study of PMWS. Vet. Microbiol. **98**:169–174.
64. **Yang Z.** 2007. PAML4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.
65. **Ypelaar I, Bassami MR, Wilcox GE, Raidal SR.** 1999. A universal polymerase chain reaction for the detection of psittacine beak and feather disease virus. Vet. Microbiol. **68**:141–148.