Virus Genes (2019) 55:802–814
https://doi.org/10.1007/s11262-019-01702-x

ORIGINAL PAPER

Assessing circovirus gene flow in multiple spill-over events

Shubhagata Das[1] · Kate Smith[2] · Subir Sarker[3] · Andrew Peters[1] · Katherine Adriaanse[4] · Paul Eden[4] · Seyed A. Ghorashi[1] · Jade K. Forwood[2] · Shane R. Raidal[1]

Received: 31 March 2019 / Accepted: 19 August 2019 / Published online: 28 August 2019

Abstract

The establishment of viral pathogens in new host environments following spillover events probably requires adaptive changes within both the new host and pathogen. After many generations, signals for ancient cross-species transmission may become lost and a strictly host-adapted phylogeny may mimic true co-divergence while the virus may retain an inherent ability to jump host species. The mechanistic basis for such processes remains poorly understood. To study the dynamics of virus–host co-divergence and the arbitrary chances of spillover in various reservoir hosts with equal ecological opportunity, we examined structural constraints of capsid protein in extant populations of *Beak and feather disease virus* (BFDV) during known spillover events. By assessing reservoir-based genotype stratification, we identified co-divergence defying signatures in the evolution BFDV which highlighted primordial processes of cryptic host adaptation and competing forces of host co-divergence and cross-species transmission. We demonstrate that, despite extensive surface plasticity gathered over a longer span of evolution, structural constraints of the capsid protein allow opportunistic host switching in host-adapted populations. This study provides new insights into how small populations of endangered psittacine species may face multidirectional forces of infection from reservoirs with apparently co-diverging genotypes.

Keywords Circovirus · Psittacine beak and feather disease · PBFD · Microbiome · Viral ecology

Introduction

Forces of co-divergence and cross-species transmission have been suggested to play key roles in the evolution of viruses and emergence of disease in naive populations [1]. However, emerging evidence suggests that true viral co-divergence is a rarity, whereas cross-species transmission plays a greater role in the evolution of both DNA and RNA viruses [2]. Successful cross-species transmission occurs more frequently among phylogenetically related hosts because it is easier to infect and replicate in genetically similar hosts that share less divergent cell receptors [3]. Moreover, related hosts may sometimes occupy the same geographic region or overlapping ecological niches which increases the probability of cross-species transmission through more frequent exposure [4]. Host radiation may also allow a virus to expand into a new ecological niche, although the viruses' ability to infect a diverse range of hosts may decrease due to different replication rates according to host physiology and tissue tropism; altered virulence; and concurrent adaptation to multiple habitats [5].

On a timescale of many millions of years, host jumps and subsequent cross-species transmission events can result in new host-adapted genotypes which can mimic co-divergence since temporal signals may become lost in phylogenetic reconstructions [6]. Estimating such co-divergence defying signals requires co-phylogenetic analyses that can

Edited by Takeshi Noda.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11262-019-01702-x) contains supplementary material, which is available to authorized users.

✉ Shane R. Raidal
shraidal@csu.edu.au

[1] School of Animal and Veterinary Sciences, Charles Sturt University, Boorooma Street, Wagga Wagga, NSW 2650, Australia

[2] School of Biomedical Sciences, Charles Sturt University, Boorooma Street, Wagga Wagga, NSW 2650, Australia

[3] Department of Physiology, Anatomy and Microbiology, School of Life Sciences, La Trobe University, Melbourne 3086, Australia

[4] Healesville Sanctuary, Zoos Victoria, Badger Creek Road, Healesville, VIC 3777, Australia

interrogate the degree of topological congruence between host and virus in which a clear congruence provides strong evidence for co-divergence while incongruence is compatible with cross-species transmission [7].

Following a host-switch any subsequent genetic diversity can arise through a range of ecological and evolutionary mechanisms [8]. Under suitable ecological conditions when multiple host species are equally available, a balance between forces of host co-divergence and host-switching should result from episodes of selection and counter-selection [9, 10]. Viral genotypes from reservoir hosts may require to go through selective 'sieve' to be established into new host population after initial spillover, where adaptive substitutions may occur owing to long-term selection into the new host followed by adaptive phenotypic changes [8, 11]. These adaptive changes often affect the likelihood of successful emergence of a virus into a new host species, therefore, surveillance for adaptive genetic markers are crucial for predicting the risk of disease emergence [8]. Nevertheless, empirical studies have shown that the evolution and adaptation of viruses in fluctuating host environments can lead to fitness improvements for both host and virus with alternating host transmission cycles and not necessarily affecting inter-host fitness [12, 13]. Stable virus-host co-divergence sits in marked contrast to processes of deeper phylogenetically based cross-species transmission that characterizes many emerging infectious viral diseases. The latter is the likely method by which *Beak and feather disease virus* (BFDV), the dominant pathogen of Australasian psittacine birds, has adversely affected several critically endangered psittacine species to the verge of extinction [14, 15].

BFDV is a member of the Circoviridae family and has a relatively simple but compact circular, ambisense single-stranded DNA (ssDNA) genome of approximately 2000 nucleotides which encode a replicase (Rep) and a single-capsid protein (Cap). The Cap is the only exposed protein of the virion that forms the capsid shell in $T=1$ icosahedral symmetry [16]. BFDV Cap is genetically diverse and prone to mutation but relatively antigenically conserved with little evidence of different antigenic serotypes [17]. Whilst BFDV is subject to a high mutation rate, approaching that of RNA viruses, spillovers and genetic recombination occurs frequently amongst contemporary genotypes circulating in psittacine reservoirs [15, 18, 19]. Recent research have highlighted an ancient cretaceous origin of BFDV in Gondwanaland from Psittacopasserae progenitors [20], while the contemporary BFDV lineages circulating within extant psittacine reservoirs act like host generalist [18], although strong host-dependent phylogenetic clades has also been discovered [20, 21].

With multi-host ecological niches, frequent spillover events and evidences of host adaptation make BFDV a suitable model for studying competing forces of co-divergence, host-switches and associated phenotypic plasticity. In this study, we survey the genetic population structure of extant BFDV genotypes in the psittacine reservoir hosts in light of host-virus co-divergence and host-switches. We used homology models based on empirical capsid structure for understanding the phenotypic plasticity of circulating BFDV lineages. We argue that inherent structural constraints of the BFDV capsid shell allow flexible host switches against forces of genetic co-divergence. We also highlight some adaptive markers on the capsid shell that might be useful for monitoring emergence of new infection.

Materials and methods

Compliance with ethical standards

Animal sampling was obtained using guidelines set by the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (1997) and authorised by the Charles Sturt University Animal Care and Ethics Committee (permit 09/046). All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. The authors declare no competing financial or other conflict of interest.

Sampling and extraction of genomic DNA

Samples from different species of wild psittacine birds in close proximity to Healesville sanctuary ($n=97$) were collected to detect BFDV infection in wildlife and to evaluate the divergence of the BFDV population. This included samples obtained from wild sulphur-crested cockatoo (*Cacatua galerita*), musk lorikeet (*Glossopsitta concinna*), rainbow lorikeet (*Trichoglossus haematodus*), little corella (*Cacatua sanguinea*), eastern rosella (*Platycercus eximius*), crimson Rosella (*Platycercus elegans*) and Australian king parrot (*Alisterus scapularis*). Total genomic DNA was extracted from blood or tissue samples using standard protocols [22–24].

The extracted DNA samples were screened for BFDV infection using routine BFDV diagnostic PCR at the Veterinary Diagnostic Laboratory (VDL), Charles Sturt University using established protocols [23] and also using HRM-Cap [25]. From positive samples, full-length BFDV genomes ($n=23$) were amplified using multiple overlapping PCR fragments using previously developed primers and thermal conditions [15, 20, 25] and sequenced by Sanger dideoxy sequencing in Australian Genome Research Faculty (AGRF). To obtain a comprehensive assessment of BFDV population structure across Australia all usable (without ambiguous base call) full-length sequences from previous

studies were compiled and 160 full-length genomes were used for bioinformatics analysis.

Reconstructing phylogeny and median joining network

Individual sequences were annotated with accession number, geographic origin, host species and taxonomic tribe and sampling year. Genomes were aligned in Geneious (Geneious version 9.2, Biomatters, New Zealand) with MAFFT v7.017 using G-INS-i (gap open penalty 1.53; offset value 0.123) alignment algorithm [26]. For the BFDV phylogeny, a general time reversible model with gamma distribution rate variation and a proportion of invariable sites (GTR + I + G4) provided the lowest Akaike information criterion using the programme jModelTest 2.1.3 [27]. Maximum-likelihood (ML) phylogenetic tree of full-length BFDV genomes across Australian landscape ($n = 160$) were estimated using the program PhyML v3.1 [28]. Raven circovirus was used as an out-group for the full-genome phylogenetics reconstruction.

Host-dependent genetic stratification in Australian BFDV population

A series of sequence-based multivariate analyses were performed to investigate stratification in the BFDV population across Australia. The non-hierarchical statistical parsimony network was used to explore genealogical relationships between BFDV haplotypes and their respective host population using TempNet, a freely available R script [29]. An unsupervised Bayesian genetic stratification approach was also used to discriminate the Australian BFDV subpopulation structures to derive an unbiased estimation of probable genetic clustering of BFDV sequences across the Australian landscape. This was achieved by analysing full-length BFDV sequences using STRUCTURE v 2.3.4 [30]. XMFA2Struct (http://www.xavierdidelot.xtreemhost.com/index.htm) was used to convert the genomic sequences to STRUCTURE compatible files. To estimate the number of population structures (the *K* parameter), the BFDV dataset was analysed allowing the value of *K* to vary from 1 to 12 with an initial burn-in of 10,000 iterations followed by 50,000 iterations. Five independent runs were carried out for each *K* value (equating to 60 runs in total). Default parameters and an admixture model with the option of correlated allele frequencies between populations were used since the model could account for BFDV recombination and producing some individuals with mixed ancestry or for allele frequencies in sub-populations being similar due to admixture or shared ancestry. Structure Harvester was used to detect the optimum number of population structures [31] by inferring the appropriate ΔK (highest change of likelihood function). After obtaining the optimum *K* value at least 30 independent reruns in STRUCTURE were performed to obtain the final result. Genetic stratification was visualised using CLUMPAK [32] using a suitable distract setting and default colour parameters. The number of genetic sub-populations thus obtained by unsupervised clustering was subsequently coupled with spatial or host-dependent analysis and subjected for molecular variance analysis (AMOVA) implemented in Arlequin v 3.1 using locus by locus format with 1000 permutations for each [33]. Genetic differentiation parameters such as segregating sites (*S*), haplotype frequency (*H*), haplotype diversity (*Hd*) and nucleotide diversity (Pi) for each putative population were determined using DnaSP v 5. Tajima's *D* neutrality test and was also performed in each population.

Assessing host–virus coevolutionary association

To test the significance of the phylogenetic congruence between psittacine species and BFDV genotypes in Australian landscape, the Procrustes approach to co-phylogeny (PACo) was used [34]. This procedure performs a principal coordinate analysis on the host genetic distance matrix followed by a Procrustes rotation of the BFDV genetic distance matrix, while retaining the information of host-virus association using an empirically observed host–pathogen pairing matrix (H–P link). A sum of squared residuals was then calculated that represents the fitness of host-virus co-phylogeny. The host-BFDV pairing matrix, i.e., which BFDV is paired with which host, was randomised 10,000 times and the sums of square values recalculated. The observed sum of squares value was then compared to the distribution of values from the randomisations to determine the probability of obtaining the observed result under random expectation. In this study, patristic distance of mitochondrial DNA (partial *Cyt-B* and *Cox-1* genes) was used for host matrix while patristic distance of BFDV genotypes were based on whole-genome, *Cap* and *Rep* sequences. Both host and viral genome were carefully selected from each STRUCTURE-inferred genetic sub-populations to obtain maximum coverage. Aside of determining significance of individual host-BFDV association, ParaFit analysis implemented in CopyCat [35] was also used to obtain the global fit statistics between psittacine hosts and BFDV genotypes.

Directional evolution and structural plasticity of BFDV lineages

Directional selection pressure in milieu of structural divergence of BFDV lineages was assessed on capsid protein sequences ($N = 45$, representing all sub-populations) using site-specific detection tools (SLAC, FEL, REL and FUBAR) in datamonkey web server. Positively selected sites detected

by all methods were analysed and compiled together using an integrated multi-program platform IMPACT_S [36]. In addition, the episodic diversifying selection on each branch of the phylogeny was computed by MEME [37] and Branch site RAL (BSR) analysis. To explore the evolutionary trajectories of these selected residues on BFDV structure over the host specific lineages, a capsid protein phylogeny (Alignment Algorithms: MUSCLE, RAxML Model: PROTCA TWAG) was constructed using PhyloBot interactive server [38] to generate ancestral state of capsid protein sequences in every bifurcating node.

Homology modelling and illustration of structural conservation

Homology modelling of representative capsid protein sequences from each putative population structures was conducted using the SWISS-MODEL server [39]. Empirical X-Ray Crystal structure of 60-mer BFDV capsid assembly (PDB id. 5j36) was used as the template model and the best output models were selected according to Global Model Quality Estimation (GMQE) and QMEAN statistical parameters. The resulted models were checked for overall stereochemical quality, including backbone torsional angles through the Ramachandran plot using PROCHECK [40] as implemented in Structural analysis and verification (SAVES) server (https://services.mbi.ucla.edu/SAVES). Lastly, each model structure was minimized using molecular modelling toolkit (MMTK) implemented in UCSF Chimera [41]. To visualise the comparative conservation of the BFDV capsid structure across the different lineages, amino acid sequence from each lineage was aligned using clustal omega [42]. The structural conservation was depicted in Chimera using the reference X-ray crystal model (PDB: 5j36) as pentameric protomer for higher-order structural symmetry. In addition, superposition of nine newly generated homology models representing of genotypic populations were conducted in Chimera. The root mean-squared deviation (R.M.S.D) and conservation scores among the models were calculated based on Needleman–Wunsch algorithm with block substitution matrix BLOSUM62.

Results

Host-dependent genetic structure and spillover infection

The updated maximum-likelihood (ML) phylogenetic tree from all publicly available full-length BFDV genomes in Australian landscape including those generated from this study highlighted a clear pattern of host-dependent clustering with frequent spillover infections amongst the reservoir species (Fig. 1). The BFDV genomes sampled from a single geographical area (Healesville, VIC) did not form any spatially segregated clade, rather they were distributed throughout the phylogenetic tree (shown in blue taxa) and clustered with closely related reservoir clades highlighting co-diverging population. For example, BFDV isolated from sulphur-crested cockatoos were incorporated within the cockatoo clade whereas those from rainbow lorikeets clustered with BFDV-Loriini clade (Fig. 1). The three-dimensional statistical parsimony networks (Fig. 2) also demonstrated shared BFDV haplotypes between reservoir host tribes (amongst Pezoporini, Cacatuidae and Platycercini) as evidence of multiple intrusions of diverse BFDV genotypes as spillover infections.

The unsupervised Bayesian clustering approach revealed nine ($K=9$) statistically supported host family/tribe-dependent BFDV genetic population structures across Australia (Fig. 3). Each individual genome was given a membership probability to the respective population using a distinct colour while multiple colour patterns represented probability of genetic admixture between populations. Based on STRUCTURE membership probability, three distinct populations within cockatoo hosts (Cacatuidae 1, 2 and 3); one Platycercini, one Polytelini, two distinct Pezoporini (Pezoporini-1 and Pezoporini-2), one Melopsittacinae and one Loriini population were identified (Fig. 3). Analysis of molecular variance (AMOVA) which uses genetic differentiation index or fixation index as a predictor of population structure validated the STRUCTURE-defined populations demonstrating positive values for all fixation indices (F_{CT}, F_{SC} and F_{ST}) with high statistical significance (Table S1). Hence, the overall F_{ST} value within populations was 0.27 demonstrating frequent exchange of genetic material or gene flow among the populations resulting from spillover and host switch events. Molecular diversity and genetic parameters were also calculated in each of the STRUCTURE-defined populations (Table S2).

Co-evolutionary trajectory of BFDV lineages in reservoir hosts

Since the population genetic analyses pointed towards host-adapted clustering of BFDV across Australia, the coevolutionary trajectory of BFDV genotypes in respective host species was assessed. The patristic distance among host mtDNA sequence data and BFDV genome data, respectively, sampled from each host were used to calculate contributions of individual host–virus links to the Procrustean fit and visualised these in Jacknifed-squared residuals plots (Fig. 4). The bar plots of squared residuals indicated that several links of psittacine reservoirs and BFDV genotypes were relatively low (for example in lorikeets and budgerigars) and thus representative for strong host–virus

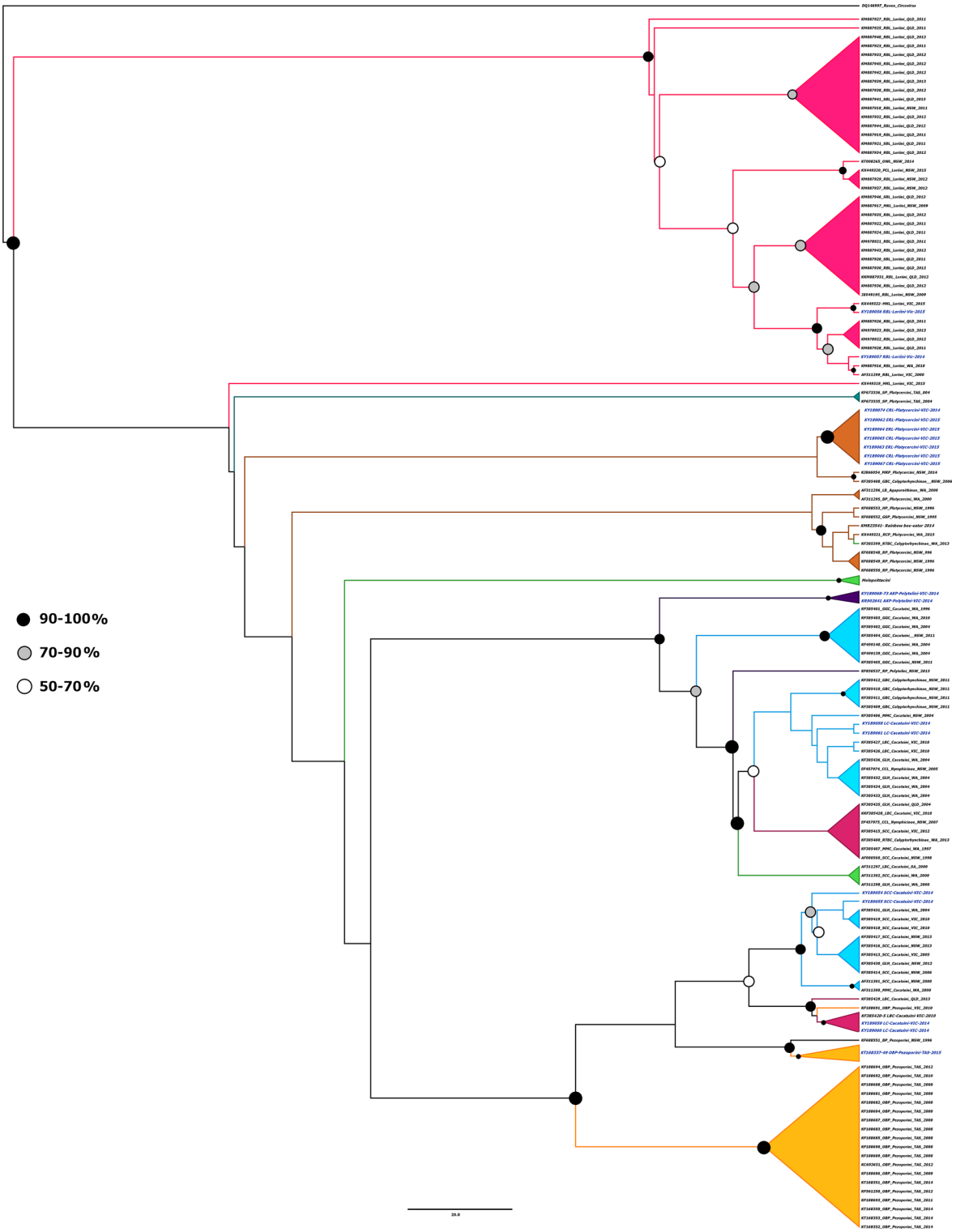

Fig. 1 Maximum-likelihood (ML) tree generated from beak and feather disease virus (BFDV) whole-genomes ($n = 160$) from Australian birds rooted to raven circovirus. Different clade colours highlight different host taxonomy. Labels at branch tips refer to GenBank accession numbers with species abbreviation state of origin, Victoria (VIC), New South Wales (NSW), Tasmania (TAS), Queensland (QLD) and Western Australia (WA) and year of sampling. Blue text highlights newly sequenced full BFDV genomes from this study. Percentage bootstrap values are highlighted as shaded circles

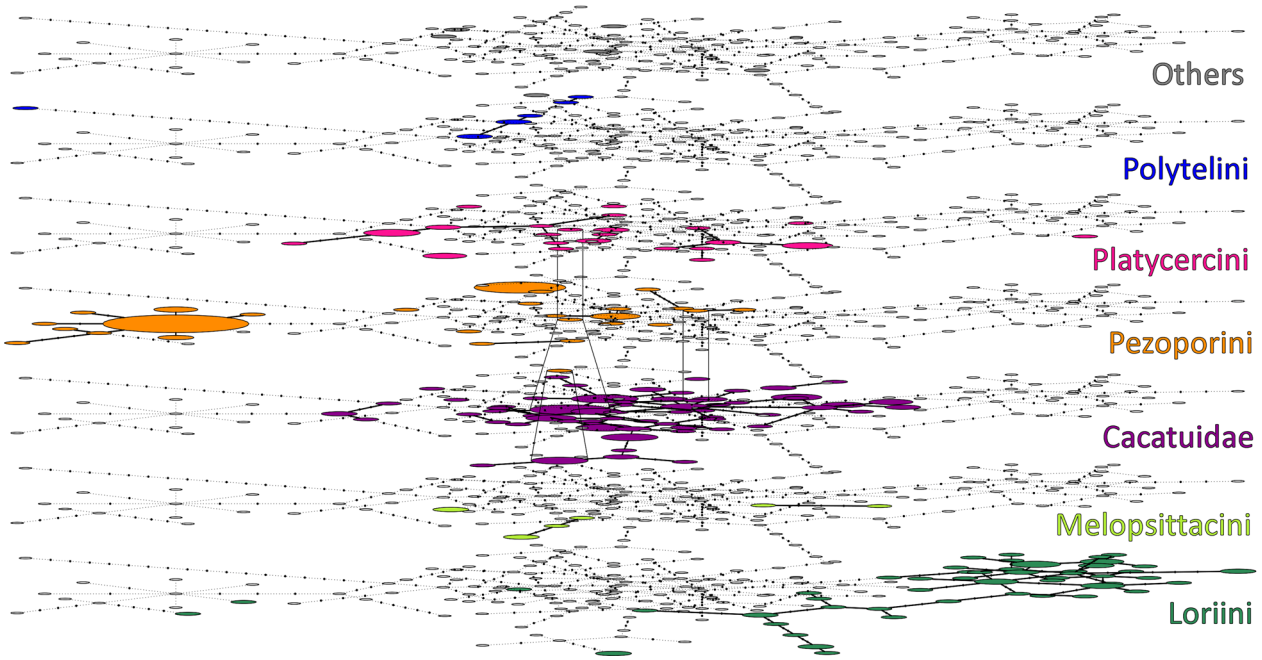

Fig. 2 Relationship between BFDV Rep sequences grouped into host-dependent populations illustrated using a three-dimensional statistical parsimony network. Unique haplotypes are represented by circles, colours indicate the presence of haplotypes in a sample from a particular host population. While size of the circle represents the haplotype frequency. Within each layer, haplotypes are connected by a horizontal line if they are separated by one mutation, each additional mutation is indicated by small black dot. Shared haplotypes between adjacent layers are joined by vertical lines

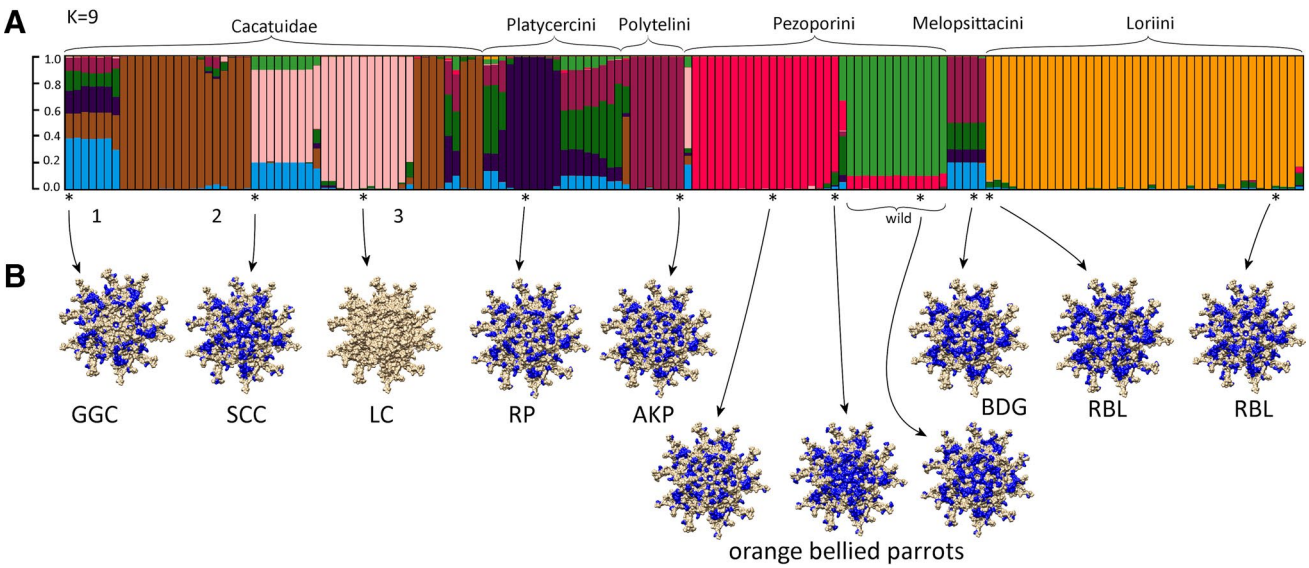

Fig. 3 Inference of BFDV population structure by Bayesian clustering methods using complete genome sequences ($n = 160$). STRUCTURE analysis inferred nine genetic populations ($K = 9$) where Individual genomes are shown by a vertical bar plot and colour coded according to the membership probability (percentage) to a particular genetic sub-population. Multiple colours within individual bars are indicative of admixture of genetic material between populations. **a** Genomes grouped using host variables (host taxonomic family/tribe) showing clear evidence of host-dependent genetic population structure. Asterisks represents individual genomes from gang gang cockatoo (GGC), sulphur-crested cockatoo (SCC), little corella (LC), twenty-eight parrot (RP), Australian king parrot (AKP), budgerigar (BDG) and rainbow lorikeet (RBL) resampled from each structured population for coevolutionary hypothesis testing. **b** Individual figures beneath the selected genomes (arrows) show amino acid diversity (blue shading) and conservation (grey shading) of respective capsid proteins against empirical crystallographic structure (PBD ID: 5j36). Note the structural plasticity of BFDV in a single host species (orange-bellied parrots) indicating three separate spillover infections from diverse genotypes

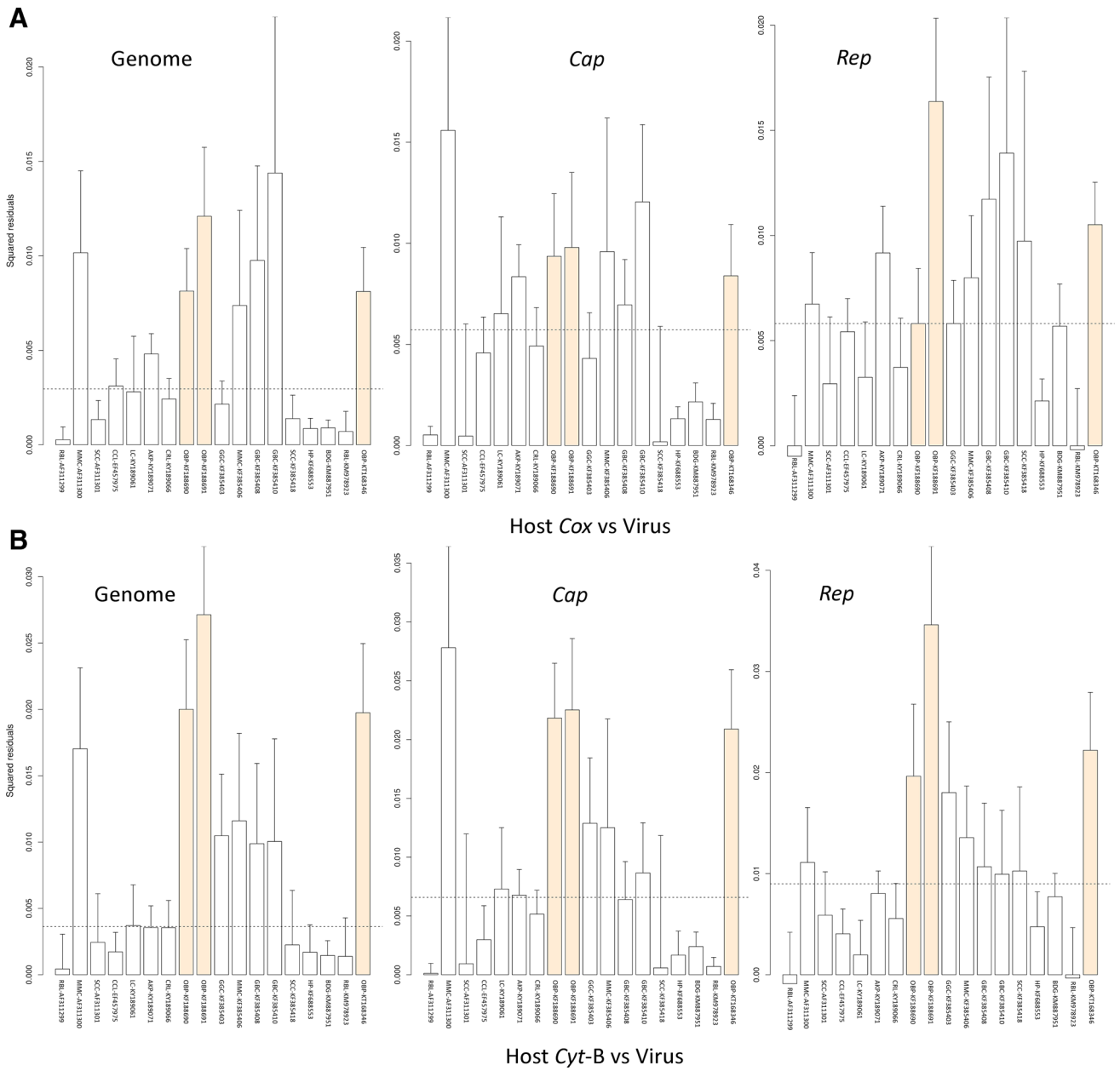

Fig. 4 Psittacine hosts and BFDV co-phylogeny analysis; contributions of individual host-parasite links to the Procrustean fit. Host mtDNA (*Cox*-1 and *Cyt*-B) and complete genome, *Cap* and *Rep* genes were used to test the host–virus coevolution. Jacknifed-squared residuals (bars) shows individual host–virus link with upper 95% confidence intervals (error bars) by applying PACo to patristic distances. Orange-bellied parrots (host) and BFDV associations are highlighted as shaded bars. To ease comparisons the median-squared residual value is shown (dashed line). Squared residuals below the median value indicates significant co-divergence, whereas those above the median indicate likely host jumps or spill over

coevolutionary associations. However, it is notable that, residuals from several other host-virus links, particularly those with orange-bellied parrot and cockatoos, were far above the median-squared residual value representing co-phylogenetic incongruence or dispersion from coevolutionary trajectory. AxParafit results for overall co-phylogenetic structures were mostly non-significant ($p > 0.02$). The value of significance in BFDV full genome, *Cap* and *Rep* against host mitochondrial *Cox-1* gene was 0.01, 0.03 and 0.04, respectively, whereas against host *Cyto-B* gene was 0.03, 0.03 and 0.04, respectively. This approach demonstrated that although some BFDV lineages might have enjoyed undisrupted co-evolution with their respective psittacine hosts to some extent, most lineages showed evidence of relatively chaotic evolutionary history resultant from frequent cross-species transmission and host jumps.

Surface plasticity and structural constraints of BFDV Cap

To better understand the phenotypic traits of BFDV in host-adapted populations and to recognise the plausible cause of host flexibility and frequent interspecies transmissions, the structural conservation of the Cap protein were analysed. Overall, different host-adapted BFDV population demonstrated extensive diversity on the surface residues of capsid pentameric protomers in both 60 mer and 10 mer assembly (Figs. 3 and 5), also watch as a movie URL; https://youtu.be/6R2Sr0WatVY The degree of diversity was depicted by surface view using gradients of blue where the bright blue and light blue illustrated maximum and minimum diversity, respectively; the conserved residues were represented by pale grey. Most striking differences were found around in the surface residues around the central pore region of the five-fold axis symmetry (Fig. 5a).

Several residues that mediated intra-pentamer (positions 119, 120, 121, 123, 124, 157 and 162) and inter-pentamer interfaces (positions 177, 179, 182, 207, 212 and 214) were highly diverse from different hosts while residues on the inner surface of the capsid were relatively more conserved (Fig. 5 and Table S3). Amino acid diversity on the outer surface of the capsid was not confined or exclusive to the protruding loop region, as degeneracy was also detected in

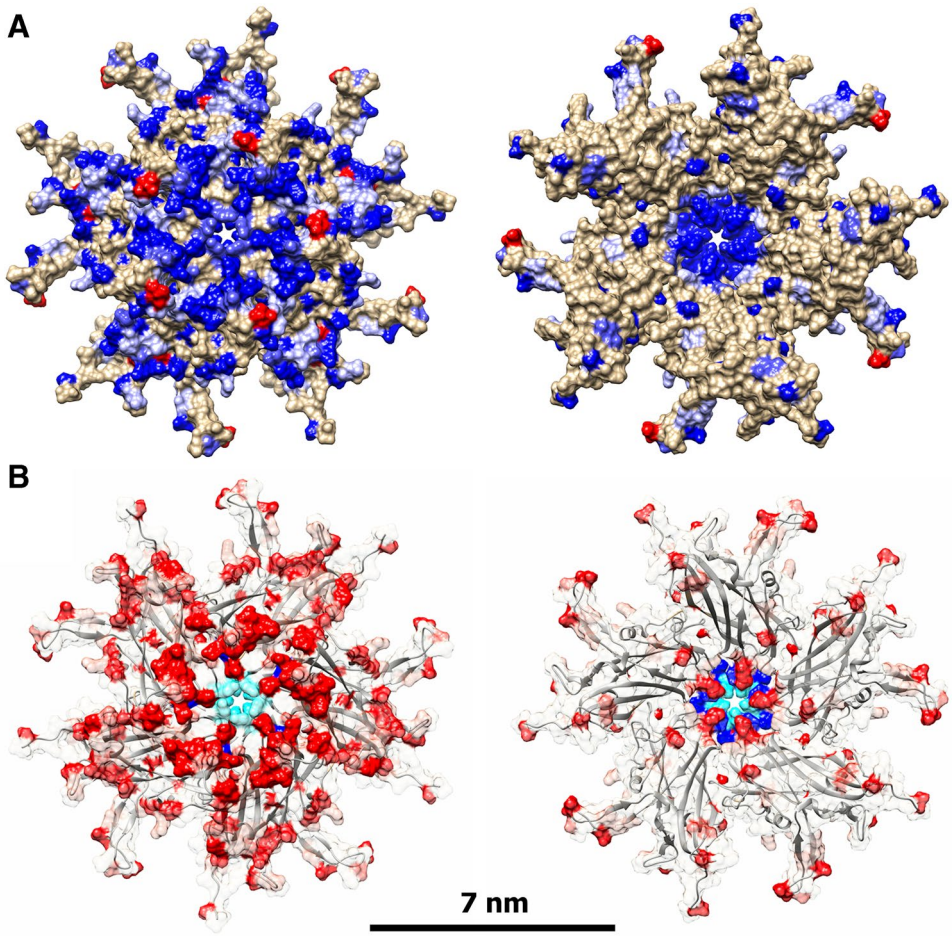

Fig. 5 Structural conservation of BFDV capsid protein across host-adapted populations depicted as capsid pentameric protomers obtained from empirical 60-mer assemblies of a BFDV crystallography structure (PDB ID: 5j36). Each model is made of five identical and interlocking capsid proteins centred on a star-shaped central pore. **a** The outer surface of the capsid fragment is shown on the left and the internal surface on the right. Light brown surfaces show conserved amino acid (aa) residues retained in all BFDV genotypes with blue-shaded surfaces showing the degree of aa diversity among BFDV clades with darker blue sites having the greatest aa diversity. The external surface on the left displays much greater diversity than the internal surface on the right. Bright red sites indicate those evolving under positive selection pressure. Conservation scores were calculated using clustal omega matrix. **b** Transparent models of the outer and internal surfaces as above showing the binding interfaces between individual capsid proteins. The diversity within each pentameric morphological units (involving all five subunits) are shown as shades of blue, whereas inter-pentameric interfaces (involving neighbouring subunits) are shown as shades of red. Intensity of colour refers to the conservation score among the 11 models shown in Fig. 3. Strands that comprise β-sheets are shown as grey ribbons with directional arrows deep within each capsid protein

the β-sheets which form the core of capsomere (Fig. 5b and Table S3). Although evolutionary fingerprinting of *Cap* protein sequences (data not shown) pointed towards an overall trend of purifying selection, consensus of codon-based site-specific likelihood methods identified some particular sites (positions 71, 84, 182 and 211) consistently evolving under directional positive selection (Fig. 6 and Table S4). These particular sites were examined in the BFDV *Cap* phylogeny at host adapted clades in response to their radiation from ancestral nodes (Fig. 7). The capsomere structure for each BFDV population were also modelled and compared for any evidence of structural incongruence. As shown in Fig. 7, the BFDV lineage from 2015 outbreak in wild orange-bellied parrots (Pezoporini-2) involved a directional mutation from threonine (T) to proline (P) at position 71, radiating from an ancestor shared with Cacatuidae-2.

These directional mutations were represented as red ball and stick (also circled red) in the ribbon view in their respective models. It is notable that the residues evolving under positive selection were located on the loop regions protruding externally on the capsid surface and therefore possibly involved with extensive host receptor–virus interactions. Other mutations (depicted blue) were most likely transient mutations and under ongoing purging process by purifying selection. The phylogeny could not resolve the closest relative of BFDV in captive orange-bellied parrot population as the posterior probability support at this node was quite low (> 60%), however, the mutations from an ancestral node were detected in loop regions as well as the β-sheets of capsomere structure. Similar to the complete genome phylogeny, spillover events between populations were also detected in the structural phylogeny as a Cap phenotype from host Polytelini (KF850537) was positioned in a monophyletic clade of Cacatuidae-3. Similarly, close phenotypic relationship between BFDV from Platycercini hosts (KY189066-67) and Pezoporini-1 was detected. Despite large variation in surface residues BFDV capsid proteins from different host adapted populations demonstrated structural robustness as

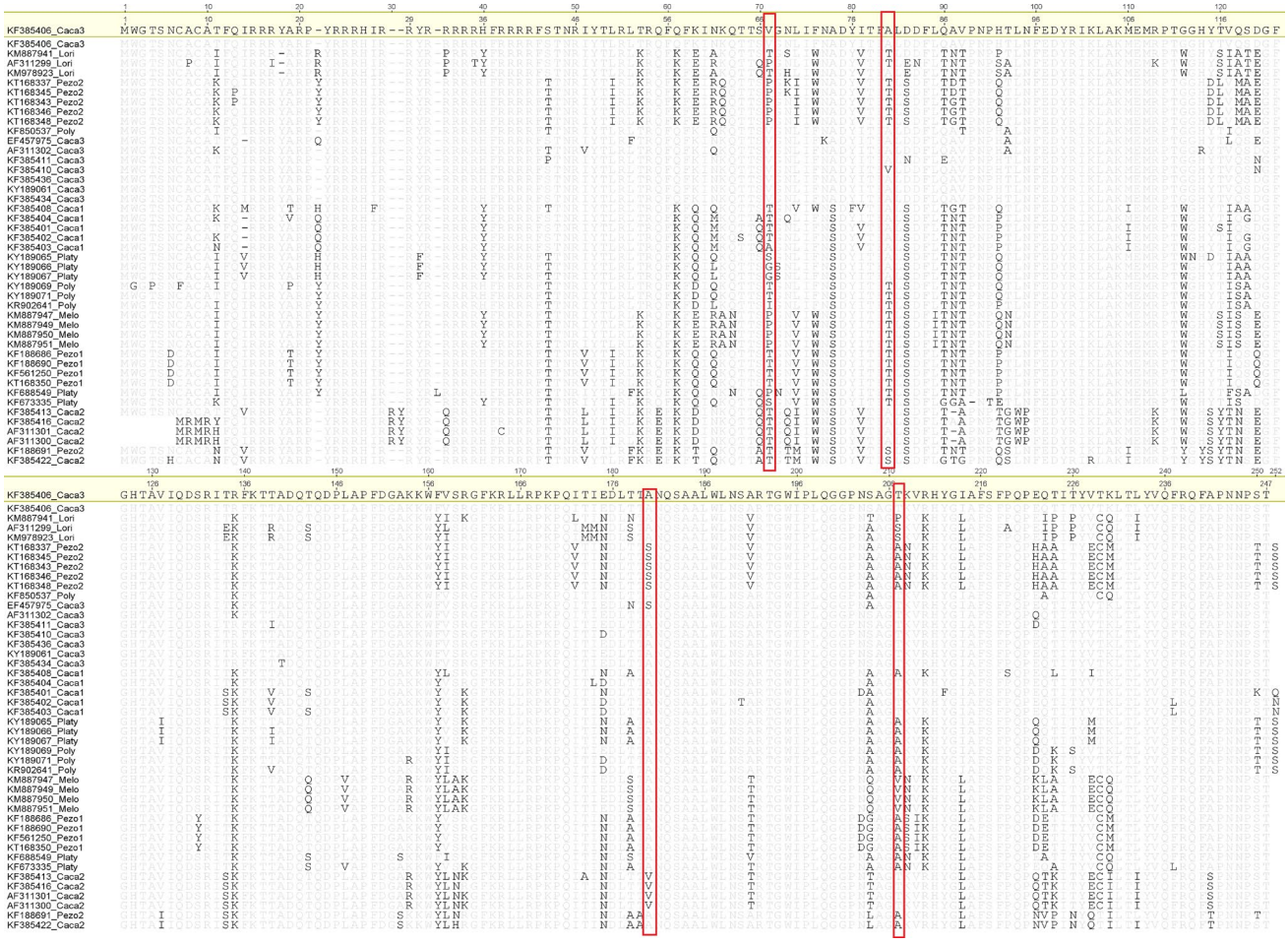

Fig. 6 BFDV capsid protein alignment including representative number of sequences from all nine STRUCTURE-inferred host-adapted BFDV populations. Consensus of site-specific selection pressure detection tools (SLAC, FEL, REL, FUBER) identified residues at alignment position 71, 84, 182 and 211 (red boxed) are evolving under positive selection pressure

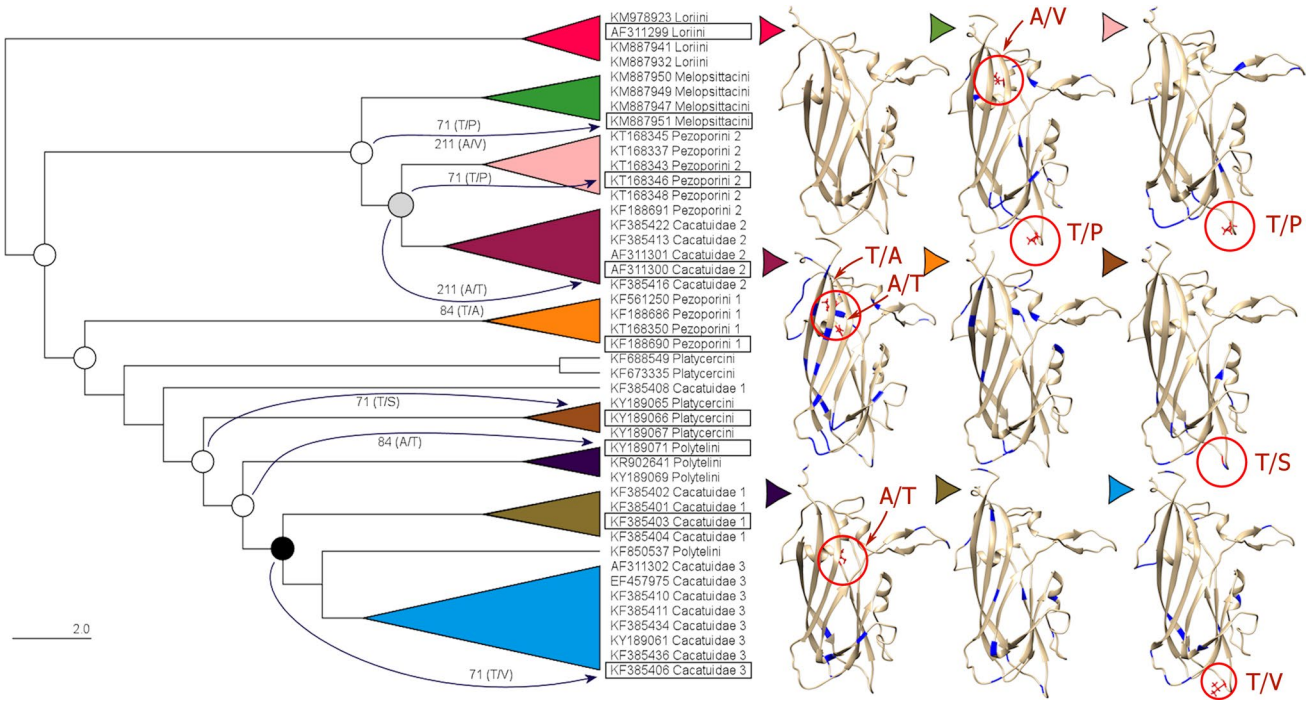

Fig. 7 Evolutionary trajectory of capsid protein and structural plasticity of host-adapted BFDV populations. Representative capsid protein sequences (at least four) from each population were used for reconstructing phylogeny by MrBayes (rate matrix = Wag, gamma category = 4 with 10,00000 MCMC chain length). As for Figs. 1 and 2, each coloured clade represent a particular host-adapted population (STRUCTURE inferred) and clade Loriini used as out-group. Posterior probability value presented by node circles (dark circle > 90%, light grey circle > 80%, empty circle > 60%). Directional mutations from most recent common ancestors (MRCA) to the extant populations at positively selected sites are indicated using alignment position and abbreviation of mutated residue (on each arrowhead). Homology models from each population (colour coded triangle) shows the amino acid diversity radiation from the ancestral state while the positively selected residues shown by ball and sticks view in red (and circled red). For example, radiation of BFDV KM887951 in a budgerigar (*Melopsittacus*) involved two directional mutations shown at sites 71 and 211 each with different aa mutations T/P (threonine to proline) and A/V (alanine to valine), respectively, from the most recent progenitor. Residues undergoing transient mutations are also shown as blue segments

the average root mean-standard deviation (R.M.S.D) value was 0.08 Å (ranging from 0.05 to 0.14 Å between any two models) reflecting high structural similarity without any steric clashes (Table S5).

Discussion

This study utilized all available BFDV sequence data sampled over two decades across Australia and incorporated novel whole-genome sequences from hitherto unrepresented reservoir hosts to ascertain the background forces for infection and genetic distribution of BFDV. The results uncovered strong host-based BFDV stratification across the Australian landscape highlighting primordial processes of cryptic host adaptation and competing forces of host co-divergence and cross-species transmission while phenotypic or structural plasticity of the virus might explain the high degree of host-switch flexibility against a background of reservoir host-convergence.

Numerous interacting strains [5] may dominate natural populations of pathogens which may impact on host-pathogen dynamics and the evolution of virulence. Geographic considerations need to be considered as well since recent work has shown that spatial structure can influence the shape of host-pathogen dynamics and transmission-virulence trade-offs [43]. Recent phylogeographic studies have highlighted the Gondwanan origin of BFDV [20, 44] with an association in parrots and cockatoos that is likely to have existed for at least 10 million years [20]. This is sufficient time to allow the development of host-based stratification despite BFDV retaining an ability to jump host species. The co-divergence defying signature shown in multiple host-based BFDV populations (Fig. 4) is highlighted as a marked incongruence generated between virus and host phylogeny as a result of spillover events and host jumps. Such intrusions may create epidemic expansion of certain viral genotypes such as in the critically endangered orange-bellied parrots (as shown in Table S2) and are often associated with increased virulence in naïve populations [15].

Understanding the speed by which viruses evolve and how evolutionary rates might differ among viruses or fluctuate through time has been difficult to accurately determine because high-contemporaneous mutation rates create an illusion of fast evolution [45]. Great discrepancies have been recently shown for evolutionary rates of extant DNA and RNA viruses; however, advances in analytic methods that include a variety of host biology and paleobiological factors showed that the cumulative evolution of viruses more likely extends over millions of years [20]. The recent demonstration of sympatric sequestration of BFDV in lorikeet hosts alongside the paleovirological and genetic recombination events points to an ancient Cretaceous origin of BFDV in Australia [24]. This high mutation rate of BFDV maximizes genetic diversity which undoubtedly broadens the capacity to opportunistically change hosts [15, 18, 46]. In fact, psittacine beak and feather disease is well recognised across Australia where it occurs in relatively high prevalence among the more common cockatoo, lorikeet and parrot species, leading to high rates of endemicity [47].

Like many other ssDNA viruses, BFDV has a broad available sequence space among potential host populations (Table S2) which is not detrimental to capsid structure. Unsurprisingly, the outer most surface of the virus demonstrated highest diversity along with inter-pentameric interfaces [48, 49]. The outer most loops on the capsid surface demonstrated maximum plasticity even between the host-adapted lineages and their putative common ancestors (Fig. 5). This suggests either for a deep sympatric speciation or existence of undiscovered intermediate host in the evolutionary trajectory of BFDV. In fact, a recent study demonstrated high prevalence of BFDV in non-psittacine birds which could expand the horizon of known reservoir populations [50].

Despite high-surface degeneracy, capsid protein models from different BFDV populations superimposed perfectly with low average root-mean-square deviation (RMSD) suggesting for a robust structural constraint on different genotypes (Table S5). This highlights the plasticity of BFDV capsids, flexible enough to adapt to surface mutations exerted by adaptive drivers during host jumps or sympatric speciation without being altered by minor rearrangement or subtle conformational changes. Phenotypic innovation among circoviruses is likely heavily constrained due to threshold structural limitations for encapsidation and the multiple functions of Cap in attachment, uncoating, nuclear localization and compartment shuttling, then DNA binding in virus assembly [16].

Conventional wisdom dictates that changes in the Cap are more likely to reflect host–viral interactions and immunological selective pressure with recent research demonstrating the Cap gene evolving more rapidly than Rep [15, 44]. While there is a high degree (76–83%) of amino acid diversity within the BFDV Cap there is little serological evidence that this is driven by immune evasion given the lack of evidence of different serotypes even between sympatrically segregated BFDV lineage in lorikeets [17, 20, 51, 52]. Nevertheless, the extensive diversity in surface architecture of BFDV capsids may allow subtle difference in antibody affinity since only a few epitopes are important for antibody recognition as shown in porcine circovirus 2 capsids [53].

Adaptation processes dictate whether a virus will be able to successfully evolve in a new host range or fadeout stochastically. Genetic markers of adaptive drivers often emerge following host jumps or cross-species transmission [8]. As the results presented here have shown that the BFDV genetic population across the Australian landscape was structured by host adaptation (Figs. 1, 2, 3, 4, 5, Table S1), such adaptive genetic changes can originate either in the new host or in the reservoir hosts associated with phenotypic traits such as receptor binding, tissue tropism, antigenicity and or virulence [8]. Surveillance for genetic markers for adaptation could help to predict the risk of disease emergence in a multi-host-pathogen such as BFDV. The positively selected amino acid residues on the capsid surface are plausible candidates for adaptive genetic markers of BFDV (Table S4, Figs. 6, 7) as they have evolved directionally and fixed into host-adapted lineages from their most recent common ancestors. Site-directed mutagenesis of these predicted residues in cell culture or animal models experiments could unveil exciting new insights for receptor attachment, tissue tropism and in general pathogenesis of BFDV in the near future.

Our BFDV population genetic analyses has demonstrated that BFDV populations in parrots and cockatoos are highly admixed with signatures of cross-species transmission and genome-wide recombination (Figs. 2, 3). When a population replicates in a heterogeneous environment, genotypic variation arises through mutation and recombination, and the fate of this variation depends on the natural selection and random drifts where different outcomes are plausible [54, 55]. The population may either become adapted to a specific niche with fitness landscape similar to those in homogeneous environments, or the population may replicate across the heterogeneous environment and adapt to all the niches within the environment, rather than adapting to a single niche and thus become a host generalist [54, 56, 57]. In this case, the population may experience cost of generalisation, i.e. suboptimal adaptation to each niche within the environment. The population is able to survive in each niche, but cannot reach the fitness level achieved by a specialist for that niche [57]. The level of host adaptation is influenced by the degree of variation amongst niches and the amount of time spent in specific niches [58]. The host generalist BFDV population could be stratified into host tribe specific minor sub-populations with significant fixed polymorphisms for each population. However, these populations also possessed signature of frequent

genetic admixture as a result of genome-wide inter-population recombination following cross-species transmission or spillover events (Figs. 2 and 3). Therefore, it is conclusive that the extant BFDV population has inherent tendency towards attaining host specificity with narrower niches, but because of the evolutionary memory exerted by capsid structural constraint and the varying fitness, multi-host niches can be explored in the form of cross-species transmission when environmental condition suits. It is therefore safe to predict that novel BFDV infection in a particular psittacine host will more likely to be transmitted from closely related reservoir species.

Acknowledgements The authors would like to thank the contributions of Peter Copley, Sheryl Hamilton, Jocelyn Hockley, Kristy Penrose, Judy Clark and Annika Everaardt of the Orange-bellied Parrot Recovery Team, and the staff of the Australian Wildlife Health Centre (Healesville Sanctuary). SD received Charles Sturt University writing up award for completion of this manuscript.

Author contributions The submitting author confirms that all individual co-authors have met the criteria of authorship. SD, SS, SRR, AP, JKF contributed to the conception and design. KA, PE contributed to sample collection and data analysis. SD, SS, SRR, AP, SAG contributed to method development, data analysis and interpretation. SD, SRR, AP, drafted the manuscript. SR and AP critically revised the manuscript. SRR and JKF gave final approval and agree to be accountable for all aspects of work ensuring accuracy and integrity.

Compliance with ethical standards

Conflict of interest The authors declare no competing financial or other conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. Animal sampling was obtained using guidelines set by the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (1997) and authorised by the Charles Sturt University Animal Care and Ethics Committee (permit 09/046). All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. This article does not contain any studies with human participants performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Holmes EC, Zhang YZ (2015) The evolution and emergence of hantaviruses. Curr Opin Virol 10:27–33
2. Geoghegan JL, Duchene S, Holmes EC (2017) Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog 13:e1006215
3. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P (2008) Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 72:457–470
4. Murthy S, Couacy-Hymann E, Metzger S, Nowak K, De Nys H, Boesch C, Wittig R, Jarvis MA, Leendertz FH, Ehlers B (2013) Absence of frequent herpesvirus transmission in a nonhuman primate predator-prey system in the wild. J Virol 87:10651–10659
5. Redman EM, Wilson K, Cory JS (2016) Trade-offs and mixed infections in an obligate-killing insect pathogen. J Anim Ecol 85:1200–1209
6. Sullivan J, Joyce P (2005) Model selection in phylogenetics. Annu Rev Ecol Evol Syst 36:445–466
7. Page RDM (2004) In: Page RDM (ed) 2003: tangled trees. phylogeny, cospeciation and coevolution. The University of Chicago Press, Chicago. Paperback: $28. ISBN 0-226-64467-7. Ethology 110, 577–578
8. Pepin KM, Lass S, Pulliam JR, Read AF, Lloyd-Smith JO (2010) Identifying genetic markers of adaptation for surveillance of viral host jumps. Nat Rev Microbiol 8:802–813
9. Dennehy JJ, Friedenberg NA, Yang YW, Turner PE (2007) Virus population extinction via ecological traps. Ecol Lett 10:230–240
10. Shin J, MacCarthy T (2016) Potential for evolution of complex defense strategies in a multi-scale model of virus-host coevolution. BMC Evolut Biol 16:233
11. Abedon ST (1999) Bacteriophage T4 resistance to lysis-inhibition collapse. Genet Res 74:1–11
12. Turner PE, Elena SF (2000) Cost of host radiation in an RNA virus. Genetics 156:1465–1470
13. Weaver SC, Brault AC, Kang W, Holland JJ (1999) Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. J Virol 73:4316–4326
14. Peters A, Patterson EI, Baker BG, Holdsworth M, Sarker S, Ghorashi SA, Raidal SR (2014) Evidence of psittacine *beak and feather disease virus* spillover into wild critically endangered orange-bellied parrots (*Neophema chrysogaster*). J Wildl Dis 50:288–296
15. Sarker S, Patterson EI, Peters A, Baker BG, Forwood JK, Ghorashi SA, Holdsworth M, Baker R, Murray N, Raidal SR (2014) Mutability dynamics of an emergent single stranded DNA virus in a naïve host. PLoS ONE 9:e85370
16. Sarker S, Terrón MC, Khandokar Y, Aragão D, Hardy JM, Radjainia M, Jiménez-Zaragoza M, de Pablo PJ, Coulibaly F, Luque D, Raidal SR, Forwood JK (2016) Structural insights into the assembly and regulation of distinct viral capsid complexes. Nat Commun 7:13014
17. Khalesi B, Bonne N, Stewart M, Sharp M, Raidal SR (2005) A comparison of haemagglutination, haemagglutination inhibition and PCR for the detection of psittacine beak and feather disease virus infection and a comparison of isolates obtained from loriids. J Gen Virol 86:3039–3046
18. Sarker S, Ghorashi SA, Forwood JK, Bent JS, Peters A, Raidal SR (2014) Phylogeny of beak and feather disease virus in cockatoos demonstrates host generalism and multiple-variant infections within *Psittaciformes*. Virology 460–461:72–82
19. Sarker S, Forwood JK, Ghorashi SA, Peters A, Raidal SR (2015) Beak and feather disease virus genotypes in Australian parrots reveal flexible host-switching. Aust Vet J 93:471–475
20. Das S, Sarker S, Peters A, Ghorashi SA, Phalen D, Forwood JK, Raidal SR (2016) Evolution of circoviruses in lorikeets lags behind its hosts. Mol Phylogenet Evol 100:281–291
21. Varsani A, Regnard GL, Bragg R, Hitzeroth II, Rybicki EP (2011) Global genetic diversity and geographical and host-species distribution of beak and feather disease virus isolates. J Gen Virol 92:752–767
22. Bonne N, Clark P, Shearer P, Raidal SR (2008) Elimination of false-positive polymerase chain reaction results resulting from hole punch carryover contamination. J Vet Diagn Investig 20:60–63

23. 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
24. Das S, Subir S, Adriaanse K, Forwood JK, Ghorashi SA, Raidal SR (2016) Characterization of beak and feather disease virus genomes from wild musk lorikeets (*Glossopsitta concinna*). Genome Announc 4:e01107–e01116
25. Das S, Sarker S, Ghorashi SA, Forwood JK, Raidal SR (2016) A comparison of PCR assays for beak and feather disease virus and high resolution melt (HRM) curve analysis of replicase associated protein and capsid genes. J Virol Methods 237:47–57
26. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066
27. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772
28. Guindon S, Gascuel O (2003) A Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
29. Prost S, Anderson CNK (2011) TempNet: a method to display statistical parsimony networks for heterochronous DNA sequence data. Methods Ecol Evol 2:663–667
30. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
31. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361
32. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191
33. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolut Bioinform Online 1:47–50
34. Balbuena JA, Míguez-Lozano R, Blasco-Costa I (2013) PACo: a novel procrustes application to cophylogenetic analysis. PLoS ONE 8:e61048
35. Meier-Kolthoff JP, Auch AF, Huson DH, Göker M (2007) CopyCat: cophylogenetic analysis tool. Bioinformatics 23:898–900
36. Maldonado E, Sunagar K, Almeida D, Vasconcelos V, Antunes A (2014) IMPACT_S: integrated multiprogram platform to analyze and combine tests of selection. PLoS ONE 9:e96243
37. Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. Nucleic Acids Res 43:W39–W49
38. Hanson-Smith V, Johnson A (2016) PhyloBot: a web portal for automated phylogenetics, ancestral sequence reconstruction, and exploration of mutational trajectories. PLoS Comput Biol 12:e1004976
39. Bienert S, Waterhouse A, de Beer Tjaart A P, Tauriello G, Studer G, Bordoli L, Schwede T (2017) The SWISS-MODEL repository—new features and functionality. Nucleic Acids Res 45:D313–D319
40. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
42. Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105–116
43. Messinger SM, Ostling A (2009) The consequences of spatial structure for the evolution of pathogen transmission rate and virulence. Am Nat 174:441–454
44. Harkins GW, Martin DP, Christoffels A, Varsani A (2014) Towards inferring the global movement of beak and feather disease virus. Virology 450:24–33
45. Aiewsakun P, Katzourakis A (2016) Time-dependent rate phenomenon in viruses. J Virol 90:7184–7195
46. Raidal SR, Sarker S, Peters A (2015) Review of psittacine beak and feather disease and its effect on Australian endangered species. Aust Vet J 93:466–470
47. Raidal SR (2016) In: Speer BL (ed) Current therapy in avian medicine and surgery. Elsevier, St. Louis, pp 51–58
48. Bahadur RP, Janin J (2008) Residue conservation in viral capsid assembly. Proteins Struct Funct Bioinform 71:407–414
49. Mintseris J, Weng Z (2005) Structure, function, and evolution of transient and obligate protein–protein interactions. Proc Natl Acad Sci USA 102:10930–10935
50. Amery-Gale J, Marenda MS, Owens J, Eden PA, Browning GF, Devlin JM (2017) A high prevalence of beak and feather disease virus in non-psittacine Australian birds. J Med Microbiol 66:1005–1013
51. Bonne N, Shearer P, Sharp M, Clark P, Raidal SR (2009) Assessment of recombinant beak and feather disease virus capsid protein as a vaccine for psittacine beak and feather disease. J Gen Virol 90:640–647
52. Raidal SR, McElnea CL, Cross GM (1993) Seroprevalence of psittacine beak and feather disease in wild psittacine birds in New South Wales. Aust Vet J 70:137–139
53. Lekcharoensuk P, Morozov I, Paul PS, Thangthumniyom N, Wajjawalku W, Meng XJ (2004) Epitope mapping of the major capsid protein of type 2 porcine circovirus (PCV2) by using chimeric PCV1 and PCV2. J Virol 78:8135–8145
54. Smith-Tsurkan SD, Wilke CO, Novella IS (2010) Incongruent fitness landscapes, not tradeoffs, dominate the adaptation of vesicular stomatitis virus to novel host types. J Gen Virol 91:1484–1493
55. Clarke DK, Duarte EA, Elena SF, Moya A, Domingo E, Holland J (1994) The red queen reigns in the kingdom of RNA viruses. Proc Natl Acad Sci USA 91:4821–4824
56. Kassen R, Rainey PB (2004) The ecology and genetics of microbial diversity. Annu Rev Microbiol 58:207–231
57. Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. J Evol Biol 15:173–190
58. Whitlock MC (1992) Temporal fluctuation in demographic parameters and the genetic variance among populations. Evolution 46:608–615