Piasecki, T. et al. (2013). Avihepadnavirus diversity in parrots is comparable to that found amongst all other avian species. Virology, 438: 98-105

https://dx.doi.org/10.1016/j.virol.2013.01.009

Avihepadnavirus diversity in parrots is comparable to that found amongst all other avian species

Tomasz Piasecki, Gordon W. Harkins, Klaudia Chrzastek, Laurel Julian, Darren P. Martin and Arvind Varsani

Abstract:
Avihepadnaviruses have previously been isolated from various species of duck, goose, stork, heron and crane. Recently the first parrot avihepadnavirus was isolated from a Ring-necked Parakeet in Poland. In this study, 41 psittacine liver samples archived in Poland over the last nine years were tested for presence of Parrot hepatitis B virus (PHBV). We cloned and sequenced PHBV isolates from 18 birds including a Crimson Rosella, an African grey parrot and sixteen Ring-necked Parakeets. PHBV isolates display a degree of diversity (>78% genome wide pairwise identity) that is comparable to that found amongst all other avihepadnaviruses (> 79% genome wide pairwise identity). The PHBV viruses can be subdivided into seven genetically distinct groups (tentatively named A-G) of which the two isolated of PHBV-G are the most divergent sharing ~ 79% genome wide pairwise identity with all their PHBVs. All PHBV isolates display classical avihepadnavirus genome architecture.

Introduction

The family *Hepadnaviridae* contains enveloped DNA viruses with partially double-stranded relaxed circular DNA (rcDNA) genomes of approximately 3 kb. These viruses have an unusual replication strategy involving reverse transcription of an RNA intermediate (Seeger and Mason, 2000). There are two currently defined hepadnavirus genera: The genus Orthohepadnavirus, which contains species that infect mammals, and the genus Avihepadnavirus, which contains species that infect birds. Although species in these two genera share very little sequence similarity they have similar genome organisations, with species in both groups containing three homologous open reading frames (ORFs), encoding surface proteins (preS/S), the nucleocapsid protein and the e-antigen (pre C/C), and the viral polymerase (P) (Seeger and Mason, 2000). A fourth ORF, encoding a protein called X, is found in all orthohepadnaviruses, but is apparently only present in some avihepadnaviruses (Chang et al., 2001; Guo et al., 2005). The exact role of the X protein during a natural infection in mammals is unknown. Although not for *in vitro* replication (Blum et al., 1992) the X protein is apparently required for *in vivo* replication (Zoulim et al., 1994). While both orthohepadnaviruses and avihepadnaviruses are hepatotropic and can cause

transient or chronic liver infections, only orthohepadnaviruses are known to cause liver disease and cancer (Funk et al., 2007).

As with orthohepadnaviruses, avihepadnaviruses have very narrow host ranges. It is however, likely that host range switches do occur in that the relatedness of viruses infecting different hosts is not always reflective of the relationships of those hosts (Funk et al., 2007). Host specificity is thought to be determined by the preS domain of the L-envelope protein (Ishikawa and Ganem, 1995). Evidence of recombination amongst Duck hepatitis B virus (DHBV) isolates (Liu et al., 2010; Piasecki et al., 2012) suggests the possibility that strains with either expanded or larger host ranges might arise through recombinational transfers of this genome region between distantly related viruses.

To date avihepadnaviruses have been found infecting various species of duck (Duck hepatitis B virus, DHBV), geese (Snow goose hepatitis B virus, SGHBV, and Ross goose hepatitis B virus, RGHBV), cranes (Crane hepatitis B virus, CHBV), herons (Heron hepatitis B virus, DHBV), and storks (Stork hepatitis B virus, SHBV; (Chang et al., 1999; Guo et al., 2005; Prassolov et al., 2003; Pult et al., 2001; Sprengel et al., 1988). Most recently an avihepadnavirus was discovered for the first time in a Ring-necked Parakeet (*Psittacula krameri*) in Poland (Piasecki et al., 2012). In the light of the discovery of this Parrot hepatitis B virus (PHBV), we decided to test a library of archived parrot liver samples collected in Poland between 2003 and 2011 for similar avihepadnaviruses in order to determine the prevalence of this virus, its genomic diversity and the extent of its psittacine species host range.

Endogenous hepadnaviruses have also been discovered in various avian species including finches, olive sunbird, dark-eyed junco (Gilbert and Feschotte, 2010; Katzourakis and Gifford, 2010) and more recently in the budgerigar (Cui and Holmes, 2012; Liu et al., 2012). Analysis of the endogenous viruses indicates that viruses distantly related to extant avihepadnaviruses likely became integrated into avian genomes at least as long ago as 19 million years (Gilbert and Feschotte, 2010; Katzourakis and Gifford, 2010). Cui and Holmes, (2012) suggest that despite the fact that endogenous avihepadnaviruses have been inherited between certain bird species, the circulating exogenous avihepadnaviruses do not show obvious evidence of virus-host co-divergence and the fact that there has been some cross species transmission, will complicate attempts to accurately date the origin of the avihepadnaviruses based on the divergence times of their host species. Nonetheless, earlier estimates by Zhou and Holmes (2007) suggest that, based on hepadnavirus substitution rates of between ~10^{-4}-10^{-6} substitutions/site/year the time to the most recent ancestor (TMRCA) of all then known avihepadnaviruses (i.e. excluding the divergent parrot infecting group) is less that 6000 years ago. In light of this, and given the diversity of the PHBV genomes that we have sequenced, we additionally decided to re-estimate the nucleotide substitution rates and associated TMRCA's using a larger sample of 77 avihepadnaviruses including 18 new sequences of PHBV isolates sampled in Poland.

Results and discussion

Parrot avihepanavirus infections and disease pathology

DNA was extracted from liver samples (*n* ¼ 41) representing 12 species (see Table 1 for details) from birds that had died from unknown diseases. Of the 41 samples screened, we found 18 to be PHBV positive. These comprised 15 samples from Ring-necked Parakeets and one sample each from an African Grey Parrot (*Psittacus erithacus*), an Alexandrine Parakeet (*Psittacula eupatria*) and a Crimson Rosella (*Platycercus elegans*). We determined the full genomic sequences of one cloned genome from each of these 18 birds.

A review of the pathology of the PHBV positive birds revealed that infection was mainly observed in young birds (2–5 weeks: *n* ¼ 9; 2–5 months: *n* ¼ 4; 6–9 months: *n* ¼ 6) with a high incidence amongst Ring-necked Parakeets (85% of analysed birds; pathology observations are summarised in Tables 1 and 2). Interestingly, these birds succumbed to sudden deaths without any physical symptoms. In summary, gross lesions were frequently observed in the livers and spleens of these birds, while in some cases splenomegaly and marble spleens were also noted.

Table 1
Details of PHBV infected parrots, full genome accession number and specific pathology related to individual birds.

PHBV strains	Lab ID	GenBank accession #	Common name	Sampling date	Age of birds	Liver	spleen	others	PHBV	APV	BFDV
PHBV-A	**328**	**JN565944**	Ring necked parakeet	**16.03.2007**	**unknown**	**none**	–	–	+		+
PHBV-A	337	JX274026	Ring necked parakeet	29.03.2007	6 months	Swollen marble-mottled haemorrhage congestion	Splenomegaly	–	+		+
PHBV-B	360	JX274025	Ring necked parakeet	18.05.2007	6 months	Small hepatomegaly congestion	Splenomegaly	–	+		
PHBV-B	410	JX274024	African grey parrot	22.08.2007	–	–	–	–	+		+
PHBV-C	755	JX274023	Ring necked parakeet	12.01.2009	< 1 year	–	–	–	+		
PHBV-D	830	JX274022	Alexandrine parakeet	26.03.2009	3–4 weeks	Swollen, marble congestion	Splenomegaly	Subcutaneous petechiae, pale muscles	+	+	+
PHBV-G	902	JX274018	Crimson rosella	23.06.2009	6 months	–	–	–	+		+
PHBV-E	904	JX274021	Ring necked parakeet	23.06.2009	< 1 year	–	–	–	+	+	
PHBV-F	1032	JX274020	Ring necked parakeet	12.03.2010	3 weeks	Hepatomegaly marble -mottled haemorrhage congestion		Subcutaneous petechiae, Pale muscles	+		+
PHBV-A	1035	JX274035	Ring necked parakeet	12.03.2010	2–4 weeks	Swollen, pallor		Muscle petechiae	+		+
PHBV-A	1036	JX274034	Ring necked parakeet	12.03.2010	2–4 weeks	Swollen, pallor		Muscle petechiae	+		+
PHBV-A	1037	JX274033	Ring necked parakeet	12.03.2010	2–4 weeks	Swollen, pallor		Muscle petechiae	+		
PHBV-A	1194	JX274032	Ring necked parakeet	4.03.2011	3–6 weeks	Swollen, pallor		Muscle petechiae	+		
PHBV-A	1195	JX274031	Ring necked parakeet	4.03.2011	3–6 weeks	Swollen, pallor		Muscle petechiae	+		
PHBV-A	1220	JX274030	Ring necked parakeet	7.09.2011	< 1 year	Hepatomegaly marble - mottled haemorrhage	Splenomegaly		+	+	
PHBV-A	1232	JX274029	Ring necked parakeet	07.06.2011	3–5 weeks	Hepatomegaly congestion	Splenomegaly	Muscle petechiae	+		+
PHBV-G	1233	JX274019	Ring necked parakeet	07.06.2011	3–5 weeks	Hepatomegaly congestion	Splenomegaly	Muscle petechiae	+	+	+
PHBV-A	1248	JX274028	Ring necked parakeet	09.09.2011	< 1 year	Swollen			+		+
PHBV-A	1288	JX274027	Ring necked parakeet	24.11.2011	6 months	Swollen marble-mottled haemorrhage Pallor	Splenomegaly Marble appearance	Subcutaneous petechiae	+		+

Muscles and subcutaneous petechiae were observed more frequently than petechiae on the heart (Table 1). These symptoms were not observed in birds that were found to be negative for PHBV. However, it must be noted that we are unable to verify that the observed pathology is linked directly to pathological PHBV infections.

Therefore the overall disease pathology resulting in psittacines from PHBV infection is currently unknown and there is an urgent need to address this gap in our knowledge. Nonetheless, we noted that at least 13/18 of the PHBV infected birds were coinfected with either *Avian polyomavirus* (APV) or *Beak and feather disease virus* (BFDV; Table 1). BFDV is the main causative agent of Psittacine beak and feather disease (PBFD), and frequent coinfections involving APV and PHBV, might indicate that these viruses together constiute an important hitherto unappreciated disease complex.

Avihepadnavirus genome organisation and conserved domains

As with the PHBV isolate previously described, the 18 PHBV full genome sequences determined here have similar genome organisations to all other avihepadnaviruses, with three main open reading frames encoding the PreC/C, PreS/S and polymerase polyproteins. These genome sequences also have many previously identified sequence motifs know to be essential for nucleic acid synthesis (such as the epsilon motif, the DR1 and DR2 motifs, TATA cis regulatory elements, transcription factor binding sites; Fig. 1, Supplementary Fig. 1) (Mueller-Hill and Loeb, 2002).

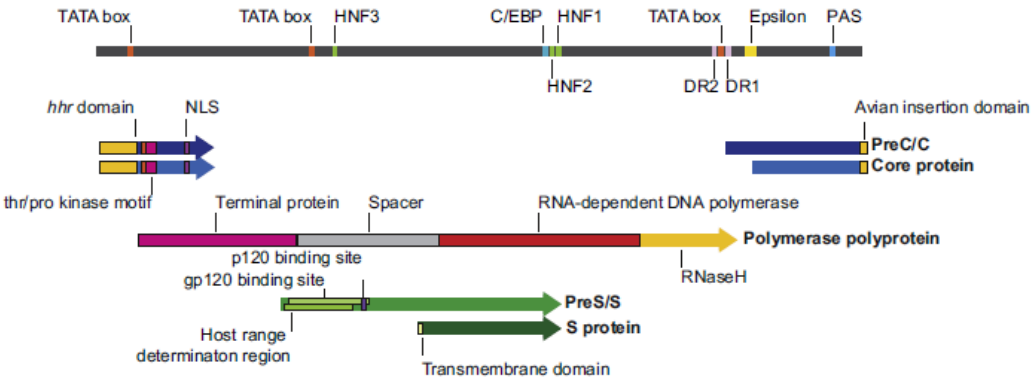
Various domains within the predicted PHBV proteome such as the terminal domain, the reverse transcriptase domain (RNA-dependent DNA polymerase), the RNase H domain, the putative myristylation sites, the host range determination region, the gp120 binding site, the p120 binding region, the transmembrane domain, the conserved avian insertion domain, the *hhr* domain and a X-like domain were identified within the polyprotein, preS/S and the PreC/C ORFs of the PHBVs (Lilienbaum et al., 1993; Liu et al., 1994; Zoulim and Seeger, 1994) (supplementary Fig. 2). As is the case with HHBV and STHBV, some of the PHBV isolates (labelled A, B, C, D and E in Figs. 2 and 3) have an X-like ORF. Similarly, PHBV-G has a homologous X-like element (as do DHBV, SHBV, CHBV and RGHBV) with a potential alternative start codon (supplementary Fig. 3). In the case of woodchuck hepatitis B virus (WHV) the X-like ORF, through *in vivo* experiments, has been shown to be required for the establishment of chronic infections and could contribute to hepatocarcinogenesis (Chang et al., 2001). The balance of evidence, however, suggests that by itself this gene does not cause orthohepadnavirus related carcinogenesis (reviewed by (Seeger and Mason, 2000).

Maximum likelihood phylogenetic and pairwise identity analyses of avihepadnavirus isolates

A maximum likelihood (ML) phylogenetic tree of the full genomes of avihepadnavirus (Fig. 2) indicated that the 18 new PHBV isolates (Table 1; Genbank accession # JX274018– JX274035; supplementary Fig. 1) could be subdivided into two major clades (Fig. 2); one with isolates sharing 495% pairwise identities (1- *p*-distances with pairwise deletion of gaps; isolates labelled A–F in Figs. 2 and 3) with the PHBV isolate described by Piasecki et al., (2012), whereas the second is more distantly related (79% pairwise identity) with the two isolates from Ring-necked Parakeet and Crimson Rosella sharing 99.4% pairwise identity (isolates P902 and P1233 labelled G in Figs. 2 and 3; Table 1; GenBank accession # JX274018 and JX274019).

Table 2
Summary of pathological conditions found to be potentially associated with PHBV.

Gross lesion	Percentage PHBV infected parrots (%)
Swollen Liver	60.0
Marble-mottled haemorrhage	
Liver	33.3
Spleen	6.7
Haemorrhage	
Myocardial	26.7
Subcutaneous	40.0
Muscle	20.0
Liver congestion	40.0
Hepatomegaly	33.3
Pallor	
Liver	33.3
Muscle	20.0
Splenomegaly	46.7
Hydropericardium	6.7



HNF1:GTTAATGRTTAAACC
HNF2:CAGTGTTTRCTTTY
HNF3:GTGCAGRCAG (PHBV-A- F) /ATCGACGCCA (PHBV-G)
Epsilon motif:GCCAAGGTATYTTTACGTCTTGCTTGCTGTTGYCTGCATTGRCTGTACCTTYGG
PAS:AATAAA
DR1 / DR2:TACACCCCTCTC
C/EBP: RCATGGCGCAATAT

Fig. 1. Illustration of the genome organisation of PHBV isolates with various domains and motifs highlighted.

A comparative analysis of pairwise genetic distances of the full genomes of avihepadnaviruses revealed that the pairwise identity of the two most distantly related PHBV isolates analysed here (22%; based on genome wide pairwise distance calculated with pairwise deletion of gaps) is similar to that of the two most divergent non-PHBV avihepadnaviruses (i.e. DHBV, SGHBV, SHBV, CHBV, RGHBV, HHBV and STHBV; 21%; Fig. 3). Similarly, pairwise comparisons (2927 pairwise comparisons) of 77 avihepadnavirus full genome sequences indicated that they all share >75% pairwise identities (Fig. 3). The diversity amongst the 39 DHBV isolates (Fig. 3) is relatively low (*p*-distances o 10%), as is the case for sparsely sampled isolates of CHBV (*p*-distances o 2%), SGHBV (*p*-distances o 1%) and STHBV (*p*-distances o 1%).

The distribution of pairwise identities clearly indicates troughs at 80–81% and 96–95% which could easily be used for strain and genotype demarcations. This can easily be expanded beyond the current criteria for species demarcation set by ICTV for avihepadnaviruses based on DHBV and HHBV isolates, which state that isolates with

>21.6% pairwise distance should be classified as new species. Based on the information available and the distribution of pairwise identities (Fig. 3), we propose that avihepadnavirus sequences with <80% pairwise identity should perhaps be considered as new species. Finally, we can easily further classify these sequences based on a 95% genome-wide sequence identity clustering rule, i.e. isolates sharing <95% pairwise identity belong to the variant grouping. For descriptive purposes we have tentatively assigned the names A–G to the various PHBV variant groupings and A–L to the DHBV variant groupings (Figs. 2–4, supplementary Figs. 1–3, Table 1).

Polyprotein, PreC/C and PreS/S maximum likelihood phylogenetic analyses

ML phylogenetic trees of the polymerase polyprotein, Prec/C and Pres/S amino acid sequences show that while the overall tree topology inferred from these different proteins is similar to that of the full genome trees, there are some noteworthy subtle differences (Fig. 4, supplementary Fig. 2). For example, the PreC/C and PreS/S proteins of SHBV isolates cluster in-amongst those of DHBV, whereas the polyproteins of the SHBV isolates cluster basal to those of the DHBV and SGHBV isolates. Most of the other differences between the trees are in the patterns of branching within the DHBV (A–L) and PHBV (A–F) clusters (Fig. 4, Supplementary Fig. 2). The PreC/C, PreS/S and polymerase polyprotein sequences of the PHBV-A, -B, -C, -D, -E, -F isolates share <98.5% amino acid similarity, which is similar to that shared by the two PHBV-G isolates. However, when PHBV-A, -B, -C, -D, -E, -F are collectively compared to PHBV-G, the PreC/C, PreS/S and polymerase polyproteins of these share only "" 82%, 65% and 70% similarity respectively. A comparative analysis of PHBV with all other avihepadnaviruses indicates that they respectively share >76%, 65% and 65% similarity in the PreC/C, PreS/S and polymerase polyproteins. The similarities in comparisons of proteins of PHBV-A, -B, -C, -D, -E, -F and -G between themselves and with all other avihepadnaviruses is striking but not unexpected given that similar degrees of diversity were observed with their full genome nucleotide sequences.

Recombination analysis

Other than the recombination events reported by Piasecki et al. (2012) we found no new events of recombination amongst the avihepadnaviruses. However, large scale screening and determination of PHBV genomes from infected parrots from all around the world will assist in identifying new recombinants, recombination hotspots and over-all recombination patterns that can be compared to those of orthohepadnaviruses (Bollyky et al., 1996;)

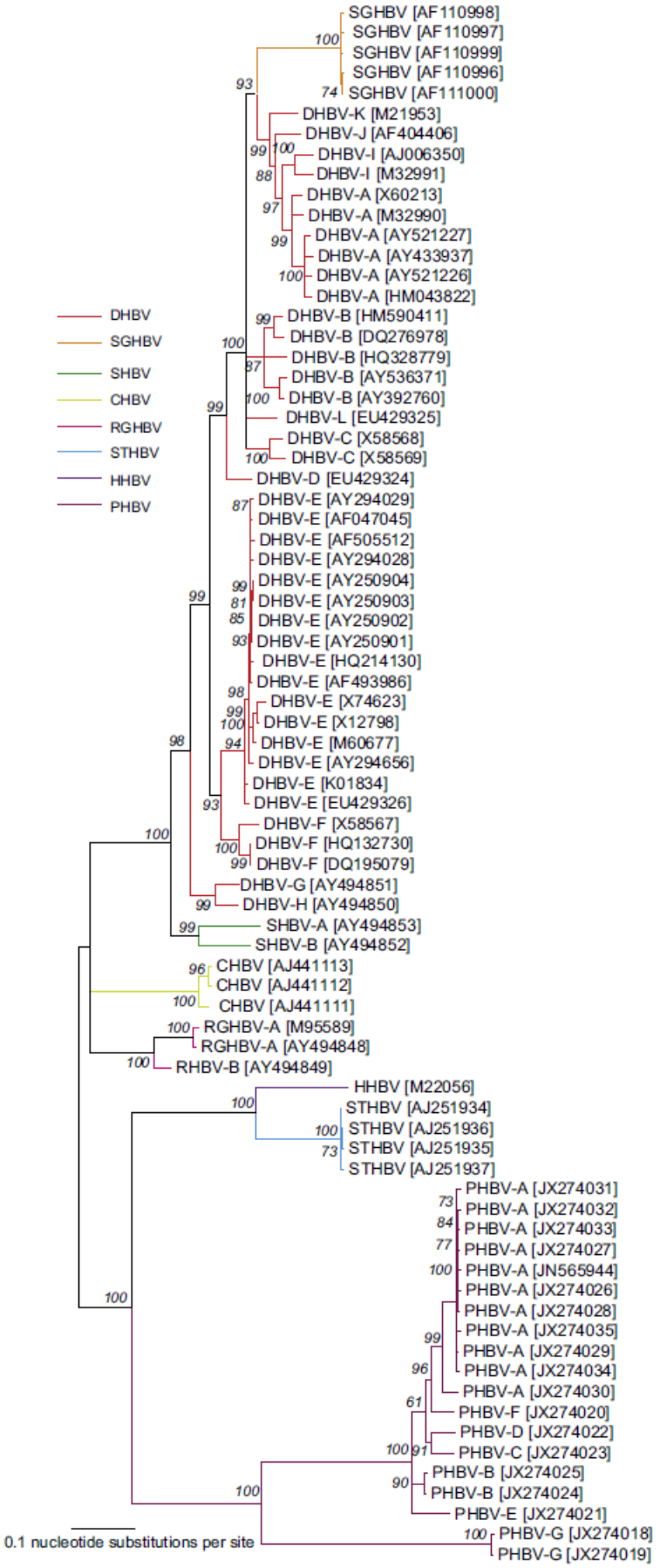


Fig. 2. Maximum likelihood phylogenetic tree based on full genomes illustrating the evolutionary relationships between 77 avihepadnaviruses (rooted with orthohepadnavirus isolate Hepatitis B virus subtype ayr, GenBank accession # X04615). Branch supports based on 1000 bootstrap replicates are provided next to each branch.

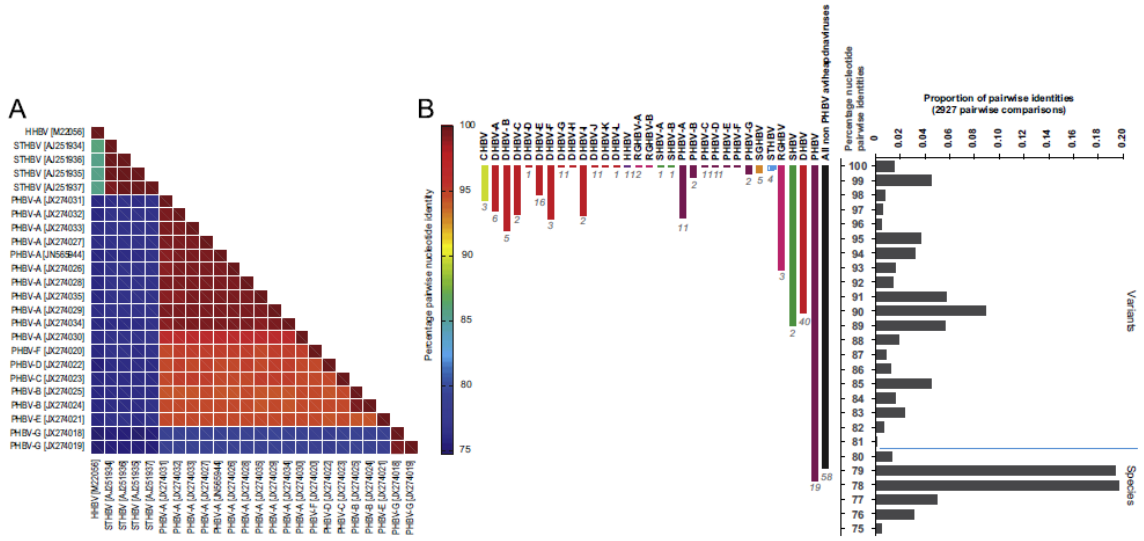


Fig. 3. (A) Two-dimensional graphical illustration of genome-wide percentage pairwise nucleotide sequence identities (calculated with pairwise deletion of gaps) of the PHBV isolates and those of the divergent HHBV and STHBV isolates. (B) Distribution of percentage pairwise identities of avihepadnaviruses (right) and comparison of diversity amongst various avihepadnaviruses species (left; number at the bottom of the bars denote number of isolates in each pairwise comparison).

We attempted to infer the nucleotide substitution rates of avihepadnaviruses using our expanded dataset of full genome sequences (all of which had known sampling dates) using the molecular clock analysis methods implemented in the computer programme BEAST v1.7.3 (Drummond and Rambaut, 2007). We tested a range of plausible demographic and molecular clock models and attempted to identify which of these fitted the data best using Bayes factor tests (Kass and Raftery, 1995). These tests indicated that for both the full genome dataset and a C gene dataset containing no evidence of recombination, a relaxed-clock provided a better fit to the data than a strict-clock model irrespective of the demographic model employed (full genome $BF_{\log 10}$ =70.13, capsid 12.89), but was equivocal between different demographic models under the relaxed clock (full genome $BF_{\log 10}$ =0.21).

For the sake of brevity, the maximum clade credibility (MCC) trees for the full genome and C gene datasets under a constant population size relaxed-clock model are shown in supplementary Figs. 4 and 5. The full genome trees had identical topologies but differed from the ML tree in Fig. 2 in that the rooting achieved with the MCC trees (supplementary Fig. 4) indicates reasonable support (posterior probability ~0.83) for all the PHBV sequences branching from the root of the tree (as is indicated in the preS/S tree of Fig. 4).

For the recombination free C gene dataset the constant (supplementary Fig. 5) and expansion model relaxed-clock MCC trees had identical topologies to their full genome counterparts with the exception that sequence, RGHBV-B [AY494849_USA_2001], which groups with the DHBV clade rendering the RGHBV sequences paraphyletic on the tree. The posterior probabilities of this split were, however, poorly supported (posterior probability¼ 0.53). In contrast, the recombination-free C gene exponential relaxed-clock MCC tree (not shown) the STHBV and the HHBV sequences form a monophyletic clade at the base of the tree that is reminiscent of the tree presented by Zhou and Holmes (2007). The statistical support for this arrangement was however, extremely weak (posterior probability¼ 0.33).

The estimated mean substitution rates with 95% highest probability density (HPD) intervals for the full genome and C gene datasets ranged between 1.01 x 10^{-4} (95% HPD, 4.22 x 10^{-5} – 1.77 x 10^{-4}) and 9.38 x 10^{-5} (95% HPD, 4.19 x 10^{-5} – 1.61 x 10^{-4}) (supplementary Table 1). These nucleotide substitution rates overlap with the estimated mean rates of 7.32 x 10^{-4} and 4.85 x 10^{-4} subs/site/year reported previously by Zhou and Holmes (2007) for the low- and high recombinant regions of the genome respectively from 35 avihepadnaviruses sampled between 1981 and 2001.

Under the constant population size, exponential growth and expansion models and a relaxed-clock, the estimated median time to the most recent common ancestor (TMRCA) of all the analysed avihepadnaviruses ranged between 1485 (95% HPD 468-3587) years (observed with the C gene dataset under an exponential growth model) and 4753 (95% HPD 1935-8535) years (observed with the full genome dataset under a constant population size model) (supplementary Table 1).

It should be emphasised however, that the degree of clock-like evolution evident in the data is low. Using a regression approach whereby the root-to-tip genetic distances inferred from neighbour-joining phylogenetic trees are regressed against sampling times using the programme PATH-O-GEN v1.3 (http://tree.bio.ed.ac.uk/software/pathogen/), indicated that there is no strong evidence for a significant temporal signal in either the C gene (r^2 ¼ 0.08916, residual mean squared= 9.32 x 10^{-5}) or full genome (r^2 ¼ 5.1324 x 10^{-3}, residual mean squared= 2.41 x 10^{-5}) datasets. Thus any attempt to date the origin of these isolated would be invalid with the current available data.

Conclusions

This study provides verification of the initial discovery of PHBV infections in parrots. While the only cases of infection so far have been found in Poland, further testing of psittacine species in other countries will determine how widespread PHBV infection is among these birds. Such testing will hopefully provide information on the potential global distribution of these viruses, and facilitate the detection of any new recombinants amongst them. Perhaps testing for HBV should be routinely carried out in aviaries, especially when psittacine birds die unexpectedly and where the presence of BFDV or APV infections are also suspected or detected. Although we have attempted to estimate the TMRCAs of all the currently sampled avihepadnavirus full genomes, the analysis of more avihepadnavirus sequence data, particularly from archived samples, will be required to obtain credible estimates of the substitution rates of these viruses. Without such data it will not be possible to properly reconcile the apparently high evolution rates of avihepadnaviruses with the fact that close relatives of these viruses likely became integrated into bird genomes at least as long ago as 19 million years.

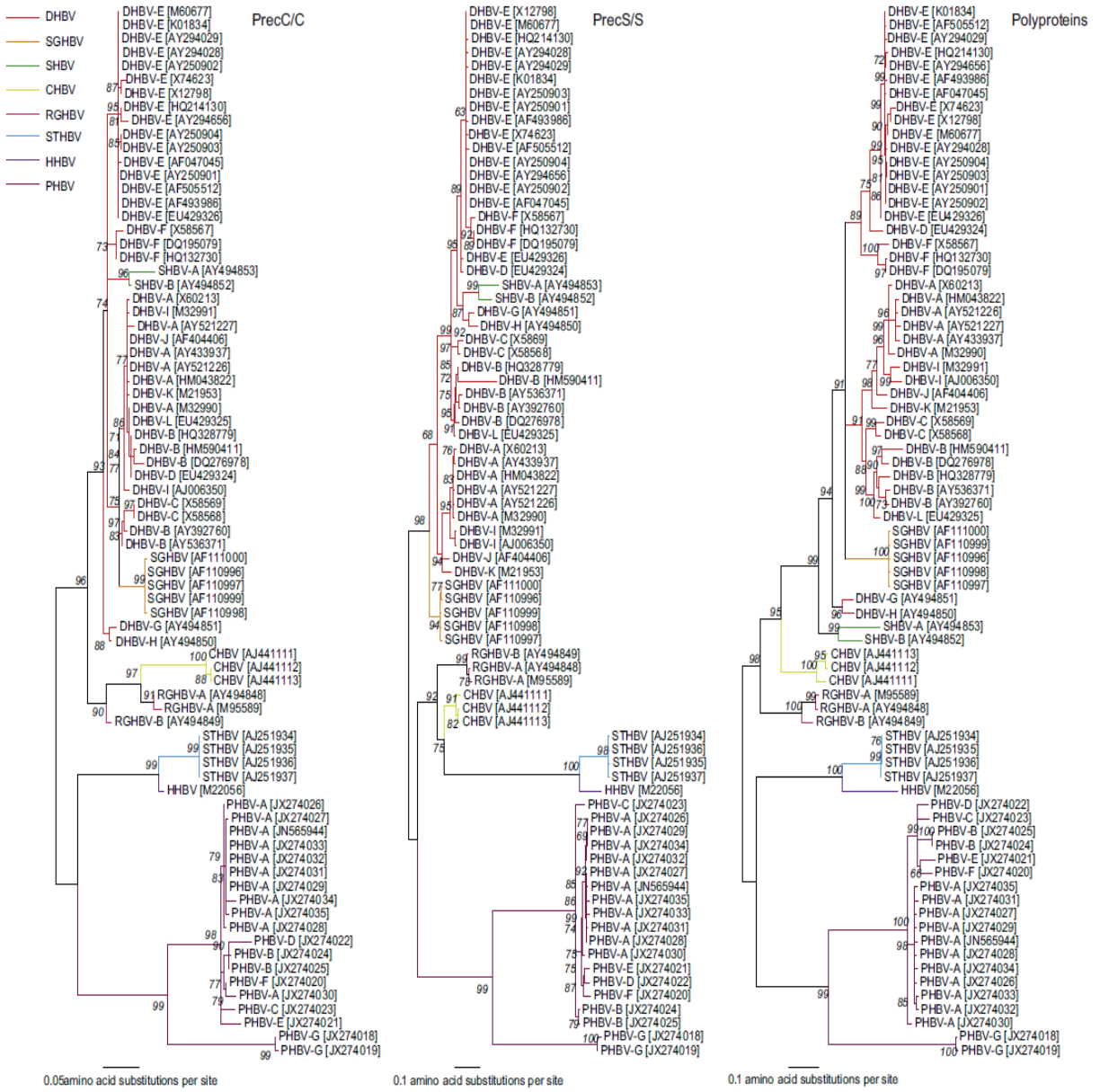


Fig. 4. Maximum likelihood phylogenetic trees of the PreC/C, PreS/S and polyproteins of 77 avihepadnaviruses (rooted with orthohepadnavirus isolate GenBank accession # X04615). Branch supports based on aLRT are provided next to each branch.

Material and methods

DNA extraction, amplification, cloning and sequencing of parrot hepadnavirus isolates

DNA from liver samples was extracted using commercially available Genomic Mini kit (A&A Biotechnology, Gdynia, Poland), according to the manufacturer's instructions. Total DNA was extracted from samples and subjected to rolling-circle amplification (RCA) using TempliPhi™ (GE Healthcare, USA), to non-specifically amplify any circular DNA (such as circular viral genomes) as described in Piasecki et al. (2012). The RCA concatemers were digested with *Bam*HI, *Hind*III and *Xmn*I restriction enzymes. The resulting fragments ranging in size from 1–3 kb were cloned into plasmid vectors (*Bam*HI and *Hind*III restricted pGEM3ZF and pJET1.2 for *Xmn*I restricted products) and sequenced. Based on the resulting sequence information we designed back-to-back primers to amplify full PHBV genomes. One micro litre of the RCA product was used for PCR amplification

using primers PHBV-F 5O - GGAYTTCTCTCAGTTYTCCAAAGG - 3O and PHBV-R 5O - ACCACRARTCTAGCYTCCYCAGT - 3O with Kapa HiFi HotStart DNA polymerase (Kapa Biosystems, USA) using the following protocol: initial denaturation at 94 1C for 2 min, 25 cycles of 98 1C for 20 s, 52 1C for 30 s, 72 1C for 3 min, followed by a final extension at 72 1C for 5 min and a final renaturation step at 4 1C for 10 min. The amplicons were resolved on a 0.7% agarose gel, ~3 kb bands excised, cleaned with Mega-spin agarose gel extraction kit (Intron, Korea) and cloned into pJET1.2 vector (Fermentas, USA). The plasmid isolated from a single *E. coli* colony was sequenced by Macrogen Inc (Korea) by primer walking.

DNA sequence manipulation and dataset preparations

The sequence contigs were assembled using DNAMAN (Version 5.2.9, Lynnon Biosoft, Canada). The full genomes were aligned with other full length genomes of avihepadnaviruses available in GenBank using MUSCLE (Edgar, 2004) as implemented in MEGA 5 (Tamura et al., 2011) with manual editing carried out by eye.

Phylogenetic and recombination analyses

A maximum likelihood phylogenetic tree (Fig. 2) of the full genomes was constructed with an orthohepadnavirus isolate (Hepatitis B isolate, GenBank accession # X04615) as an out-group using nucleotide substitution model GTR þ I þ G_4 (selected as the best substitution model by jModelTest, (Posada, 2008)) with PHYML. Branches with less than 60% approximate likelihood (aLRT) support were collapsed using MESQUITE (Version 2.75). Maximum likelihood phylogenetic trees of the polymerase polyprotein, preC/C and preS/S amino acid sequences (aligned using MUSCLE, Edgar, 2004) were constructed with PHYML (Guindon et al., 2010) using the LG model substitution model selected using ProTest (Abascal et al., 2005) with aLRT being used to assess branch support (Anisimova and Gascuel, 2006). The 77 avihepadnaviruses were analysed for recombination using RDP4 (Martin et al., 2010) with default settings. Nucleotide pairwise identities ([1 – *p*-distances] x 100;with pairwise deletion of gaps) were calculated by aligning two sequences at a time using MUSCLE (Edgar, 2004) implemented in our software SDTv1.0 (Muhire et al., in press) and amino acid identities were calculated using MUSCLE (Edgar, 2004).

Estimation of nucleotide substitution rates

Two avian hepadnavirus alignments comprising 77 full genome (3116 bp) and C gene (786 bp) sequences respectively, obtained from samples collected between 1981 and 2011 were analysed. The Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.7.3 (Drummond and Rambaut, 2007) was employed to co-estimate the nucleotide substitution model parameters, phylogeny and time to the most recent common ancestor (TMRCA) of the avihepadnavirus full genome and preC/C datasets. Three different coalescent parametric demographic models were investigated (constant population size, exponential population growth and expansion growth) with both strict and relaxed molecular clock models and a general time reversible nucleotide substitution model with gamma distributed rate variation (with 4 rate categories) and a proportion of invariable sites (GTRþ G_4 þ I). For each evolutionary model, five independent replicate runs of length 2 x 10^8

steps in the Markov chain were performed using BEAST and the estimated effective sample sizes (ESS) were always4200.

Acknowledgments
Part of the research was supported by a University of Cape Town, South Africa block grant awarded to AV. Part of this research was supported by the Polish Ministry of Education and Science (grant no. N308 036 32/3369).

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.01.009.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Blum, H., Zhang, Z.S., Galun, E., Von Weizsa¨cker, F., Garner, B., Liang, T., Wands, J., 1992. Hepatitis B virus X protein is not central to the viral life cycle in vitro. J. Virol. 66, 1223–1227.

Bollyky, P.L., Rambaut, A., Harvey, P.H., Holmes, E.C., 1996. Recombination between sequences of hepatitis B virus from different genotypes. J. Mol. Evol. 42, 97–102.

Chang, S.F., Netter, H.J., Bruns, M., Schneider, R., Fro¨lich, K., Will, H., 1999. A new avian hepadnavirus infecting snow geese (Anser caerulescens) produces a significant fraction of virions containing single-stranded DNA. Virology 262, 39–54.

Chang, S.F., Netter, H.J., Hildt, E., Schuster, R., Schaefer, S., Hsu, Y.C., Rang, A., Will, H., 2001. Duck hepatitis B virus expresses a regulatory HBx-like protein from a hidden open reading frame. J. Virol. 75, 161–170.

Cui, J., Holmes, E.C., 2012. Endogenous hepadnaviruses in the genome of the budgerigar (Melopsittacus undulatus) and the evolution of avian hepadna-viruses. J. Virol. 86, 7688–7691.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. B.M.C Evol. Biol. 7, 214.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Fallot, G., Halgand, B., Garnier, E., Branger, M., Gervais, A., Roque-Afonso, A.M., Thiers, V., Billaud, E., Matheron, S., Samuel, D., Feray, C., 2012. Recombination of hepatitis B virus DNA in patients with HIV. Gut 61, 1197–1208.

Funk, A., Mhamdi, M., Will, H., Sirma, H., 2007. Avian hepatitis B viruses: molecular and cellular biology, phylogenesis, and host tropism. World J. Gastroenterol. 13, 91.

Gilbert, C., Feschotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol., 8.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Guo, H., Mason, W.S., Aldrich, C.E., Saputelli, J.R., Miller, D.S., Jilbert, A.R., Newbold, J.E., 2005. Identification and characterization of avihepadnaviruses isolated from exotic anseriformes maintained in captivity. J. Virol. 79, 2729–2742.

Ishikawa, T., Ganem, D., 1995. The Pre-S domain of the large viral envelope protein determines host-range in avian Hepatitis-B viruses. PNAS 92, 6259–6263.

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795. Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6, e1001191.

Lilienbaum, A., Crescenzochaigne, B., Sall, A.A., Pillot, J., Elfassi, E., 1993. Binding of nuclear factors to functional domains of the duck hepatitis-B virus enhancer. J. Virol. 67, 6192–6200.

Liu, C., Mason, W.S., Burch, J.B.E., 1994. Identification of factor-binding sites in the duck hepatitis-B virus enhancer and in-vivo effects of enhancer mutations. J. Virol. 68, 2286–2296.

Liu, W., Pan, S., Yang, H., Bai, W., Shen, Z., Liu, J., Xie, Y., 2012. The first full-length endogenous hepadnaviruses: identification and analysis. J. Virol. 86, 9510–9513.

Liu, W., Zhai, J., Liu, J., Xie, Y., 2010. Identification of natural recombination in duck hepatitis B virus. Virus Res. 149, 245–251.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Mueller-Hill, K., Loeb, D.D., 2002. Cis-acting sequences 5E, M, and 3E interact to contribute to primer translocation and circularization during reverse transcription of avian hepadnavirus DNA. J. Virol. 76, 4260–4266.

Muhire B., Martin D.P., Brown J.K., Navas-Castillo J., Moriones E., Zerbini F.M., Rivera-Bustamante R., Malathi V.G., Briddon R.W., Varsani A., A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae), Arch. Virol., http://dx.doi.org/10.1007/s00705-012-1601-7, in press.

Piasecki, T., Kurenbach, B., Chrzastek, K., Bednarek, K., Kraberger, S., Martin, D.P., Varsani, A., 2012. Molecular characterisation of an avihepadnavirus isolated from Psittacula krameri (ring-necked parrot). Arch. Virol. 157, 585–590.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.