Global phylogeography and evolution of chelonid fibropapilloma-associated herpesvirus

A. R. Patrício,1 L. H. Herbst,2 A. Duarte,3 X. Vélez-Zuazo,1,4 N. Santos Loureiro,5 N. Pereira,6 L. Tavares3 and G. A. Toranzos1

Correspondence
A. R. Patrício
caldaspatricio.ar@gmail.com

Received 19 October 2011
Accepted 17 January 2012

1Department of Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, PR 00931, Puerto Rico
2Department of Pathology and Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, NY, USA
3Interdisciplinary Centre of Research in Animal Health, Faculty of Veterinary Medicine of the Technical University of Lisbon, 1300-477 Lisbon, Portugal
4ecOceanica, Lima 41, Peru
5Faculty of Sciences and Technology (DCTMA) of the University of Algarve, Gambelas Campus, 8005-139 Faro, Portugal
6Lisbon Oceanarium, 1999-005 Lisbon, Portugal

A global phylogeny for chelonid fibropapilloma-associated herpesvirus (CFPHV), the most likely aetiological agent of fibropapillomatosis (FP) in sea turtles, was inferred, using dated sequences, through Bayesian Markov chain Monte Carlo analysis and used to estimate the virus evolutionary rate independent of the evolution of the host, and to resolve the phylogenetic positions of new haplotypes from Puerto Rico and the Gulf of Guinea. Four phylogeographical groups were identified: eastern Pacific, western Atlantic/eastern Caribbean, mid-west Pacific and Atlantic. The latter comprises the Gulf of Guinea and Puerto Rico, suggesting recent virus gene flow between these two regions. One virus haplotype from Florida remained elusive, representing either an independent lineage sharing a common ancestor with all other identified virus variants or an Atlantic representative of the lineage giving rise to the eastern Pacific group. The virus evolutionary rate ranged from $1.62 \times 10^{-4}$ to $2.22 \times 10^{-4}$ substitutions per site per year, which is much faster than what is expected for a herpesvirus. The mean time for the most recent common ancestor of the modern virus variants was estimated at 192.90–429.71 years ago, which, although more recent than previous estimates, still supports an interpretation that the global FP pandemic is not the result of a recent acquisition of a virulence mutation(s). The phylogeographical pattern obtained seems partially to reflect sea turtle movements, whereas altered environments appear to be implicated in current FP outbreaks and in the modern evolutionary history of CFPHV.

INTRODUCTION

Chelonid fibropapilloma-associated herpesvirus (CFPHV) is the most likely aetiological agent of fibropapillomatosis (FP) (Arthur et al., 2008a; Greenblatt et al., 2005; Lackovich et al., 1999; Quackenbush et al., 1998), a neoplastic disease of sea turtles, characterized by recent outbreaks (Diez et al., 2010; Foley et al., 2005; Work & Balazs, 1999). The tumours can be both external and internal and, although benign, depending on location and size, they may obstruct crucial functions such as swimming, feeding and sight, or may impede organ function (Herbst, 1994; Herbst & Klein, 1995). Severe FP also leads to bacteraemia (Work et al., 2003). The most susceptible life stages appear to be neritic juveniles and subadults, whereas in adults the disease is rare (Ene et al., 2005; Herbst & Klein, 1995; Work et al., 2004). A high prevalence of FP is common in anthropogenically altered environments (Aguirre & Lutz, 2004; Herbst, 1994; Van Houtan et al., 2010), suggesting that factors in these environments promote disease outbreaks, for example, through the enhancement of virus transmissibility and/or the enhancement of disease expression via substances that

The GenBank/EMBL/DDJB accession numbers for the CFPHV sequences determined in this study are JN580279–JN580296 and JN625251–JN625262.

Further details about samples are available with the online version of this paper.
modulate the immune system or promote tumour formation (Chaloupka et al., 2008). Although more frequent among green turtles (Hirama & Ehrhart, 2007), FP has been reported in all species of sea turtles (Foley et al., 2005; Herbst et al., 2004; Lackovich et al., 1999). These are listed as vulnerable, endangered or critically endangered (IUCN, 2010), and therefore FP is one more threat to their survival.

CFPHV DNA has been consistently detected by PCR analysis in tumour tissue samples from sea turtles (Arthur et al., 2008a). So far, over 90,000 bp from the genome of CFPHV have been assembled (Herbst et al., 2004). Based on genome organization and sequence similarity with other herpesviruses, CFPHV is classified within the subfamily Alphaherpesvirinae, although not within any of the current reference genera of alphaherpesviruses (Greenblatt et al., 2005; McGeoch & Gatherer, 2005).

In their phylogenetic analysis of CFPHV variants, Herbst et al. (2004) identified two major clades, each with Atlantic and Pacific representatives. A study of the CFPHV variants in coastal sites from Florida detected strong spatial heterogeneity in the distribution of variants, suggesting that infection occurs locally after arrival in coastal foraging areas (Ene et al., 2005). Furthermore, Ene et al. (2005) and Herbst et al. (2004) found that sympatric species of sea turtles shared some variants of the virus at each locality. Greenblatt et al. (2005) also found spatial heterogeneity among virus variants on a larger geographical scale and identified four regional groups: Atlantic, west Pacific, mid-Pacific and east Pacific. Remarkably, the CFPHV variant from Puerto Rico included in their study did not cluster with any of the groups identified in that analysis (Greenblatt et al., 2005).

Phylogenies of herpesviruses typically have topologies congruent with those of their tetrapod hosts, suggesting a host-linked evolution (Davison, 2002; Firth et al., 2010; Umene & Sakaoka, 1999). Following this assumption, the evolutionary rate of herpes simplex virus type 1 (HSV-1) has been estimated to be between $3.5 \times 10^{-8}$ and $3.0 \times 10^{-9}$ substitutions per site per year (McGeoch et al., 2000; Sakaoka et al., 1994). Using the published rates for HSV-1 and the amount of nucleotide divergence among CFPHV haplotypes, Herbst et al. (2004) estimated that the two most divergent virus clades separated 1.6–4.0 million years ago (mya). They also used the rise of the Isthmus of Panama (3.1–3.5 mya) to separate the Florida and Hawaii lineages, and estimated, independent of host-linked evolution, that the two most divergent virus clades separated 7.9–8.9 mya. In both cases, they concluded that CFPHV has been associated with sea turtles for millions of years (Herbst et al., 2004).

Although it is widely accepted that herpesviruses co-diverge with their hosts, this assumption is seldom tested (Duffy et al., 2008; Firth et al., 2010). Adding temporal information from heterochronous sequences (i.e. from different points in time) using molecular clock models allows an estimation of the rate and timescale of virus evolution independent of host evolution (Drummond & Rambaut, 2007). In fact, these methods have been widely employed to estimate the evolutionary and epidemiological characteristics of a broad range of rapidly evolving RNA and ssDNA viruses, retrieving estimates that agree well with the known epidemiology of these pathogens (Firth et al., 2009; Rambaut et al., 2008; Shackelton et al., 2005).

In the present study, sampling time, spanning 20 years, was incorporated in a Bayesian Markov chain Monte Carlo (MCMC) analysis using BEAST version 1.6 (Drummond & Rambaut, 2007) to: (i) obtain an estimate of the mean evolutionary rate of CFPHV independent of co-evolution with its host species, and (ii) infer a global phylogeny for CFPHV, including newly identified haplotypes from Puerto Rico and the Gulf of Guinea, West Africa. Ultimately, we aim to contribute to expanding knowledge of the evolution and demographic history of CFPHV.

**RESULTS**

**Sequence recombination assessment**

Maximum-likelihood (performed in PhyML version 3.0) and Bayesian gene phylogenies inferred for the DNA polymerase (483 bp), UL18 (1212 bp), UL34 (861 bp) and glycoprotein B (2486 bp) genes of CFPHV, sampled from four species of sea turtles from nine geographical areas (Fig. 1), resulted in similar topologies for each gene (data not shown). Hence, further analyses were based on the

![Fig. 1. Location of sampling sites for CFPHV](map created using http://www.seaturtle.org/maptool/). CR, Costa Rica (Pacific coast); FL, Florida; HW, Hawaii; AUS, Australia (Pacific coast); BRB, Barbados; MEX, Mexico (Pacific coast); SD, San Diego (CA); GG, Gulf of Guinea (West Africa); PR, Puerto Rico.
Bayesian inferred trees (Fig. 2). Although the phylogenies of the four analysed genes were mostly in agreement, some samples changed branches between gene trees. The sample PR1 (Greenblatt et al., 2005) from Puerto Rico, was placed in different branches, corresponding to different geographical locations, in each gene tree (Fig. 2). The conclusion that PR1 was a recombinant was well supported by five of six recombination analysis methods performed using the RDP3 (Martin et al., 2010) with P < 0.001 (see Methods for the recombination detection methods used). Samples PR3 and AUS1 (from Australia) also swapped positions between some gene trees (Fig. 2). Based on recombination analyses in RDP3, gene tree topologies and preliminary MCMC analyses using a concatenated matrix (‘supermatrix’) of the four analysed genes, samples PR1, PR3 and AUS1 were determined as recombinants, potentially artefacts resulting from mis-identification of sequences in GenBank, and removed from further analyses, as the tree topology and branch support, but not evolutionary parameter estimates, were influenced by these samples.

**Convergence diagnostics and prior sensitivity**

Bayesian MCMC analyses were performed with a relaxed molecular clock model and under four prior distributions, using the generalized time reversible (GTR) (Tavaré, 1986) and HKY (Hasegawa et al., 1985) models of nucleotide substitution, each with the coalescent parameters of constant population size and exponential growth (Ho et al., 2005a). The relaxed clock model was preferred, as, unlike the strict molecular clock model, which assumes that nucleotide substitutions accumulate at a fixed rate over time, the former allows rates of nucleotide substitution to vary among genomes, genes and lineages, providing better estimations derived from divergences among sequences (Drummond et al., 2006; Ho & Larson, 2006; Ho et al., 2005b). For both the supermatrix and individual gene alignments, MCMC analyses produced high effective sample sizes (>400) for all parameters of interest, leading to accurate parameter estimates with small SDs from the mean (Drummond et al., 2002). The trace plots showed no obvious trends indicating that the MCMC was still converging, and there were no large-scale fluctuations in the trace suggesting poor mixing. Parameter estimates were highly congruent among independent runs and prior distributions, indicating the robustness of the data and suggesting that the sampled genes evolved at similar rates (see Table 1 for summary of MCMC analysis per gene). The mean evolutionary rate estimates were highly similar between genes and the supermatrix, and the range of the 95 % HPD, which is a measure of precision, did not increase with reduced sequence length. The most precise estimates, corresponding to the narrowest 95 % HPD, were produced for the DNA polymerase gene, whereas the glycoprotein B and UL18 genes had the least precise estimates, corresponding to the broadest 95 % HPD intervals.

**Phylogeography of CFPHV**

Four geographical groups were identified: (i) the eastern Pacific group, consisting of samples from San Diego, Costa Rica and Mexico; (ii) the western Atlantic/eastern Caribbean group, including most samples from Florida and Barbados; (iii) the Atlantic group, consisting of samples from the Gulf of Guinea and Puerto Rico; and (iv) the mid-west Pacific group, formed from Australian and Hawaiian samples (Fig. 3). The Florida sample FL3 (most distant variant D from Herbst et al., 2004) formed a distinct Atlantic group. In two of the three gene-specific analyses (glycoprotein B and UL18 in Fig. 2), this sample formed its own clade, sharing a common ancestor with all other Atlantic and Pacific CFPHV groups. In the third gene-specific analysis (DNA polymerase gene, Fig. 2) and in the supermatrix analysis (Fig. 3), it clustered with the eastern Pacific group, which shared a common ancestor with all other groups, similar to the phylogeny shown by Herbst et al. (2004), but with very little support in posterior probabilities (65 %).

**Fast evolutionary rate for CFPHV**

The mean evolutionary rate of CFPHV estimated by MCMC analysis of the supermatrix ranged from 1.62 × 10⁻⁴ to 2.22 × 10⁻⁴ substitutions per site per year. For all independent MCMC runs, the 95 % HPD varied by no more than one order of magnitude from their corresponding posterior mean in the downward direction (a summary of the MCMC results for the supermatrix is shown in Table 2). The mean TMRCA ranged from 192.90 to 429.71 years, with the corresponding 95 % HPD ranging from 48.56 to 781.32 years (Table 2).

For further support of the estimated rate for CFPHV, we ran new MCMC analyses in BEAST version 1.6, this time fixing the substitution rates to our estimated value (~2×10⁻⁴) in one analysis and to the previously predicted rate for HSV-1 in the other analysis (~3×10⁻⁵; Sakaoka et al., 1994) to test the prior likelihoods. The MCMC analysis using the faster rate had a higher prior likelihood and subsequent higher marginal posterior mean: −9662 (95 % HPD = −9678 to −9647) compared with −9920 (95 % HPD = −9935 to −9903) when the MCMC was run under the slower rate, demonstrating a better fit of our substitution rate to the dataset.

**DISCUSSION**

**Host life history implications for the phylogeography of CFPHV**

The diversity and regional distribution of the CFPHV variants found in this study may be explained partially by the sea turtle’s complex life history. Sea turtles undergo an oceanic stage during the ﬁrst few years of their lives, while as juveniles they recruit to coastal foraging habitats where they remain for several years. Mature individuals move periodically from neritic bays to nesting beaches and mating areas,
often separated by hundreds to thousands of kilometres (Arthur et al., 2008b; Bowen & Karl, 2007; Bowen et al., 2007; Casale et al., 2008; Plotkin et al., 1995). Adults return to their region of origin for mating and nesting, whereas immature turtles typically form near-shore aggregations of individuals from multiple nesting colonies (Bowen & Karl, 2007). It has been suggested that oceanic individuals free of FP express the disease only after entering neritic areas (Herbst et al., 2005). FP-associated herpesvirus: phylogeography and evolution

The inferred phylogeny showed that the CFPHV variants were fairly homogeneous among nearby foraging grounds and that in general there was a clear geographical isolation between distant sites and a rapid divergence of virus variants. Additionally, different species of sea turtles in the same locality were shown to share the same haplotypes (also found by Ene et al., 2005 and Herbst et al., 2004). Our results support the conclusion that horizontal transmission in neritic bays is the major route by which individuals become infected, which is further supported by the demonstration that infectious CFPHV particles are transmitted along with FP through skin lesions (Herbst et al., 1995, 1996).

Our phylogeny differed from a previous analysis by Greenblatt et al. (2005). First, the Puerto Rican haplotypes were efficiently placed in the phylogeny of CFPHV. Secondly, with the additional samples included in this study and the elimination of recombinants, a different phylogeographical pattern for the CFPHV variants became apparent. We established that the Puerto Rican and Gulf of Guinea variants were closely related, suggesting recent virus gene flow between these regions. Although initially unexpected considering the geographical distance between the two locations, this connection could be explained by the strong equatorial current from the Gulf of Guinea to Brazil, with its northern arm continuing to the eastern Greater Antilles, known to influence the dispersal and distribution of sea turtle hatchlings (Lahanas et al., 1998). Furthermore, a mixed-stock analysis of the mitochondrial DNA control region of green turtles indicated the presence of shared haplotypes between Puerto Rican foraging bays and West African rookeries (Velez-Zuazo et al., 2010). A plausible hypothesis is that individuals infected with CFPHV at Puerto Rico foraging sites could return to their natal home in the Gulf of Guinea, facilitating virus gene flow. The possibility that virus gene flow occurred in a westward direction by transport on the prevailing current of an infected host or as free virus particles is less likely, as there is no population genetic evidence to support host movement in this direction or to support transoceanic transport of free virus particles. However, waterborne transmission is known to occur with enveloped viruses of fishes (Bowser et al., 1999), and a different but related herpesvirus of sea turtles was shown to retain \textit{in vitro} infectivity for at least 5 days in seawater at environmental temperatures (Curry et al., 2000), leaving open the possibility of long-distance waterborne transmission.

This study identified four phylogeographical groups of CFPHV haplotypes, which can be understood in terms of documented host life history patterns and movements. The position of sample FL3 (from Florida) is a puzzling exception. FL3 was the most divergent CFPHV variant included in this study with \textgreater{}5% nucleotide differences from other Florida variants (Herbst et al., 2004). Although, in agreement with Herbst et al. (2004), it clustered with the eastern Pacific group in the Bayesian phylogeny using the supermatrix of four genes, the support for this grouping was very weak (posterior probability of 65%). Furthermore, in

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|c|}
\hline
Gene & Length (bp) & Marginal posterior mean & UCED mean* & UCED 95% HPD & TMRCA (years) & TMRCA 95% HPD \\
\hline
DNA pol & 482 & -1 107.086 & 4.97 \times 10^{-4} & 1.90 \times 10^{-4} & 59.88 & 27.08–105.16 \\
UL34 & 861 & -1 573.182 & 2.12 \times 10^{-4} & 1.82 \times 10^{-6} & 162.35 & 20.86–433.34 \\
Glycoprotein B & 2486 & -4 945.681 & 1.32 \times 10^{-4} & 1.61 \times 10^{-7} & 519.51 & 38.98–1 165.59 \\
UL18 & 1212 & -2 414.565 & 1.65 \times 10^{-4} & 1.10 \times 10^{-7} & 989.10 & 40.14–15 138.84 \\
\hline
\end{tabular}
\caption{Mean and 95\% highest probability density (HPD) intervals of the Bayesian posterior estimates of substitution rate (substitutions per site per year) and time to most recent common ancestor (TMRCA) of each gene alignment of CFPHV, using the best model of substitution per gene, a relaxed clock model with uncorrelated exponential rates and constant size for the coalescent prior. UCED, uncorrelated exponential distribution.}
\end{table}

*Substitutions per site per year.
two of the three single-gene phylogenies, FL3 formed its own clade, sharing a node with all other variants examined. Therefore, it is likely that the FL3 variant represents a distinct Atlantic lineage that branched off very early from all the other CFPHV lines included in this study. Whether Atlantic haplotype FL3 is more closely related to eastern Pacific variants or represents its own distinct lineage, it raises interesting questions regarding how the lineage evolved. Herbst et al. (2004) proposed that the FL3 variant, to date found only in Atlantic loggerheads, and the Mexican and Costa Rican variants, isolated from eastern Pacific olive ridley turtles, could represent a distinct clade segregating primarily in members of the subfamily Carettinae. The sample SD (from San Diego) from a green turtle (subfamily Cheloniinae), would represent the first exception to this host–virus association. Interspecific transmission of CFPHV has been documented previously in localities where multiple turtle species co-occur (Ene et al., 2005; Herbst et al., 2004), but the degree to which different variants segregate independently within host species and the relative effect of cross-species transmission on host–virus co-evolution remain unclear. Clarification of the relationships between CFPHV variants and host species must await further studies with more samples from localities where multiple turtle species with FP co-occur.

**Faster evolutionary rate for CFPHV**

The evolutionary rate of CFPHV estimated by MCMC analysis was much higher than would be expected under
the theory of co-evolution with host species and was highly congruent among the four genes analysed independently and the supermatrix.

Potential biases inherent to the analysis conducted have been pointed out (Debruyne & Poinar, 2009; Emerson, 2007; Ho et al., 2005a). In particular, it was demonstrated that molecular rates are accelerated at short timescales on population level studies (Ho et al., 2007b), the most likely explanation for this phenomenon being the persistence of slightly deleterious mutations, which, through purifying selection, are eliminated over long timescales (Ho & Larson, 2006; Ho et al., 2005a, 2007a, b).

To test that our results were not a consequence of the time dependency of the molecular clock due to the short time range of our dataset (i.e. 20 years), we conducted selective pressure analyses for each gene, fitting a codon substitution model to HypHY (Pond & Muse, 2005). The analyses revealed that the DNA polymerase, UL34 and UL18 genes of CFPHV were already under purifying selection with an overall ratio of non-synonymous to synonymous substitutions (dN/dS) of 0.17, 0.24 and 0.28, respectively, and only the glycoprotein B gene had an overall dN/dS of ~1, indicative of no significant selective pressure. Moreover, our findings are consistent with recent studies supporting the suggestion that the evolutionary rates of some dsDNA viruses are comparable to those of RNA and ssDNA viruses (Babkin & Shchelkunov, 2008; Firth et al., 2010; Hughes et al., 2010).

Substitution rates of viruses are a complex product of the underlying mutation rate, generation time, effective population size and fitness (Duffy et al., 2008). Factors such as changing environments, high transmission rates and high replication rates can inflate virus evolutionary rates (Duffy et al., 2008; Firth et al., 2010; Hughes et al., 2010). The FP panzootic has emerged over the last few decades, with the disease spreading rapidly and reaching high prevalence in many localities (Herbst, 1994; Herbst et al., 2008), consistently associated with recent environmental changes (Aguirre & Lutz, 2004; Herbst, 1994; Herbst & Klein, 1995; Van Houtan et al., 2010). It is possible that virus transmission and replication rates have been enhanced during this pandemic with an inherent increase in evolutionary rates.

The estimated TMRCA of the CFPHV variants ranged from 192.90 to 429.71 years, which is much more recent than previous estimates placing the emergence of modern CFPHV variants on a scale of several thousand to several million years ago (Herbst et al., 2004), predicted under assumptions of co-divergence with the host species (McGeoch et al., 2000; Sakaoka et al., 1994) or indicating a vicariance event between the eastern Pacific and western Atlantic (Herbst et al., 2004). Nevertheless, our estimates still place the TMRCA previous to the emergence of the current FP panzootic, agreeing with the conclusion by Herbst et al. (2004) that the CFPHV variants acquired the shared mutation(s) responsible for FP oncogenesis prior to their divergence and that this panzootic is not due to the emergence and global spread of a single virulent CFPHV strain, but instead is due to recent environmental changes that favour disease expression. The possibility that contemporary CFPHV variants diverged from a common ancestor within historic times is none the less novel and begs further investigation. It should also be noted that the current analysis estimates the TMRCA of the dataset only, not of the CFPHV lineage, and hence our results do not preclude a much older emergence of this pathogen and its long association with sea turtle hosts.

Table 2. Mean and 95 % HPD intervals of the Bayesian posterior estimates of substitution rate (substitutions per site per year) and TMRCA of CFPHV

<table>
<thead>
<tr>
<th>Clock model</th>
<th>Substitution model</th>
<th>Demographic model</th>
<th>Marginal posterior</th>
<th>UCED rate*</th>
<th>UCED rate 95 % HPD</th>
<th>TMRCA (years)</th>
<th>TMRCA 95 % HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCED</td>
<td>HKY</td>
<td>Constant size</td>
<td>−9820.849</td>
<td>1.62 × 10⁻⁴</td>
<td>2.48 × 10⁻⁵–3.38 × 10⁻⁴</td>
<td>429.71</td>
<td>58.79–752.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential growth</td>
<td>−9808.804</td>
<td>2.22 × 10⁻⁴</td>
<td>5.61 × 10⁻⁵–4.20 × 10⁻⁴</td>
<td>192.90</td>
<td>51.93–427.40</td>
</tr>
<tr>
<td>GTR</td>
<td></td>
<td>Constant size</td>
<td>−9825.397</td>
<td>1.69 × 10⁻⁴</td>
<td>1.53 × 10⁻⁵–3.43 × 10⁻⁴</td>
<td>367.49</td>
<td>58.16–781.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exponential growth</td>
<td>−9800.454</td>
<td>2.20 × 10⁻⁴</td>
<td>5.15 × 10⁻⁵–4.23 × 10⁻⁴</td>
<td>202.552</td>
<td>48.56–447.30</td>
</tr>
</tbody>
</table>

*Substitutions per site per year.

Conclusions

On a larger scale, the phylogeographical pattern obtained for CFPHV appears to reflect movements of the host, whereas altered environments seem to be implicated in the current FP outbreaks and in the modern evolutionary history of CFPHV. Further work on the evolution of
viruses infecting threatened organisms, including additional analyses of CFPHV, should be conducted. Including more geographical areas and a wider time frame will no doubt improve our understanding of CFPHV evolution and expansion routes. More interestingly, a population structure analysis will not only improve parameter estimates but will also open up the opportunity for epizootic studies and transmission rate estimations.

METHODS

DNA extraction and amplification. Biopsies of fibropapilloma tissue and skin were collected from nine green turtles at Puerto Rico, between 2005 and 2010, and from six green turtles at Principe Island, Gulf of Guinea, in 2009. DNA was extracted using a DNeasy Blood & Tissue kit (Qiagen) and eluted in a final volume of 50 µl per sample. PCR primers used for the amplification of each target gene sequence have been published previously for the DNA polymerase, UL34, glycoprotein B and UL18 genes (Ene et al., 2005; Greenblatt et al., 2005; Quackenbush et al., 2001). Amplifications were performed in a total volume of 8 µl, with 1 µl lesion genomic DNA (including both target and green turtle DNA) at a concentration of ~10 ng µl⁻¹, 4 µl Fermentas DreamTaq PCR Master Mix or SPriime Master Mix, 2.2 µl MilliQ water and 0.4 µl each primer at 10 µM. The DNA polymerase gene was amplified following the conditions described by Ng et al. (2009). PCR conditions for the glycoprotein B and UL34 genes were 5 min at 94 °C, followed by 35 cycles of 94 °C for 30 s, 57 °C for 30 s and 72 °C for 45–160 s, with a final hold at 72 °C for 5 min (Greenblatt et al., 2005). The same conditions but with an annealing temperature of 58 °C were used for amplification of the UL18 region. PCR products were purified with ExoSAP-IT (Affymetrix) and sequenced.

Sequence data mining and alignment. Partial sequences of the DNA polymerase (483 bp), UL18 (1212 bp), UL34 (861 bp) and glycoprotein B (2486 bp) genes of CFPHV from nine geographical areas [Florida, Barbados, Puerto Rico, San Diego (CA), Hawaii, Mexico (Pacific coast), Costa Rica (Pacific coast), Australia (Pacific coast) and Gulf of Guinea (Principe Island)] [Fig. 1] were acquired from GenBank. Samples were collected from 1990 to 2010, and came from four host species: green sea turtle (Chelonia mydas), loggerhead sea turtle (Caretta caretta), olive ridley sea turtle (L. olivacea) and Kemp’s ridley sea turtle (L. kempii). Turtles carrying tumours were for the most part captured or found stranded in neritic bays. In total, 30 samples were unique per species and/or per year, but not all of these were represented for the four genes (see Table S1 for sample information).

Nucleotide sequences were individually translated to protein sequences and aligned using CLUSTAL W (Thompson et al., 2002), implemented in MEGA4.0 (Tamura et al., 2007), with the following parameters for multiple alignment: gap opening penalty=3, gap extension penalty=1.8, keeping the other default parameters (Hall, 2007). Protein alignments were uploaded onto the online interface of PAL2NAL (Suyama et al., 2006), together with nucleotide sequences, to retrieve the correct codon alignment for each gene. Each codon alignment was then screened manually using Bird View in Mesquite (Maddison & Maddison, 2001) for potential misalignments. A supermatrix of the four genes was built and exported both as NEXUS and PHYLIP files for further analyses; the same file formats were created for each individual gene alignment. Recombination in our dataset was assessed through the comparison of gene trees and using the rdp3 software package (Martin et al., 2010), which includes a range of different recombination detection methods that do not require prior identification of a non-recombinant set of reference sequences. We used the RDP (Martin & Rybicki, 2000), BOOTSCAN (Martin et al., 2005), GENECONV (Padidam et al., 1999), MAXCHI (Maynard Smith, 1992; Posada & Crandall, 2001), CHIMAERA (Posada & Crandall, 2001) and SISCAN (Gibbs et al., 2000) methods, keeping default parameters.

Phylogenetic reconstruction. Maximum-likelihood phylogenies for each gene alignment were inferred via the online interface of PhyML version 3.0 (Guindon et al., 2005) using bootstrap resampling (Felsenstein, 1985) for branch support evaluation with 1000 replications and using the most suitable model of substitution as evaluated through Akaike Information Criteria in the HYPHY program (Pond & Muse, 2005). Sampling year was used as calibration time to infer the phylogeny and evolutionary rate of CFPHV, using the MCMC inference methods available in BEAST version 1.6 (Drummond & Rambaut, 2007). MCMC analyses included γ-distributed rate heterogeneity among sites and partition into codon positions, which performs better for datasets with multiple genes (Drummond et al., 2007; Shapiro et al., 2006). Genealogy was estimated under a relaxed molecular clock model, enabling terminal branches to have faster rates (Ho & Larson, 2006; Ho et al., 2005b). The distribution of rates was determined as the UCED, as this method has been shown to fit virus datasets well (Chen & Holmes, 2006; Ho et al., 2005b; Smith et al., 2009). MCMC analyses were run using the GTR (Tavaré, 1986) and HKY (Hasegawa et al., 1985) models of nucleotide substitution, with each model run under the assumption of constant population size and exponential growth (Ho et al., 2005a).

For each prior combination, two independent MCMC analyses were run for 100 million generations, subsampling every 10 000 generations. After a 10% burn-in, the chains were combined and the convergence of chains was examined using Tracer version 1.5 (Rambaut & Drummond, 2003). Marginal posterior parameter means, their associated 95% HPD and the effective sample size for each parameter were analysed to assure statistically robust parameter estimates (Drummond et al., 2002). After confirming MCMC convergence and congruence among parameter estimates from different prior sets, summary trees were created using TreeAnnotator version 1.6.0 (Rambaut & Drummond, 2009) and edited in FigTree version 1.3.1 (Rambaut & Drummond, 2010). MCMC analyses were also conducted for each individual gene alignment, using a best-fit substitution model and through the procedure described above.

Parameter estimates and phylogenies from gene alignments were compared for congruence and to assess the presence of recombinant sequences.

ACKNOWLEDGEMENTS

We acknowledge Dr Tierry Work, Dr George Balazs, Dr James Casey and Dr Anny Chaves for useful information on sampling dates of sequences used in this study. We thank Carlos E. Diez for his support in sample collection at Puerto Rico, conducted under permits by the National Marine Fisheries Service (NMFS) – National Oceanic and Atmospheric Administration (NOAA) (Permit nos 1253 and 1518-01) and by the Department of Natural and Environmental Resources of Puerto Rico (DNER-PR) (06-EPE-016) and funded by the DNER-PR, US NMFS (NMFS-NOAA, Section 6, grant #NA08NMF4720436), US-Fish and Wildlife Service (USFWS), Sociedad Chelonia and WIDECAST. Sample processing was funded by the US Environmental Protection Agency (US-EPA). Sample collection at the Gulf of Guinea was conducted under the Program SADA funded by the Lisbon Oceanarium, Portugal, and sample processing was funded by the Interdisciplinary Research Center for Animal Health of the Faculty of Veterinary Medicine of the Technical University of Lisbon (FMV/UL).

1042
REFERENCES


