LETTER

Phylogenetic concordance analysis shows an emerging pathogen is novel and endemic

Andrew Storfer,[1]* Michael E. Alfaro,[1] Benjamin J. Ridenhour,[1] James K. Jancovich,[2] Stephen G. Mech,[1,3] Matthew J. Parris[4] and James P. Collins[5]

[1]*School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA*
[2]*Molecular and Cellular Biology Program, School of Life Sciences, The BioDesign Institute, Arizona State University, Tempe, AZ 85287-4601, USA*
[3]*Biology Department, Albright College, Reading, PA 19612-5234, USA*
[4]*Department of Biology, University of Memphis, Memphis, TN 38152, USA*
[5]*School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA*
Correspondence: E-mail: astorfer@wsu.edu

Abstract

Distinguishing whether pathogens are novel or endemic is critical for controlling emerging infectious diseases, an increasing threat to wildlife and human health. To test the endemic vs. novel pathogen hypothesis, we present a unique analysis of intraspecific host-pathogen phylogenetic concordance of tiger salamanders and an emerging *Ranavirus* throughout Western North America. There is significant non-concordance of host and virus gene trees, suggesting pathogen novelty. However, non-concordance has likely resulted from virus introductions by human movement of infected salamanders. When human-associated viral introductions are excluded, host and virus gene trees are identical, strongly supporting coevolution and endemism. A laboratory experiment showed an introduced virus strain is significantly more virulent than endemic strains, likely due to artificial selection for high virulence. Thus, our analysis of intraspecific phylogenetic concordance revealed that human introduction of viruses is the mechanism underlying tree non-concordance and possibly disease emergence via artificial selection.

Keywords

INTRODUCTION

Emerging infectious diseases are widely recognized as a threat to public health and are increasingly appreciated as having major effects on biological communities (McCallum & Dobson 1995). In some cases, pathogens can even cause host extinctions (DeCastro & Bolker 2005). Emerging pathogens are either recently introduced to new regions or naïve hosts, or are already endemic but have increased in impact because of environmental changes or genetic changes in the host or pathogen (Daszak *et al.* 2000). Distinguishing whether emerging infectious diseases are novel or endemic is critical because each scenario necessitates different avenues for further research and possible mitigation strategies. In the case of emerging endemic pathogens, research may focus on identification and mitigation of possible environmental cofactors or assessing whether recent genetic changes increased pathogen virulence or decreased host resistance (Daszak *et al.* 2000; Jancovich *et al.* 2005; Rachowicz *et al.* 2005). In the novel pathogen case, research may focus on reasons for spread, such as: pathogen host switching, recent introductions of infected hosts to new areas, or pathogen range expansion. Efforts may then focus on identifying the source localities or species and controlling spread, such as for sudden acute respiratory syndrome (SARS; Li *et al.* 2005). However, distinguishing between the novel vs. endemic hypotheses is difficult for any pathogen and has thus rarely been accomplished because of difficulties gathering adequate data to identify the source of the pathogen, molecular information to disentangle the geographic relationships of pathogen strains, and dating the age of the pathogen (Rachowicz *et al.* 2005).

One way to determine whether pathogens are novel or endemic is to assess the extent to which they are coevolved with their host. Analyses of phylogenetic

concordance that use the degree of alignment of phylogenies to examine patterns of coevolution and cospeciation have yielded key insights into many interspecific host–parasite relationships (Brooks 1979; Hafner *et al.* 1994; Huelsenbeck *et al.* 2000; Page 2002). Interspecific assessment of phylogenetic concordance analysis is useful in discriminating between hypotheses of cospeciation (Brooks 1979; Hafner *et al.* 1994; Huelsenbeck *et al.* 2000; Page 2002; Smith *et al.* 2004) vs. host switching (Page 1994, 2002; Huelsenbeck *et al.* 1997, 2000) . However, it is usually not possible to distinguish the mechanisms underlying coevolutionary events themselves in interspecific studies because the time scale of speciation events is too broad. *Intra*specific analyses of host-pathogen phylogenetic concordance, however, can reflect more recent evolutionary patterns than interspecific studies. Although intraspecific phylogenetic concordance has rarely been assessed, such analyses provide the opportunity to investigate mechanisms underlying disease emergence. Concordance suggests a pathogen is coevolved and likely endemic, and environmental cofactors or genetic changes may underlie emergence. Conversely, non-concordance may indicate pathogen novelty or other reasons for emergence, such as decreased dispersal of infected hosts relative to uninfected hosts. It is also possible to have concordance in one part of the evolutionary trees and non-concordance in another, resulting in apparent pathogen novelty in some areas and endemism in others.

Notably, novel or endemic pathogens are among the leading hypotheses for the global decline of amphibians, a key indicator group in the global loss of biodiversity (Collins & Storfer 2003; Daszak *et al.* 2003; Stuart *et al.* 2004; Rachowicz *et al.* 2005; Jancovich *et al.* 2005). We analysed phylogenetic concordance between tiger salamanders (*Ambystoma tigrinum*) and their *Ranavirus* (*Ambystoma tigrinum* virus; ATV) strains throughout western North America (Fig. 1) to test whether ATV is novel or endemic. Ranaviruses are generally considered emerging pathogens because of their recent implication in the increase of global epizootics of ectothermic vertebrates (Chinchar 2002).

The tiger salamander-ATV system, in particular, is a model for studying the basis for disease emergence (Collins *et al.* 2004). Molecular evidence suggests that ATV may be introduced by humans into new areas via transport of infected salamanders used as fishing bait (Jancovich *et al.* 2005). Thus, possible artificial introduction of virus strains and consequent host switching events makes this system ideal for testing patterns of phylogenetic concordance and the mechanism underlying disease emergence. First, we assessed concordance between intraspecific pathogen and host phylogenies as one of the first to test the novel vs. endemic pathogen emergence hypotheses. We then conducted a cross-infection experiment to test whether a
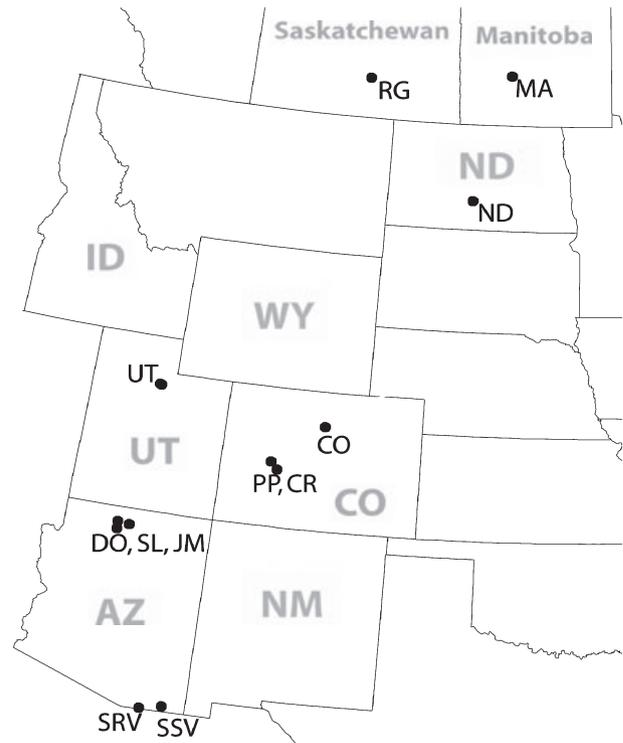

Figure 1 Sampling sites in US and Canada where salamander die-offs have occurred (see text for abbreviations). Viruses were isolated in pure culture from diseased animals at each site and salamander tissue was taken for genetic analyses.

putatively introduced novel bait strain is more virulent than putatively endemic virus strains. Bait salamanders are maintained at artificially high densities, a condition under which high virulence is expected to evolve because of enhanced opportunities for transmission (Ewald 1994). Thus, we hypothesized that novel virus strains should be more virulent than coevolved ATV strains.

MATERIALS AND METHODS

Study system

Pathogen biology

Ranaviruses (Family: Iridoviridae; Genus: *Ranavirus*) are relatively large (> 100 kb), double-stranded DNA viruses with icosahedral shaped capsid proteins and infect ectothermic vertebrates worldwide (Chinchar 2002). Tiger salamander iridoviruses, now collectively called ATV, are monophyletic relative to other ranaviruses (Jancovich *et al.* 2005). ATV is highly infectious and disease is easily transmitted horizontally from sick to healthy animals via water exposure, injection, or direct contact (Jancovich *et al.* 1997). The host range of ATV was shown to include other ambystomatid salamanders and spotted newts (*Notophthalmus viridescens*), but not predacous insects such

as diving beetles (*Ditiscus* sp.), making salamanders the primary disease vector (Jancovich *et al.* 2001). Mortality rates are dosage-dependent (Brunner *et al.* 2005), and in natural populations, epizootics follow a density-dependent pattern which, at times appears to result in localized extirpation of at least larval populations (Brunner *et al.* 2004). However, metamorphosed salamanders can act as sublethal viral carriers across years, re-infecting ponds as they enter in spring to re-colonize and breed (Brunner *et al.* 2004).

Host biology

We studied five subspecies of tiger salamanders (Fig. 1): *A. t. stebbinsi* (found only in S. AZ), *A. t. mavortium* (with a broad geographic range that includes E. AZ, CO, WY, NE, KS, OK and TX), *A. t. nebulosum* (ranging from AZ, UT, CO to NM), *A. t. melanostictum* (found in CO, ID, WA, Alberta and Saskatchewan, Canada, WY, MT, N. and S. SD, and NE) and *A. t. diaboli* (found in Manitoba and Saskatchewan, Canada, ND, SD and WI) (Petranka 1998).

ATV has varied effects on tiger salamander population dynamics, from apparent declines to recolonization and persistence. Long-term studies of populations on the Kaibab Plateau (*A. t. nebulosum*) and in the San Rafael Valley, AZ (*A. t. stebbinsi*) reveal frequent epizootics. Kaibab sites, as well as sites in Manitoba, Canada are typically recolonized within 2 years, but in Saskatchewan, at least three populations have not recovered from ATV epizootics since 1997 (D. M. Schock, unpublished data). ATV is also among the possible causes of Sonora tiger salamander (*A. t. stebbinsi*) declines and endangerment under the US Endangered Species Act, which also include introduced *A. t. mavortium* used as fishing bait (Collins *et al.* 1988; Storfer *et al.* 2004) . Other subspecies, such as *A. t. mavortium* and *A. t. nebulosum* are abundant, and although they experience die-offs from ATV, they have not noticeably declined.

Tissue collection and DNA sequencing

Salamander tissue and virus samples were collected from 12 localities ranging from southern Arizona, USA on the Mexico border to Saskatchewan, Canada (Fig. 1): JMH ("Joe's Mud Hole"; AZ, Coconino Co.); DO ("Donut Tank"; AZ, Coconino Co.); UT ("Desolation Lake"; UT, Salt Lake Co.); SL ("Snipe Lake"; AZ, Coconino Co.); SRV ("Bodie Canyon"; AZ, Santa Cruz Co.); SSV ("White Water Draw"; AZ, Cochise Co.); PP ("Pat's Pond"; CO, Gunnison Co.); CR, ("Cunningham Reservoir"; CO, Gunnison Co.); CO ("Mud Lake"; CO, Boulder Co.); ND ("Dogtown Reservoir"; ND, Burleigh Co.); MA ("Manitoba"; Manitoba, Canada); RG ("Regina Ranavirus"; Saskatchewan, Canada) (see Appendix S1 for detailed locality information).

Approximately 1 cm^2 of salamander tail was clipped from each individual (10 individuals per site) and stored in 95% EtOH before returning to the laboratory. Prior to DNA extraction and analysis, tissue was stored in a −20 °C freezer. Virus strains were given locality names, plus a "V" at the end to indicate the virus was collected from that locality. Virus was grown using standard cell culture techniques for EPC (*Epithelioma papilloma cyprini*) cells followed by a modified Hirt DNA extraction protocol (Jancovich *et al.* 2005), which is a modified phenol–chloroform extraction (Sambrook *et al.* 1989). Both viral and salamander DNA was amplified using the polymerase chain reaction (PCR) using primers for the region of interest. Then, PCR products were purified and sequenced following standard protocols for Applied Biosystems (ABI) using Big Dye 3.1 and an ABI 377 automated sequencer. Sequences were aligned using Clustal X software (Thompson *et al.* 1997).

We sequenced 1530 bp of tiger salamander mitochondrial DNA from the control region and adjacent insert region using PCR primers from Shaffer & McKnight (1996) and cytochrome B using primers from Steele *et al.* (2005). We sequenced 1086 bp of viral DNA, including the major capsid protein and two non-coding regions using PCR conditions in Jancovich *et al.* (2005). We chose the genes to be sequenced in the salamanders because they are conserved evolutionarily (Shaffer & McKnight 1996) and thus there is no intrapopulation variation that could confound results. The major capsid protein in ATV is also highly conserved (Jancovich *et al.* 2005), and thus the two intergenic spacers were also sequenced to provide phylogenetic signal. Previous studies have indicated no intrapopulation variation in these viral markers (Jancovich *et al.* 2005). Partition homogeneity tests (Farris *et al.* 1994, 1995) were performed on the two mtDNA host sequences and the three viral sequence segments to assess whether they suggested similar gene tree topologies. There were no differences in phylogenetic signal among the two host or three virus DNA sequences and they were thus analysed together for the reconstruction of host or virus gene trees, respectively (Farris *et al.* 1994, 1995).

Phylogenetic analyses

Our data consisted of two matrices of thirteen taxa for each host and virus-including outgroups chosen based on their close relationships to the ingroup taxa (*Ambystoma californiense* for the host tree and Frog virus 3, the type *Ranavirus* isolate, for the virus tree). All data sets were analysed with PAUP* 4.0b10 (Swofford 2000) and Mr. Bayes 3.1 (Huelsenbeck & Ronquist 2001) using Bayesian, parsimony and maximum likelihood methods. We performed Bayesian analyses using the program MrBayes 3.1 (Huelsenbeck &

Ronquist 2001). We used the default (flat) priors for rate parameters (Dirichlet 1,1,1,1,1,1), branch length (Exp 10), alpha parameter of the gamma distribution for rate heterogeneity (Uniform 0.1, 50), proportion of invariant sites (uniform 0, 1), base frequencies (Dirichlet 1,1,1,1) and tree topology parameters (uniform over all possible topologies). We used Metropolis coupling (running of simultaneous Monte Carlo Markov Chain samplers that are more likely to accept proposed moves to lower-likelihood states) with one cold chain (at the correct target density) and three heated chains (at higher densities and thus able to reach local optimality peaks more easily; heating parameter = 0.2) to help prevent the sampler from become trapped in one region of the tree space. We ran 20 million generation Markov chains, sampling every 2000 generations (total of 100 000 trees). We assessed convergence on an optimal tree reconstruction visually by first looking for a plateau in likelihood values. Second, after repeating each analysis and discarding the first 5000 sampled trees as burn-in, we plotted posterior probabilities from the first two runs against one another to assess variation in the confidence estimate. For virus data sets we defined two data partitions: one for the sequence characters and a second (morphological) containing two characters that described the state of the virus matrix indels (i.e. present or absent). The variable option was selected for the morphological partition to ensure the selection of only variable characters. Details of maximum likelihood (ML) and parsimony analyses can be found in Appendix S2.

Concordance testing

Four sets of analyses were used to test salamander and virus tree concordance (first with the full data sets and second with pruned data sets with host switches removed). We conducted four sets of analyses to determine consistency of results among tests, because each test has strengths and weaknesses (see Page 2002). Consistent results would thus provide the best evidence for tree concordance or non-concordance. The four tests included a modified partition homogeneity test (Farris et al. 1994, 1995; Swofford 2000; Johnson et al. 2001), reciprocal Shimodara-Hasegawa tests (Shimodara & Hasegawa 1999), reciprocal Swofford–Olsen–Waddell–Hillis (SOWH) (Swofford et al. 1996; Goldman et al. 2000) tests (i.e. parametric bootstrapping), and a comparison of the Bayesian credible intervals of both salamander and virus trees. As the Shimodara–Hasegawa test is viewed as statistically conservative (Goldman et al. 2000), it will be discussed here. The remaining three tests yielded consistent results, and the details of each are presented in supplemental methods. We performed reciprocal Shimodara–Hasegawa tests (Shimodara & Hasegawa 1999) to compare the virus and salamander

topologies on both data sets using RELL optimization and 1000 bootstrap replicates (Rambaut & Grassly 1997) in PAUP* 4.0b10 (Swofford 2000). This model-based test determines whether the fit of the host data to the virus tree is significantly worse than the fit of the host data to the host tree (or vice versa).

Molecular clock

To test for further evidence of coevolution, we analysed whether speciation times between host and virus were congruent (Douady et al. 2003). The first step in this analysis is to test for a violation of a molecular clock; if both host and pathogen are evolving according to a molecular clock, we can then test whether speciation times of congruent nodes are concordant. We tested for violation of a molecular clock on both full and pruned data sets by comparing ML scores for unconstrained topologies to those where a molecular clock was enforced (for both virus and salamander trees) using likelihood ratio tests (Felsenstein 1981). Also in the pruned data sets, we used TREEEMAP 1.0b (Page 1995) to test for significance of the number of cospeciation events with a randomization test using 1000 replicates. Further, to test for strict cospecation, we used a randomization test with 10 000 replicates in TREEEMAP 1.0b to test for a significant correlation of coalescence times of both host and virus in the pruned datasets.

Infection experiment

We conducted a cross-infection experiment to test whether putatively introduced virus strains (i.e. host-switches) are more virulent (i.e. cause more mortality), than putatively coevolved strains (i.e. concordant nodes) as indicated by the phylogenetic concordance analyses. To control the genetic background of the host, and thus test for viral differences while limiting variation in the virus-host interaction, we obtained 120 eggs from each of four inbred, full-sibship families of laboratory-bred Ambystoma tigrinum. Family lines originated from animals collected at locations along the Mogollon Rim (Coconino County, AZ, USA). These families were not originally from a locality used in this study; rather, they were chosen to control host genetic variability because they are highly inbred as a result of being maintained in captivity and inbred for over 20 years.

Eggs and larvae were reared in an environmental chamber on a 12:12-h light:dark cycle at a constant temperature of 20 ± 1 °C. Eggs were separated upon laying and reared individually in round, polyethylene containers (12.7 × 7.6 cm) containing 500 ml of artesian spring water dechlorinated with Reptisafe® (Zoo Med Laboratories, San Luis Obispo, CA, USA) and aerated for at least 24 h before each water change. Water was changed weekly and larvae were fed 0.015 g (dry weight of eggs) of hatched brine

shrimp every other day for 6 weeks and subsequently fed blackworms every 3 days *ad libitum*. Larvae were reared until 12 weeks of age prior to infection; at this age they exhibit basic immunocompetence as evidenced by their ability to produce immunoglobulins (Fellah *et al.* 1992).

We used a fully factorial 4 × 6 randomized block design, whereby 20 larvae from each of the four families were assigned to one of six virus treatments: four endemic viral strains as indicated by the phylogenetic concordance analysis (UTV, CRV, DOV, SLV) one putatively introduced strain (SRV) or a control. The "endemic" strain for these hosts is SLV due to its close geographic proximity to their source. Larvae in viral treatments were randomized and subsequently exposed to $10^{3.5}$ PFU ml^{-1} (sufficient to induce infection in previous experiments; Jancovich *et al.* 2001; Brunner *et al.* 2005) of one of the five virus strains (suspended in cell culture medium) for three days. Larvae in the control treatment were exposed to an equivalent volume of cell culture medium alone. Animals were then observed for a minimum of 4 weeks (consistent with larval die-off periods in previous experiments; Brunner *et al.* 2004, 2005) and until no larvae experienced mortality for 3 days, which resulted in conclusion at day 35. We estimated percentage mortality resulting from each virus strain as a measure of virulence, with families as replicates and arcsin-square root transformed data to meet normality assumptions of analysis of variance (ANOVA) models. We conducted an ANOVA using SYSTAT 11.0 to test the main effects of virus strain on mean mortality, with Fisher's LSD test of multiple comparisons with assess differences among treatment means.

RESULTS

We constructed host and pathogen gene trees of tiger salamanders and sympatric ATV strains from 12 localities within the Western US and Canada (Fig. 1) using maximum parsimony, maximum likelihood and Bayesian methodologies. Within a locality, all salamander DNA sequences were identical. Previous studies also indicated no within locality ATV genetic variation (Jancovich *et al.* 2005; A. T. Storfer, unpublished data) and thus, each locality represented a single operational taxonomic unit (OTU) on each of the host and pathogen trees. All three methods yielded identical topologies for both salamander and virus trees; the Bayesian tree with support values is shown in Fig. 2a. Detailed results of maximum likelihood and parsimony analyses can be found in Appendix S2.

Bayesian analyses

Visual inspection revealed that all Bayesian analyses appeared to reach stationarity within the first million
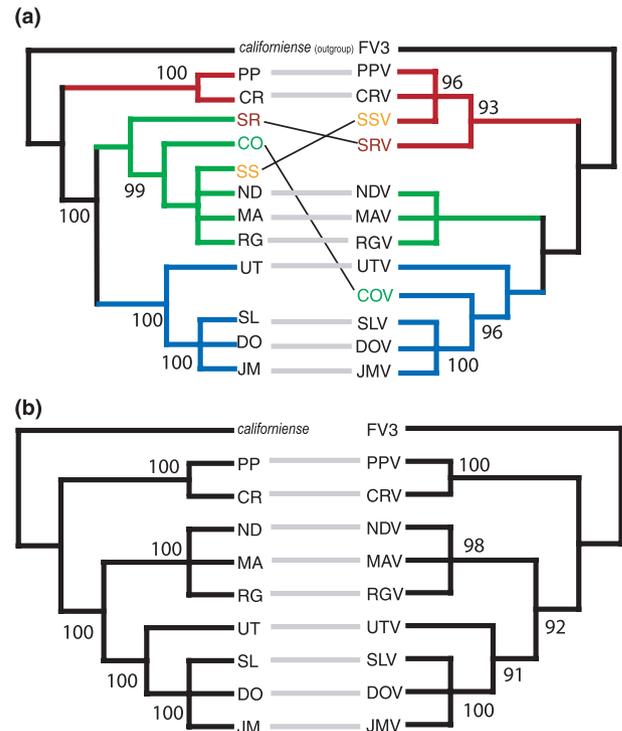

Figure 2 (a) Tiger salamander mtDNA gene genealogy (1530 bp) and gene genealogy from concatenated viral sequences (1,086 bp). The most appropriate model of evolution for salamander (GTR + I, −lnL = 2873.60, AIC = 5675.21) and viral (K81uf + I, −lnL = 1842.98, AIC = 3697.98) data sets were used. Maximum likelihood and Bayesian analysis were performed on both data sets, with both methods yielding identical tree topologies. Numerical values on nodes represent Bayesian posterior probabilities. Outgroups were chosen based on their close relationships to ingroup taxa. Trees were significantly non-concordant. Host switching events are noted in color. (b) Tiger salamander and ranavirus gene genealogies with human-implicated host switches removed. The most likely model of evolution remained K81uf + I (−lnL = 1786.50, AIC = 3583.01) for the virus and GTR + I (−lnL = 2096.21, AIC = 5610.42) for the host. Trees and node values were generated with same analyses as in (a). Trees are completely concordant.

generations, and thus we discarded the first 5000 sampled trees as burn-in. In all cases, the Bayesian consensus topology was identical to the maximum likelihood topology, though the posterior probabilities for clades were generally higher (Figs. 2a,b). Even though it has become common in recent studies to find that posterior probabilities are higher than maximum likelihood bootstrap values (Alfaro *et al.* 2003; Erixon *et al.* 2003; Douady *et al.* 2003), increased support for relationships within the virus topologies is also due to the inclusion of two morphological characters that show no homoplasy on the Bayesian consensus tree.

Concordance testing

Concordance

All four sets of independent analyses strongly supported non-concordance of salamander and virus gene trees when the full dataset was analysed. Shimodara–Hasegawa tests showed significant incongruence between the full salamander and virus data sets based on maximum likelihood topologies. When the virus tree was constrained to the most likely host tree, the likelihood score (1873.45) was significantly worse than the most likely virus tree (1842.98, $P = 0.002$). Similarly, when the salamander tree was constrained to the most likely virus tree, the likelihood score significantly decreased (3083.26) relative to the best host tree score (2873.60, $P = 0.001$). Results of the remaining three tests are presented in Appendix S2.

Molecular clock

For the full data sets, we could reject a molecular clock for both virus ($-\ln L = 1859.89$ with clock enforced; $-\ln L = 1842.98$ without clock enforced; $P < 0.05$) and salamander ($-\ln L = 2877.69$ with clock enforced; $-\ln L = 2873.60$ without clock; $P < 0.01$) data sets. However, we were unable to reject a molecular clock for either of the pruned data sets. For the pruned virus data set, the ML topology with a molecular clock enforced had a score of 1788.72, which was not significantly better than the unconstrained topology ($-\ln L = 1786.60$; $P > 0.05$). For the pruned salamander data set, the maximum likelihood score with the molecular clock enforced was 2799.73 vs. 2796.21 for the full dataset ($P > 0.05$).

When the three host switches were removed from the analysis pruned topologies were identical, and the randomization test in TREEMAP showed that the number of cospeciation events was significant ($P < 0.001$), and thus highly unlikely that the two trees were concordant by chance. There was a strong correlation of host and pathogen coalescence times ($r^2 = 0.859$), although we could not reject the hypothesis that strict cospeciation has not occurred ($P = 0.110$).

Infection experiment

The ANOVA indicated significant differences among mean mortality rates across the different viral treatments ($F_{5,18} = 17.116$, $P < 0.0001$; Fig. 3). A Fisher's LSD multiple comparison test indicated significant differences between individual treatments within the study with the control treatment resulting in no mortality and the putatively introduced viral strain (ATV) resulting in significantly higher mean mortality (60%) than all other strains ($P < 0.001$; Fig. 3). There was also variation among the putatively
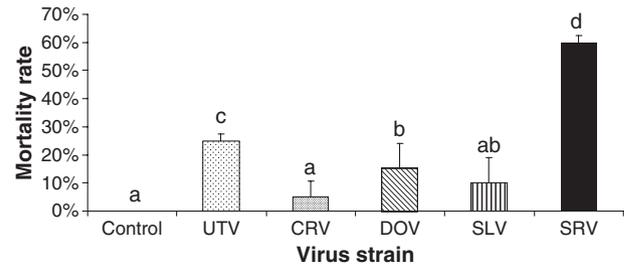

Figure 3 Mortality rates (percentages) of four families of laboratory-bred salamanders infected using five ranavirus strains and a control. According to phylogenetic analyses in Fig. 1, strains UTV, CRV, DOV and SLV coevolved with their hosts, while strain SRV represents a host switch. Shown are means + SE The overall ANOVA was significant for the main effect of virus ($F_{5,18} = 17.116$, $P < 0.0001$). Different letters above bars indicate significant differences among means.

coevolved strains, with UTV having the highest mortality (30%; $P < 0.01$).

DISCUSSION

Our results support both the novel and endemic hypotheses as explanations for disease emergence in western North American tiger salamander populations. That is, ranaviruses are likely endemic in most tiger salamander populations, but infected bait salamanders are moved by humans, resulting in apparent host-switches. Initial analyses showed that host and pathogen trees were significantly different and non-concordant, suggesting pathogen novelty (Fig. 2a). Non-concordance appears to result from three host switches, corresponding to two localities in southern Arizona (SS and SR) and one locality in Colorado (CO, Figs. 1, 2a).

We previously documented that these three host switches likely resulted from human movement of infected, non-native salamanders, probably used as fishing bait (Collins 1981; Storfer *et al.* 2004; Jancovich *et al.* 2005) . We formerly isolated ATV from infected bait shop salamanders on two separate occasions, and geographic patterns of virus genotypes are consistent with documented anthropogenic movement of bait salamanders (Jancovich *et al.* 2005). With regard to the three host switches observed in the present study, we recently showed hybridization of bait salamanders (non-native *A. t. mavortium*) and the native Sonora tiger salamander, *A. t. stebbinsi* at one of these three sites (SR, Fig. 1; Storfer *et al.* 2004). Salamanders from the other southern Arizona site (SS) have been documented as human-introduced *A. t. mavortium* (Collins 1981; Collins *et al.* 1988). The third site representing a host-switch (CO) is in an easily accessible state park in Colorado, and we showed that this strain is nearly identical genetically to an

ATV strain isolated from a bait colony (Jancovich *et al.* 2005).

Curiously, one might expect that if the three host-switch ATV strains were maintained in bait shops, that they would be more phylogenetically similar (e.g. monophyletic) than shown herein. However, salamanders from bait shops throughout western US states have tested positive for ATV (A. M. Picco, unpublished data). Thus, the geographic variation in the associated viral strains and resulting potential genetic variation is consistent with the observed non-monophyly of the three putatively introduced ATV strains.

Because of the likely human involvement in moving all three host-switch strains, we removed them from the analysis and re-tested for concordance. The resulting pruned host and virus gene genealogies are identical, strongly suggesting host-pathogen coevolution in the remaining localities (Fig. 2b). It is highly improbable that the two pruned trees are concordant by chance alone, because a post hoc-randomization analysis showed a significant number of cospeciation events. Further support for coevolution came from an observed strong correlation between host and pathogen coalescent times ($r^2 = 0.859$; Huelsenbeck *et al.* 1997, 2000). In addition, the pruned data sets had higher Bayesian support at each node than the full data set, a result consistent with human movement of infected salamanders where host-switches appear to have occurred. Collectively, these analyses provide strong evidence of salamander-virus coevolution in all but three localities.

Previous long-term ecological and experimental data also support a coevolutionary relationship. ATV is easily spread by cannibalism (Jancovich *et al.* 1997), and there is a negative correlation between disease frequency and cannibal frequency among salamander populations throughout Arizona (J. P. Collins, unpublished data; Pfennig *et al.* 1991). Common garden experimental data provide evidence that these patterns are likely the result of past selection by ATV against cannibalism (Pfennig *et al.* 1991; Parris *et al.* 2005).

In light of the strong phylogenetic, ecological and experimental evidence for coevolution, we investigated the mechanism for disease emergence. We tested the hypothesis that the artificially high densities maintained in bait colonies select for high virulence of bait associated strains, leading to their emergence. Theory suggests that in high density host environments, high pathogen virulence can evolve because virulence is associated with within-host growth rate and consequently transmission (Ewald 1994). In contrast, high virulence can not be maintained below a critical host density because transmission becomes rare and the pathogen will likely burn itself out (Lenski & May 1994). Thus, we predicted higher virulence of bait virus strains than endemic strains because tiger salamander densities are artificially high

in bait shops and transmission is virtually guaranteed. A laboratory experiment supported this prediction; an introduced strain (SRV; Fig. 3) was significantly more virulent (i.e. caused higher mortality) than endemic strains. It is noteworthy that the introduced strain used in this experiment (SRV) is implicated in the decline of the endangered Sonora tiger salamander (*A. t. stebbinsi*), supporting the virulence of this strain in natural populations (Collins *et al.* 1988).

In general, intraspecific tests of non-concordance (see Fig. 2a) or concordance (see Fig. 2b) of host and pathogen phylogenies afford a means to test the novel vs. endemic hypotheses and have broad implications for guiding further emerging disease research and control. This study is one of the first investigations of phylogenetic concordance *within* species and provides the distinct advantage relative to interspecific studies of exploring the mechanistic underpinnings for discrepancies in host-pathogen phylogenies. We propose that the mechanism underlying non-concordance of salamander and virus phylogenies is artificial movement of bait salamanders infected with ATV. Such virus introductions have likely altered the coevolutionary process and resulted in disease emergence at localities where humans introduced novel, more virulent virus strains.

Analyses of intraspecific phylogenetic concordance can provide critical information about geographic locations where host-pathogen coevolution has occurred and the pathogen is likely endemic, in contrast to places where pathogen strains may have been introduced. Distinguishing between endemic and introduced pathogens allows researchers to target appropriate host populations for possible intervention. In the case of emerging diseases that are endemic, researchers can investigate: whether there are evolved host defenses that can be used as potential treatments in other host populations, whether there are environmental cofactors of disease emergence, or whether there are recent genetic changes in the host or pathogen. In cases where pathogens are found to be novel, research could focus on identifying locations where pathogens are introduced and perhaps most virulent, or understanding the conditions under which the pathogen has spread, such as via increased movement of infected hosts or host switches to naïve species. In addition, if pathogens are found to be spread recently, such as SARS (Li *et al.* 2005), research can focus on tracing them to their source to control further introductions.

comments that improved the manuscript. This work was supported by NSF DEB 0213851 (to JPC and AS), DEB 0548415 (to AS) and DEB 0445453 (to MEA).

REFERENCES

Alfaro, M.E., Zoller, S. & Lutzoni, F. (2003). Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence *Mol. Biol. Evol.*, 20, 255–266.

Brooks, D.R. (1979). Testing the context and extent of host-parasite coevolution. *Syst. Zool.*, 28, 299–307.

Brunner, J.L., Schock, D.M., Collins, J.P. & Davidson, E.W. (2004). Intraspecific reservoirs: complex life history and the persistence of a lethal ranavirus. *Ecology*, 85, 560–566.

Brunner, J.L., Richards, K. & Collins, J.P. (2005). Dose and host characteristics influence virulence of ranavirus infections. *Oecologia*, 144, 399–406.

Chinchar, V.G. (2002). Ranaviruses (Family: Iridoviridae): emerging cold-blooded killers. *Arch. Virol.*, 147, 447–470.

Collins, J.P. (1981). Distribution habitats and life history variation in the tiger salamander *Ambystoma tigrinum* in east central and southeast Arizona, USA. *Copeia*, 1981, 666–675.

Collins, J.P. & Storfer, A. (2003). Amphibian declines: sorting the hypotheses. *Divers. Distrib.*, 9, 89–98.

Collins, J.P., Jones, T.R. & Berna, H.J. (1988). Conserving genetically distinctive populations: the case of the Huachuca tiger salamander (*Ambystoma tigrinum stebbinsi*). In: *Management of Amphibians, Reptiles and Small Mammals in North America* (eds Szaro, R.C., Severson, K.C. & Patton, D.R.). USDA Forest Service, Fort Collins, CO, pp. 45–53.

Collins, J.P., Brunner, J.L., Jancovich, J.K. & Schock, D.M. (2004). A model host-pathogen system for studying infectious disease dynamics in amphibians: Tiger salamanders (*Ambystoma tigrinum*) and *Ambystoma tigrinum* virus. *Herpetol. J.*, 14, 195–200.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2000). Emerging infectious diseases of wildlife-threats to biodiversity and human health. *Science*, 287, 443–449.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2003). Infectious disease and amphibian population declines. *Divers. Dist.*, 9, 141–150.

DeCastro, F. & Bolker, B. (2005). Mechanisms of disease-induced extinction. *Ecol. Lett.*, 8, 117–126.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, E.D. & Douzery, E.J.P. (2003). Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.*, 20, 248–254.

Erixon, P., Svennblad, B., Britton, T. & Oxelman, B. (2003). Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.*, 52, 665–673.

Ewald, P.W. (1994). *Evolution of Infectious Diseases*. Oxford University Press, New York, 298 pp.

Farris, J.S., Kallersjo, M., Kluge, A.G. & Bult, C. (1994). Testing significance of congruence. *Cladistics*, 10, 315–320.

Farris, J.S., Kallersjo, M., Kluge, A.G. & Bult, C. (1995). Constructing a significance test for incongruence. *Syst. Biol.*, 44, 570–572.

Fellah, J.S., Wiles, M.V., Charlemagne, J. & Schwager, J. (1992). Evolution of vertebrate IgM: complete amino acid sequence of the constant region of *Ambystoma mexicanum* μ chain deduced from cDNA sequence. *Eur. J. Immunol.*, 22, 2595–2601.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17, 368–376.

Goldman, N., Anderson, N.P. & Rodrigo, A.G. (2000). Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.*, 49, 652–670.

Hafner, M.S., Sudman, P.D., Villablanca, F.X., Spradline, T.A., Demastes, J.W. & Nadler, S.A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, 265, 1087–1090.

Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754–755.

Huelsenbeck, J.P., Rannala, B. & Yang, Z. (1997). Statistical tests of host-parasite cospeciation. *Evolution*, 51, 410–419.

Huelsenbeck, J.P., Rannala, B. & Larget, B. (2000). A Bayesian framework for the analysis of cospeciation. *Evolution*, 54, 352–364.

Jancovich, J.K., Davidson, E.W., Morado, J.F., Jacobs, B.L. & Collins, J.P. (1997). Isolation of a lethal virus from the endangered tiger salamander *Ambystoma tigrinum stebbinsi. Dis. Aquat. Org.*, 31, 161–167.

Jancovich, J.K., Davidson, E.W., Seiler, A., Jacobs, B.L. & Collins, J.P. (2001). Transmission of the *Ambystoma tigrinum* virus to alternative hosts. *Dis. Aquat. Org.*, 46, 159–163.

Jancovich, J.K., Davidson, E.W., Parameswaran, N., Mao, J., Chinchar, V.G., Collins, J.P. *et al.* (2005). Evidence for emergence of an amphibian iridoviral disease because of human-enhanced spread. *Mol. Ecol.*, 14, 213–224.

Johnson, K.P., Drown, D.M. & Clayton, D.H. (2001). A data based parsimony method of cophylogenetic analysis. *Zool. Scripta*, 30, 79–87.

Lenski, R.E. & May, R.M. (1994). The evolution of virulence in parasites and pathogens – reconciliation between 2 competing hypotheses. *J. Theor. Biol.*, 169, 253–265.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H. *et al.* (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science*, 310, 676–679.

McCallum, H. & Dobson, A. (1995). Detecting disease and parasite threats to endangered species and ecosystems. *Trends Ecol. Evol.*, 10, 190–194.

Page, R.D.M. (1994). Parallel phylogenies: reconstructing the history of host–parasite assemblages. *Cladistics*, 10, 155–173.

Page, R.D.M. (1995). *Treemap 1.0b*. University of Glasgow, Glasgow.

Page, R.D.M. (ed.) (2002). *Tangled Trees*. U. Chicago Press, Chicago, 350 pp.

Parris, M.P., Storfer, A., Collins, J.P. & Davidson, E.W. (2005). Pathogen effects on life history in tiger salamander (*Ambystoma tigrinum*) larvae. *J. Herpetol.*, 39, 366–372.

Petranka, J.W. (1998). *Salamanders of the United States and Canada*. Smithsonian Institution Press, Washington, DC, 587 pp.

Pfennig, D.W., Loeb, M.L.G. & Collins, J.P. (1991). Pathogens as a factor limiting the spread of cannibalism among tiger salamanders. *Oecologia*, 88, 161–166.

Rachowicz, L.J., Hero, J.M., Alford, R.A., Taylor, J.W., Morgan, A.T., Vredenburg, V.T. *et al.* (2005). The novel and endemic pathogen hypotheses: competing explanations for the origin of emerging infectious diseases of wildlife. *Conserv. Biol.*, 19, 1441–1448.

Rambaut, A. & Grassly, N.C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comp. Appl. Biosci.*, 13, 235–238.

Sambrook, J.E.F., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Press, New York.

Shaffer, H.B. & McKnight, M.L. (1996). The polytypic species revisited: genetic differentiation and molecular phylogenetics of the tiger salamander *Ambystoma tigrinum* (Amphibia: Caudata) complex. *Evolution*, 50, 417–433.

Shimodara, H. & Hasegawa, H. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.*, 16, 1114–1116.

Smith, V.S., Page, R.D.M. & Johnson, K.P. (2004). Data incongruence and the problem of avian louse phylogeny. *Zool. Scripta*, 33, 239–259.

Steele, C.A., Carstens, B.C., Storfer, A. & Sullivan, J. (2005). Testing hypotheses of speciation timing in *Dicamptodon copei* and *Dicamptodon aterrimus* (Caudata:Dicamptodontidae) . *Mol. Phylogenet. Evol.*, 36, 90–100.

Storfer, A., Mech, S.G., Reudink, M.W., Ziemba, R.E., Warren, J. & Collins, J.P. (2004). Evidence for introgression in the endangered Sonora Tiger Salamander, Ambystoma tigrinum stebbinsi (Lowe). *Copeia*, 2004, 783–796.

Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B..E., Rodrigues, A.S.L., Fischman, D.L. *et al.* (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306, 1783–1786.

Swofford, D.L. (2000). *PAUP* 4.00: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. (1996). Phylogenetic inference. In: *Molecular Systematics*. 2nd edn. (eds. Hillis, D.M., Moritz, C. & Mable, B.K.). Sinauer Associates, Sunderland, MA., pp. 407–514.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid Res.*, 25, 4876–4882.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Detailed Locality Information for virus and salamander tissue sampling sites (see Fig 1. for locations on a map).

Appendix S2 Detailed Phylogenetic Methods and Results.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/full/10.1111/j.1461-0248.2007.01102.x.

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.