

Spread of an introduced vector-borne banana virus in Hawaii

RODRIGO P. P. ALMEIDA,* GORDON M. BENNETT,* MANDY D. ANHALT,† CHI-WEI TSAI* and PATRICK O'GRADY*

*Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA, †Department of Plant and Environmental Protection Sciences, University of Hawaii at Manoa, Honolulu, HI 96822, USA

Abstract

Emerging diseases are increasing in incidence; therefore, understanding how pathogens are introduced into new regions and cause epidemics is of importance for the development of strategies that may hinder their spread. We used molecular data to study how a vector-borne banana virus, *Banana bunchy top virus* (BBTV), spread in Hawaii after it was first detected in 1989. Our analyses suggest that BBTV was introduced once into Hawaii, on the island of Oahu. All other islands were infected with isolates originating from Oahu, suggesting that movement of contaminated plant material was the main driving factor responsible for interisland spread of BBTV. The rate of mutation inferred by the phylogenetic analysis (1.4×10^{-4} bp/year) was similar to that obtained in an experimental evolution study under greenhouse conditions (3.9×10^{-4} bp/year). We used these values to estimate the number of infections occurring under field conditions per year. Our results suggest that strict and enforced regulations limiting the movement of banana plant material among Hawaiian islands could have reduced interisland spread of this pathogen.

Keywords: *Banana bunchy top virus*, emerging disease, invasive species, Nanoviridae, vector

Received 10 March 2008; revision received 29 July 2008; accepted 19 October 2008

Introduction

The potential and realized threats of emerging diseases of humans, domestic and wild animals, and plants have been well documented. Studies have now identified several factors driving the emergence of new diseases, such as climate and land-use change, ease of travel, international trade, and pathogen evolution (Woolhouse & Gowtage-Sequeria 2005). However, one of the main factors responsible for emerging and re-emerging diseases may be pathogen pollution, defined as the introduction of pathogens into new areas where susceptible hosts occur (Cunningham *et al.* 2003). Such pathogen invasions, primarily mediated by anthropogenic activities, are often identified as the factor responsible for most emerging diseases. In fact, it has been estimated that more than 50% of emerging plant diseases were driven by introduced pathogens (Anderson *et al.* 2004).

The determination that a newly introduced pathogen is the driving factor of a disease is often based on genetic analysis of the pathogen at its new location compared to other areas where it is found. Such an approach allows for

the identification of the probable origin of introductions. Recent evolutionary studies on the epidemics of West Nile virus and avian flu are examples of the successful use of phylogenetic analysis to determine the source of pathogens (Lanciotti *et al.* 1999; Janies *et al.* 2007). These and similar studies generate hypotheses concerning how pathogens are spread and provide helpful information on their ecology. However, because detection of emerging diseases occurs after introduction and initial pathogen spread, it is difficult to identify the underlying factors associated with pathogen invasion. These are often unknown and assumed to be a consequence of increased commerce, human transportation, and lack of guidelines to avoid the movement of infectious agents. Furthermore, it is not easy to demonstrate which anthropogenic factor may have driven a specific epidemic. Difficulties with determining the root cause of a given epidemic include alternative invasion hypotheses, high mobility of infected hosts, and the possibility of natural, albeit comparatively rare, spread of pathogens over large spatial scales (e.g. between continents). Nevertheless, DNA sequence-based data and analyses methods are proving to be powerful tools to reconstruct the spread of pathogens (e.g. Biek *et al.* 2006, 2007).

Correspondence: Rodrigo Almeida, Fax: (510) 643 5438;

E-mail: rodrigo@nature.berkeley.edu

Banana bunchy top virus (BBTV) is the aetiological agent of one of the most important banana diseases in the South Pacific and Southeast Asia (banana bunchy top disease, BBTD); infections result in bunched leaves and stunted, fruitless plants. BBTV is a multicomponent single-stranded DNA virus that is efficiently transmitted by an aphid vector, *Pentalonia nigronervosa* (Hemiptera, Aphididae), in a persistent, circulative and nonpropagative manner (Magee 1940; Hafner *et al.* 1995; Hu *et al.* 1996; Anhalt & Almeida 2008). There are no other means for natural dispersal of BBTV. However, as banana is vegetatively propagated, the spread of BBTV is possible by transportation of contaminated material originating from asymptomatic but infected plants. Natural spread of BBTV occurs slowly in plantations, and is often restricted to metres from a source plant (Allen 1978; Smith *et al.* 1998). Rouging of symptomatic plants and planting of virus-free plant material have been demonstrated to successfully control BBTV epidemics (Dale 1987). In 1989, BBTV was first detected on the island of Oahu, Hawaii (Conant 1992). Banana varieties cultivated in Hawaii are susceptible to BBTV, and *P. nigronervosa* has been established in the islands at least since 1924 (Zimmerman 1948). Although *P. nigronervosa* feeds on other plants, the Hawaiian BBTV strain infects only banana, reducing the importance of alternative hosts on spread of the virus (Hu *et al.* 1996). Despite its expected slow and limited spatial spread, BBTV is now present on all main Hawaiian islands, with the exception of Lanai.

BBTV in Hawaii is a good model system to study factors driving the dispersal of an emerging pathogen because of the spatial distribution of host plants (i.e. isolated archipelago system), all hosts are susceptible to the pathogen, the virus does not multiply in vectors, BBTV hosts are immobile, its Hawaiian strain does not infect other hosts, and aphid vectors were already present on all islands prior to BBTV's introduction. We used DNA sequence data to identify the region of origin of the Hawaiian BBTV infection and to determine the spatial genetic structure of BBTV within the Hawaiian Islands. Historical data on colonization of different islands in the archipelago was used to calibrate a molecular clock and calculate a rate of evolution of the Hawaiian BBTV. We compared the estimated evolutionary rate with that based on experimental data to test the reliability of such data and its applicability on estimating disease spread. We hypothesized that spread of BBTV in Hawaii could have occurred by (i) multiple introductions of BBTV through infective vectors or contaminated plant material, (ii) transportation of infected plant material between islands, with or without presence of aphids, and (iii) new interisland invasions by infected aphids. We used biogeography, evolutionary and demographic frameworks and experimental data to interpret our results and address our question of how this emerging pathogen spread in Hawaii.

Materials and methods

Virus samples

Samples of BBTV from symptomatic banana plants from all main Hawaiian Islands were collected between October and December 2005 (Table S1, Supporting Information; which includes the number of samples taken from each island). The only exception was Lanai where BBTV has not yet been reported. Although we tried to sample from as many populations as possible, samples were skewed towards the island of Oahu because of the high frequency of BBTV infections there. In addition, quarantine efforts by the state of Hawaii limited the availability of samples in certain locations, such as the island of Hawaii. Samples were freshly processed or maintained frozen in -80°C until DNA extraction. We extracted viral DNA using a QIAGEN DNeasy extraction kit (QIAGEN Inc.) following the manufacturer's instructions. Here we provide a brief description of the interisland spread of BBTV in Hawaii, based on information provided by the Hawaii Department of Agriculture for historical data (unpublished surveys, prepared by Larry Nakahara, Hawaii Department of Agriculture) and personal observations on the current status of the disease (R. Almeida, personal observations).

Oahu: BBTV was first observed on Oahu in 1989 on the east side of the island. The disease is currently widespread in commercial areas, roadways and backyards. Early eradication efforts failed to control BBTV spread.

Kauai: The disease was first identified here in 1997. BBTV is currently widespread in the island. Immediately after the 1997 introduction, eradication efforts were conducted but interpreted as unsuccessful, as years later (in 2000) BBTV was detected again on the island.

Hawaii (Big Island): BBTV was first observed on the west (Kona) side of the island in 1995. Eradication efforts reduced disease incidence but did not completely eliminate the pathogen from that area. In 2004, BBTV was identified on the east (Hilo) side of the island, the major banana-producing area in the State.

Maui and Molokai: The disease was first found in 2002 and 2005 on Maui and Molokai, respectively; BBTV spread is limited in these two islands.

Primer design and PCR amplification

We designed primers to amplify five of the six BBTV components (except DNA-2) based on sequence data available in GenBank (Table S2, Supporting Information). BBTV components from multiple strains were aligned

using MacClade version 4.06 (Maddison & Maddison 2002) and conserved regions were analysed for primer selection. Our primers amplified primarily the noncoding region of each viral component, with the exception of DNA-1, which only has a short noncoding region. Polymerase chain reaction (PCR) cycles used were 94 °C for 1 min, annealing temperature (variable, see Table S2) for 1 min, extension at 72 °C for 1 min, for 30 cycles, followed by a 5-min final extension at 72 °C. We determined the optimal annealing temperature for each primer pair using a temperature gradient block. PCR products were run on gels to confirm fragment size and purified using a PCR purification kit (QIAGEN). We sent samples for sequencing to the Greenwood Molecular Biology Facility (Pacific Biomedical Research Center, University of Hawaii at Manoa, Honolulu, HI) and to the University of California, Berkeley Sequencing Facility. Both strands were sequenced for all samples. Accession numbers for deposited sequences obtained in this study are EU292002–EU292141.

Origin of BBTV in Hawaii: a global approach

To ascertain the origin of the Hawaiian BBTV invasion, we constructed a phylogeny based on DNA-1 sequences (coding for replication initiation protein) with sequences from Hawaiian and worldwide BBTV infections (Table S1). Sequences from the other viral particles were not used in these analyses because they were (i) not determined for many non-Hawaiian taxa and (ii) designed to amplify noncoding regions, making them too rapidly evolving to address global biogeographical patterns. Alignments were checked manually in MacClade version 4.06 (Maddison & Maddison 2006), ambiguously aligned regions were removed from the analyses, and any missing sequence data were coded as 'Ns'. Because BBTV is the only member of the genus *Babuvirus* in the family Nanoviridae, outgroup selection was a nontrivial problem. Initial BLASTN (Altschul *et al.* 1990) searches retrieved *Cardamom clump virus* (CCV, AY485960), as the most significant match to BBTV (search statistics gave an 84% sequence match, *E*-value = 0). However, CCV is not currently recognized as a member of the family Nanoviridae. In order to more thoroughly test the appropriateness of CCV as an outgroup, we also investigated the phylogenetic relationships between the six recognized viruses (see Table S1 for taxa used) in Nanoviridae. Components coding the replication initiation protein (homologous to DNA-1 in BBTV) were aligned in ClustalW (Higgins *et al.* 1994) and exported to PAUP* (Swofford 2002) for analyses. A heuristic parsimony search was performed using 1000 independent searches (addition sequence = random; branch swapping = tree-bisection–reconnection, TBR); tree support was assessed using nonparametric bootstrap methods (100 replicates, other search settings as above). Results gave a single

tree (length = 1847, consistency index = 0.8733, retention index = 0.5821; tree not shown) with CCV and BBTV as sister taxa (BP = 100). Thus, CCV was chosen as suitable outgroup for subsequent analysis of BBTV phylogeny. This analysis was also performed using Bayesian methods under a Hasegawa–Kishino–Yano (HKY) model. A total of 10⁶ generations were run and convergence occurred rapidly (data not shown). CCV was selected as sister to the BBTV in both analyses.

A Bayesian phylogenetic approach, as implemented in BEAST version 1.4.2 (Drummond & Rambaut 2006), was taken. BEAST uses Markov chain Monte Carlo (MCMC) algorithm to search over tree space, integrating prior biological knowledge (e.g. dates), and a likelihood model of nucleotide substitution to sample the posterior distribution, averaging over all sampled topologies and desired parameters (Drummond *et al.* 2002; Drummond & Rambaut 2006a). We used CCV as the outgroup taxon for the BBTV data set. A general time reversible (GTR) model that allowed for a proportion of invariant sites (I) and a six-category, discrete gamma (Γ) shape distribution (Lanave *et al.* 1984; Yang 1996) was selected for the worldwide analysis using the Akaike information criterion (AIC) in ModelTest 3.8 (Posada & Crandall 1998; Posada 2006). The mutation rate was fixed at one so that the estimation of tree topology and branch lengths were a function of base-pair substitutions per site, following recommendations of Drummond *et al.* (2007). A total of five independent chains were run for 5 × 10⁶ generations under an uncorrelated relaxed clock model with a constant population size. Chains were sampled every 1000 generations, and burn-in was set to exclude the first 10% of sampled generations, thereby removing non-optimal tree topologies. Convergence was assessed using Tracer version 1.4 (<http://evolve.zoo.ox.ac.uk/tracer>), and was accepted if effective sample size (ESS) was above 200 (Drummond *et al.* 2007). Posterior probabilities were examined to estimate support for individual clades for further analyses. All tree topologies were highly similar, and hence, mean values for the five independent runs will be presented below.

The Bayesian tree, which provides an estimate of mean branch length, was used to infer the geographical origin of Hawaiian BBTV. Ancestral state reconstructions were performed using the likelihood algorithm implemented in Mesquite version 1.11 (Maddison & Maddison 2006) under the Markov *k*-state 1 parameter with parameters estimated based on the molecular data (Lewis 2001). This approach utilizes tree topology and branch lengths to produce probability values for ancestral state reconstructions throughout ancestral nodes in the phylogeny (Schluter *et al.* 1997). Reconstruction of ancestral area used a six-state, unordered matrix delineated by world regions: (i) East Asia, (ii) South Asia, (iii) Southeast Asia, (iv) Middle East, (v) Oceania/South Pacific, and (vi) Hawaii for the summary topologies

resulting from individual Bayesian runs. There was no statistically significant difference between topologies.

Molecular clock and dating of BBTV invasion

Several methods have been devised to estimate divergence times on phylogenetic trees (Kishino *et al.* 2001; Sanderson 2002; Yang & Yoder 2003). BEAST version 1.4.6 was used to infer a temporal framework of intra-island virus spread. This program uses a relaxed uncorrelated clock model and allows for probabilistic priors on node age for date calibrations (Drummond & Rambaut 2006). Because invasion may have happened prior to detection, probabilistic priors are important in order to model uncertainty in invasion times. To study the spread of BBTV within Hawaii, we concatenated the five DNA fragments sequenced for a total of 2560 bp per taxon. No evidence for recombination among fragments was found using the likelihood-based test in HyPhy (Kosakovsky Pond *et al.* 2005; data not shown). ModelTest version 3.8 (Posada & Crandall 1998) was used to determine an appropriate model under the AIC. The HKY + Γ + I model was selected for the concatenated data set.

We inferred Hawaiian BBTV colonization and divergence dates under a relaxed uncorrelated lognormal clock model simultaneously with the phylogenetic reconstruction of the Hawaiian invasion. Terminal nodes (sampled sequences) were set at time zero to reflect the fact that samples were taken contemporaneously, and phylogenetic outputs and absolute dates were interpreted from the 2005 sampling date. Three time-point calibrations for island infections were used based on available agricultural data: initial infection of the Hawaiian archipelago = 1989; first infection discovered on Hawaii = 1995; first infection discovered on Maui = 2002. Exponentially distributed priors on ancestral node ages were used and included a hard lower bound and a soft upper bound (Ho 2007). The hard lower bound constrains the date analysis such that an infection cannot be dated to occur after the first observed infection on an island. The soft upper bound indicates that an observation of an infection might lag behind the first observation of BBTV on a given island. We did not place a prior estimation of the Kauai invasion because the samples represented in our data set may not represent a single infection, and the prior may artificially constrain taxa sharing a clade with a paraphyletic Kauai infection. We examined both the standard deviation of the uncorrelated lognormal relaxed clock (ucl.d.stdev) and coefficient of variation parameters to examine the degree of rate variation within our tree (Drummond *et al.* 2007).

A total of five independent runs were sampled every 1000 generations for a total of 2×10^7 generations. The burn-in was set to exclude the first 10% of sampled generations. Posterior convergence was evaluated using

Tracer version 1.4 (<http://evolve.zoo.ox.ac.uk/tracer>), and accepted if stationarity was achieved with an ESS above 200. Each of the five output trees was examined by eye to assess consistency in topology, and individual island clades. Each individual run was then concatenated into a single tree file containing the posterior sampling from all five runs (minus burn-in) to provide a single tree estimate. The mean of the 95% high posterior density (HPD) was taken as an estimate of the absolute date for each node. We estimated the evolutionary rate of Hawaiian BBTV in base-pair substitutions per year with phylogenetic and population analyses ran under the same parameters as stated above.

Sources of spread within Hawaii

To infer the intra-archipelago source pools of BBTV spread across the Hawaiian Islands, we mapped the observed geographical data for Hawaiian BBTV distribution on to the single dated phylogeny (with branch lengths estimated as a function of time). We conducted ancestral geographical state reconstructions under the same likelihood method as above using Mesquite version 1.11. Current geographical field data of island infections was coded as a six-state, unordered state matrix (Kauai, Oahu Molokai, Maui, West Hawaii, and East Hawaii). These results are identical to those presented in Fig. 1.

BBTV in Hawaii: timing of invasion and population structure

The historical infection dynamics of BBTV spread in the Hawaii archipelago were estimated simultaneously with the dated phylogenetic reconstruction of the Hawaiian BBTV infection (above). Effective numbers of infections (N_e) in years before present (BP) were projected under a five-group ($m = 5$) Bayesian skyline plot estimation (Drummond *et al.* 2005). The skyline plot was estimated under the same likelihood model of base pair substitution (HKY) and initial priors as from above. Hawaiian BBTV is a unique data set for historical population estimation because (i) it is spread over a geographically diverse region (i.e. an island archipelago); (ii) it represents a single infection or population of BBTV; and (iii) the data generated from five viral components drawn from a worldwide sample produces a well-supported and resolved phylogeny for regional infections (several well-supported clades; Hawaii, Maui *et al.*). Although a prior substitution rate for the Hawaiian infection is not known, strong evidence for the initial Hawaiian BBTV infection in 1989 (Conant 1992) can be used to estimate mean substitution rate. The substitution rate was not fixed, and was allowed to sample the posterior distribution reflecting absent prior knowledge of a substitution rate for BBTV.

Sequential transfer of BBTV under controlled conditions

We conducted a sequential transfer experiment to estimate the rate of mutations in BBTV under controlled conditions in relation to the number of vector transmission events. For this purpose, we allowed single aphid vectors (*P. nigronervosa*) to acquire BBTV from symptomatic plants for 24 h and transferred these insects to healthy tissue-cultured banana plantlets for a 24-h inoculation access period. This process was initiated with one individual symptomatic plant, which generated 14 parallel BBTV-infection lineages that were sequentially transferred, as described above, for 10 generations (i.e. 10 new plants were sequentially infected by individual aphids, one aphid per plant). Because under greenhouse conditions plants became symptomatic with BBTV within approximately 1 month, transfers were made every 4–5 weeks after infection of a new plant generation. The experiment took 1 year to complete. Aphids were reared on healthy banana plants and randomly collected from our colony for transmission tests. Details on general insect and plant maintenance, experimental aphid infection protocols, and greenhouse conditions have been described

elsewhere (Robson *et al.* 2007; Anhalt & Almeida 2008). We kept tissue and extracted DNA samples from all lineages and passages. After 10 generations, we sequenced the same five DNA fragments used for the phylogeographical analyses for the original infected plant and all the descendants (last generation) of the 14 lineages.

Results*Origin of Hawaiian BBTV invasion: a global approach*

To determine the origin and number of BBTV invasions into Hawaii, we combined DNA-1 component sequences from material collected throughout the state with available data (Table S1). Visual inspection of five independently derived world phylogenies indicated nearly identical topologies, thus the first phylogeny was arbitrarily selected for ancestral geographical distribution analysis. Summary statistics for averaged MCMC searches are presented in Table 1, under 'World phylogeny', with posterior probability and $-\ln L$ scores presented as mean values from the five independent runs. Figure 1 depicts

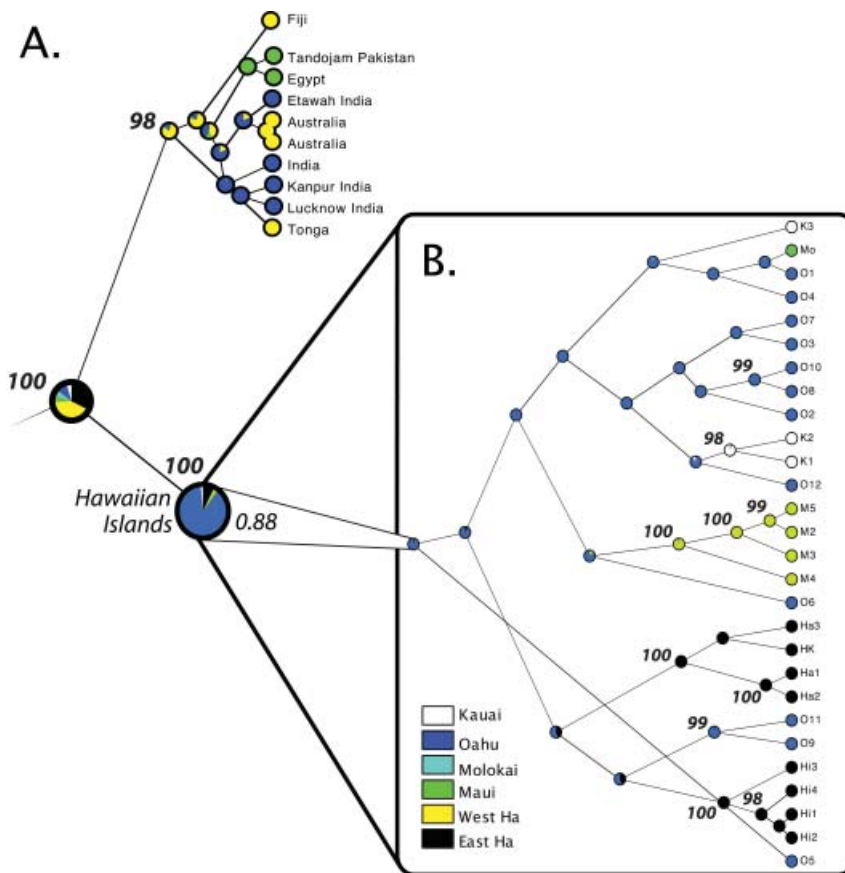


Fig. 1 (A) Bayesian phylogeny showing Hawaiian clade (PP = 1.0) and Middle Eastern – South Pacific clade (PP = 0.99) taken from a phylogeny of 64 worldwide *Banana bunchy top virus* rooted with *Cardamom clump virus* from aligned DNA-1 sequences with base-pair substitution rate fixed (likelihood = -8688.01 ; posterior = -8591.85). Phylogeny with likelihood ancestral state reconstructions of an unordered six-state character state matrix under MK-1 model ($-\ln L = 40.15$). Posterior probabilities are above each node; nodes without values are supported less than 90%. (B) Ancestral state reconstruction showing Oahu as the area of first infection of BBTV (probability = 0.88, to right of node) in Hawaii derived from one of the five Bayesian phylogenies (probability on other four trees were 0.92, 0.90, 0.69, and 0.87). Collapsing poorly supported internal nodes on all phylogenies increase reconstruction of Oahu as ancestral to ~ 0.99 . Pie charts at internal nodes show optimization for character states (proportional likelihood values) with unequal branch lengths derived from phylogenetic analysis. Probability values for Hawaiian ancestral state reconstructions infection can be found in Table 2.

Table 1 Summary statistics for combined Markov chain Monte Carlo runs taken from Tracer version 1.3 for Bayesian phylogenetic estimation of world *Banana bunchy top virus* infection and Hawaiian intra-island dated phylogeny analysis with simultaneous Bayesian skyline plot estimation of population structure over time

	Mean	95% HPD lower*	95% HPD upper†	ESS‡
World phylogeny				
Posterior	-8618.265	-8635.063	-8602.222	3445.981
Tree likelihood	-8371.317	-8727.833	-8698.982	5110.345
Hawaii infection				
Posterior	-4220.753	-4237.2271	-4204.262	2207.308
Tree likelihood	-4070.559	-4081.569	-4059.745	4985.939
Mean rate (bp/year)	1.425E-4	8.818E-5	2.07E-4	1374.235
MRCAS§ (root height)	17.324	16	19.961	8676.606
Skyline	-123.004	-132.032	-113.31	1670.71
ucl.mean¶	1.579E-4	8.975E-5	2.4E-4	1576.053
COV **	0.411	9.135E-5	0.885	3409.075

*, † high posterior density of the lower and upper (respectively) distribution of values obtained from the MCMC for each parameter.
‡ effective sampling size.

§ root height of tree to the most recent common ancestor of the Hawaiian Infection.

¶ the mean of the branch rates under the uncorrelated lognormal relaxed clock model.

** the coefficient of variation.

Table 2 *Banana bunchy top virus* clades with the mean posterior probability support, mean inferred age in years before the present (BP) with 95% high posterior density (HPD), and mean probability values for ancestral state reconstructions of infection source. Values obtained from combined Bayesian phylogenetic analysis. Ancestral state probabilities are taken from Fig. 1B

Clade	Mean posterior probability	Mean age (BP)	95% HPD	Ancestral state probability
(M4, M5, M2, M3)	0.9996	4.03	(3.0–5.804)	Oahu, 0.984*
(Ha2, Ha1, HK, Ha3)	0.9999	10.608	(10.0–11.812)	Oahu, 0.602*; W. Hawaii, 0.321
(H4, H1, H2, H4)	1.0	2.246	(1.0–4.06)	Oahu, 0.970*
(K1, K2)	0.892	4.321	[0.846–8.945]	Oahu, 0.934*
(K3)	0.28	8.1624	—	Oahu, 0.931*
(Mo)	0.7584	1.908	(0.0001–6.423)	Oahu, 0.904*

* indicate statistically significant values for ancestral geographical state as determined using Mesquite's likelihood ancestral state reconstruction model.

nodes of interest for the inferred tree, with ancestral state reconstructions for the first run for inferences of geographical origin of BBTV invasion centred on the Hawaiian infection (complete ancestral state reconstructions for the world data set are not shown). BBTV isolates from Hawaii formed a monophyletic group (posterior probability (PP) = 1.0), indicating a single invasion of the archipelago. The Hawaiian strain was found to be sister to the Middle Eastern–Oceania/South Pacific infections (PP = 0.99). Within this clade, Tonga and Fiji were basal (PP = 0.98 and PP = 0.49, respectively) to the Middle Eastern (including India) isolates. While the worldwide distribution of the sister group says little about the origin of the Hawaiian BBTV, likelihood ancestral reconstructions (Fig. 1; $-\log L = 40.15$) indicate a strong link between the Oceania/South Pacific infections, probability = 0.72 (Oceania/South Pacific = 0.41 and Hawaii = 0.31).

Intra-island spread of Hawaiian BBTV

Within the Hawaiian clade, support at basal nodes was low (PP < 0.50). However, there was reasonable support for spread of BBTV to individual islands within the chain (Table 2). Due to low branch support for deep internal nodes of the Hawaiian topology, likelihood ancestral state reconstructions were performed on both the originally resolved topologies ($-\log L = 27.11$) taken from BEAST, and on the appropriately collapsed phylogenies ($-\log L = 26.61$). Both yielded statistically significant probabilities of an Oahu origin (88% and 99%, respectively). Inspection of nodes leading to each independent infection of other islands in the archipelago within the collapsed phylogeny yielded Oahu as a statistically significant geographical source pool for infection/colonization. For clarity, Fig. 1 depicts the initial uncollapsed Hawaiian BBTV ancestral

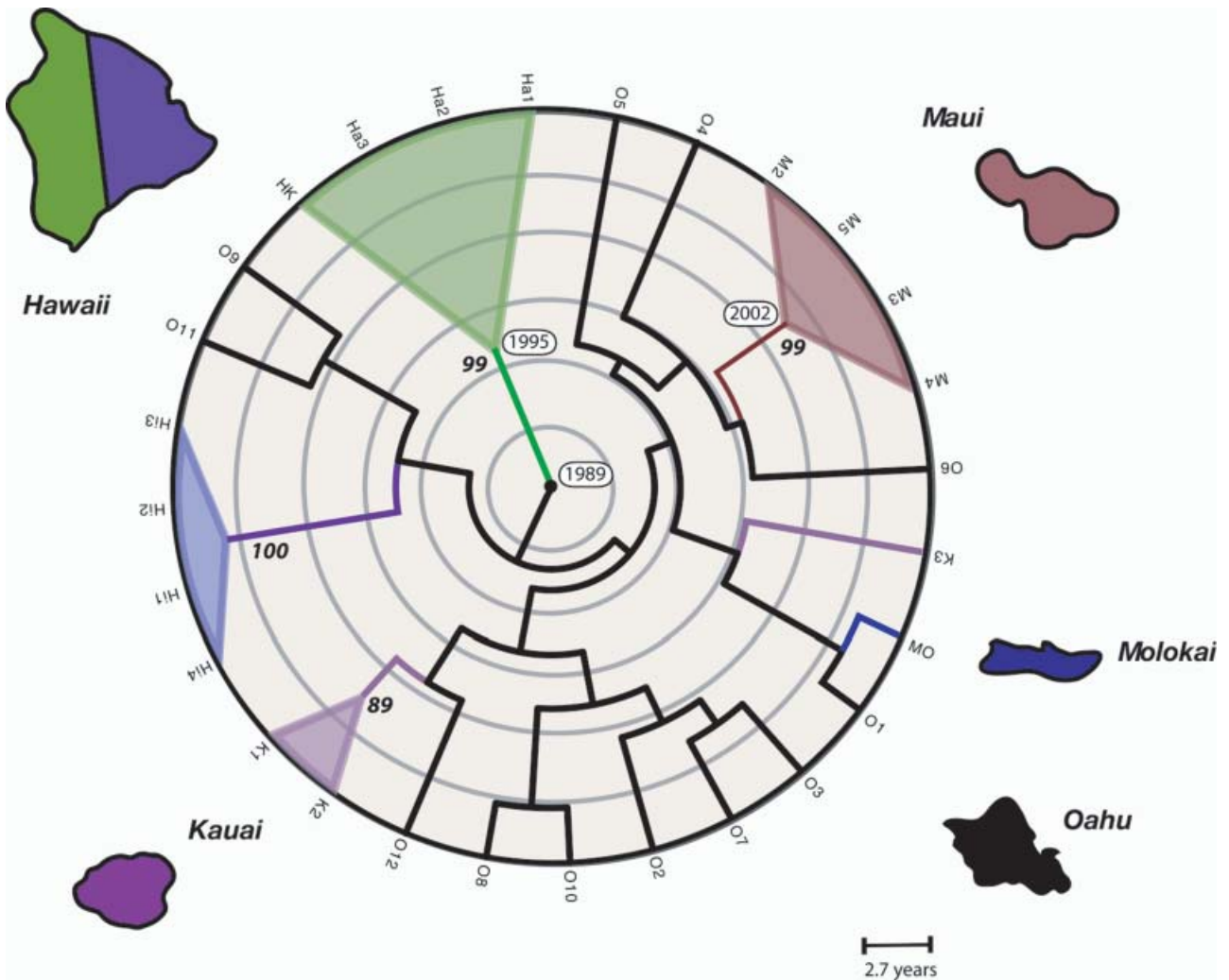


Fig. 2 Bayesian calibrated, time-orientated circular phylogeny of 28 concatenated molecular sequences (DNA-1, 3, 4, 5, 6) for Hawaiian archipelago *Banana bunchy top virus* infection (mean likelihood = -4070.559 , mean posterior = -4220.753). Dated phylogeny estimated under an uncorrelated lognormal relaxed molecular clock model in BEAST. Rings indicated time intervals of approximately 3 years between a mean MRCA of 17.324 years before sequence sampling date, 2005. The three agricultural calibration points, shown as the first recorded date of infection, are indicated in the phylogeny. Posterior probability values over 85% are shown.

state phylogeny and likelihood reconstruction of ancestral states. Ancestral state reconstructions were used to determine the ancestral nodes for that corresponded to a successful island infection for dating analyses below.

Dating and population structure of Hawaiian BBTV

Because our analysis did not identify the presence of recombinants in the data set, concatenated sequences of fragments of five BBTV components were used. Posterior estimation of the uclsd.stdev mean was 0.433, and the mean coefficient of variation was 0.411 (abutting against zero; data not shown). Therefore, we selected a relaxed uncorrelated lognormal model to infer divergence times in our

analyses (Drummond *et al.* 2007). The base-pair substitution rate (mean rate statistic; number of substitutions per time) for Hawaiian BBTV was estimated at approximately 1.425×10^{-4} substitutions per year (95% HPD = 8.818×10^{-5} – 2.07×10^{-4}).

Figure 2 illustrates the time-orientated, Bayesian phylogeny of the Hawaiian BBTV infection. Averaged summary statistics for the converged Bayesian MCMC algorithm searches of posterior tree space and skyline estimation of effective number of infections over time are in Table 1. An origination time of 17.324 years before 2005, the date of Hawaiian sequence collection, suggests that BBTV was introduced to Hawaii in about 1987, 2 years prior to its detection. A Bayesian Skyline plot (mean posterior skyline

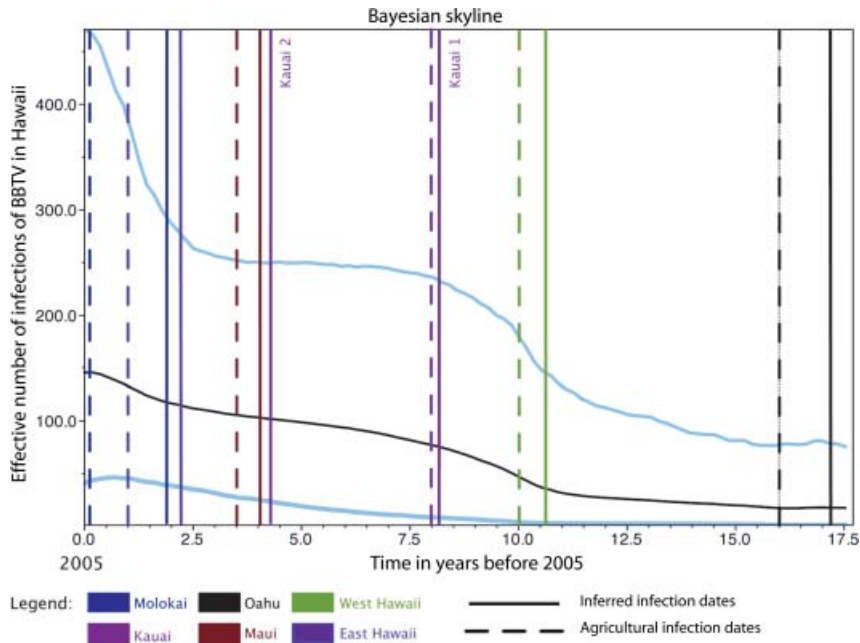


Fig. 3 Bayesian skyline plot estimate of effective infection size of *Banana bunchy top virus* infecting the Hawaiian Archipelago from 28 concatenated noncoding molecular data from the origin of infect (c. 17.324 BP) to 2005. The horizontal dark line indicates the mean estimate for population dynamics overtime in years before present, and pale blue lines indicate the 95% high posterior density. Vertical lines represent approximate times of viral infection on, respectively, labelled islands.

value = -123.004) illustrates the dynamics of the effective number of BBTV infections since 17.324 BP [the mean reconstructed origin of infection in Hawaii, 95% HPD = (16.0–19.961); Fig. 3]. We overlaid information provided by the Hawaii Department of Agriculture on the dates when BBTV was first detected, and our estimated dates of infection on the skyline plot. Pale blue lines indicate uncertainty in the number of infections as the 95% HPD for population size at any given time along the x -axis. Effective number of infections increases little between the origin of BBTV in Hawaii (~ 17.3 BP) and the colonization of West Hawaii. At approximately 10.5 BP (1994), which corresponds to a West Hawaii infection, N_e exhibits a sharp increase in slope. The sharp population expansion stabilizes at approximately 8 BP (1997), yielding to more slowly increasing population growth, during which other islands are colonized. After these colonization events, another fast increase in N_e occurred (last 2 years), probably due to epidemics in recently colonized islands, which had large populations of susceptible hosts. Results obtained with the molecular data did not completely match the Hawaii Department of Agriculture's information regarding the absolute dates of BBTV introductions to each of the islands. There were two patterns present in the results: (i) an inferred infection predating historical documentation (although 95% HPD covers agricultural detection), and (ii) a lag of the inferred date behind the documented dates for invasion of some islands. An inferred invasion date earlier than historical documentation is expected since the colonizing lineage itself may show an older age, and the presence of the disease may have gone undetected for sometime on indi-

vidual islands. The time lag shown by our inferred dates may reflect the case that for Kauai eradication efforts were conducted until the re-emergence of BBTV occurred. In the case of West Hawaii, eradication efforts may have dramatically reduced disease incidence; however, the disease seems to have persisted since its origin. Kauai is more difficult to interpret due to two possible infections of the island: the initial invasion occurring prior to the agricultural documentation, and the other more than 3 years later. Support for the reconstructed older lineage is weak, and interpretation of a single surviving infection, or an incomplete eradication effort with a subsequent independent infection is difficult to discern between. Given the reasonably supported ($pp = 0.892$) nested placement of the younger Kauai infection, it appears that a multiple infection hypothesis is feasible.

Sequential transfer of BBTV under controlled conditions

To determine an experimental rate of substitution for the BBTV genome, we conducted a sequential transfer experiment using aphids as vectors of the virus. After 10 generations of individual aphid transfer of BBTV from banana to banana, we detected one nucleotide substitution in one of the 14 lineages maintained (noncoding region of DNA-1; data not shown). Assuming this result to be the experimental substitution rate for BBTV, we estimated the rate of substitution under greenhouse conditions to be 3.91×10^{-4} bp/year (1 year to complete 10 passages). Although the single mutation was double-checked, we acknowledge that this experimentally derived estimate

of substitution rate has a large confidence interval. The substitution rate estimated by the Bayesian analysis was 1.425×10^{-4} bp/year (Table 1). Both of these substitution rates fall within the expected range for ssDNA viruses (Duffy *et al.* 2008). Based on these results (and assumptions), under field conditions in 1 year we would have 0.364 of the total number of passages compared to the greenhouse study, or 3.64 infection events (passages) where one new plant was infected by an aphid. This value can also be interpreted as an estimated basic reproductive rate for BBTv in Hawaii, i.e. $R_0 = 3.64$ (R_0 as the net reproductive rate for the pathogen in 1 year).

Discussion

We used phylogenetic inference to reconstruct BBTv's invasion of Hawaii. BBTv was introduced once into Oahu and from that island spread throughout the archipelago (Fig. 1). The lack of biogeographical pattern on pathogen dispersal (e.g. stepping stone model), and the rapidity of BBTv spread supports the hypothesis that transportation of contaminated plant material from Oahu to other islands was the driving factor in this epidemic. The evidence for multiple introductions in Kauai and Hawaii (Big Island) illustrates the difficulty in eradicating invasive species from new areas if there is a potential for reintroductions. In our case, approaches used were able to determine the presence of multiple introductions highlighting the potential of such tools for disease management. Several studies have used a similar approach in other invasion systems (Roderick 1996; Tsutsui *et al.* 2001; Nardi *et al.* 2005), most typically as a means to (i) identify the site of origin for a pest species to find natural enemies for biological control efforts or (ii) compare and contrast population structure between native and invading populations.

The potential to reliably reconstruct and time the historical spread of pathogens using molecular tools can provide valuable insights on disease ecology and pathogen evolution (Biek *et al.* 2006, 2007). However, can we use similar approaches to gain information on the temporal or spatial patterns of host-to-host pathogen spread? To address this question, we simulated a series of host infections under greenhouse conditions using aphid vectors. A similar approach, but with *in vitro*-based data, has been previously used for *Yersinia pestis* (Girard *et al.* 2004). Our goal was to determine the rate of nucleotide substitution of BBTv genomic DNA mirroring its natural mode of dispersal. We obtained an experimental substitution rate (based on a single mutation) only three times higher than that estimated by the phylogenetic analyses, which can be hypothesized to be affected by the shorter disease incubation period and potentially larger number of plant-to-plant transfers within 1 year in a greenhouse when compared to field conditions. After correcting for that difference, we estimated the growth

rate of new infections caused by one infected host would be $R_0 = 3.64$. This rate is very similar to BBTv-spread infection rate based on epidemiological data. Allen (1987) estimated the rate of new BBTv infections per infected plant per day to be $r = 0.014$, that is an $R_0 = 5.11$ /year. Although an ideal study to estimate rates of mutation should be conducted in the field to better mimic environmental conditions, it is notable that our experimental rate was similar to that estimated by the molecular analyses and an epidemiological variable (i.e. number of new infections per plant). It must be emphasized that the rate obtained here was based on a single mutation and consequently has a large confidence interval. Much larger experimental studies would be required to obtain a robust estimate of the substitution rate for BBTv. Regardless of pathogen-specific ecological differences and other limitations, our study provides support for the use of molecular clocks in the reconstruction of epidemics and the use of such data in conjunction with experimental substitution rates to determine variables of epidemiological importance. Despite concerns on the reliability of molecular clocks (Bromham & Penny 2003), they have been successfully used to reconstruct pathogen epidemics (Biek *et al.* 2007) and determine their origin (Stukenbrock *et al.* 2007).

The results we obtained were reasonably consistent with dates obtained in surveys conducted by the Hawaii Department of Agriculture. In our work, the rate of nucleotide substitution, the existence of eradication efforts and second colonization events may have impacted our conclusions. However, results indicate that most invasions occurred 1–2 years prior to their detection. In that case, with an $R_0 = 3.64$, the introduction of one BBTv-infected banana would result in ~ 4 –13 infected hosts at a new island prior to detection (no. of infected hosts_{*t*} = $R_0^t \times$ no. of infected hosts_{*t=0*}, with *t* = time; as discussed in Madden *et al.* 2007). As expected, this plant pathogen was not detected immediately after the first introduction, but its slow rate of spread allowed for detection when the epidemic were likely starting on each island.

Zuckerkandl & Pauling (1965) first proposed the idea that evolutionary events can be dated by assuming a constant rate of change over time. Since then, evolutionary biologists have inferred divergence times using either fossils or biogeography as calibration points. Computational advances have allowed for more realistic date estimation by relaxing the requirement of a constant rate across the focal group (Sanderson 1997, 2002; Thorne & Kishino 2002; Ho *et al.* 2005; Drummond *et al.* 2006; Ho & Larson 2006; Yang & Rannala 2006). A major criticism, however, has been the lack of incorporating error in either calibration or divergence time estimates (Graur & Martin 2004; Pulquerio & Nichols 2006; see also Hedges & Kumar 2004). Incorporation of error in fossil or biogeographical calibration points can be addressed with Bayesian approaches (Thorne

& Kishino 2002; Yang & Rannala 2006) that use soft calibration points. BBTV in Hawaii provides a good model system for applying the molecular clock to an agricultural system; it combines good calibration points with a wealth of sequence data in an analytical framework that allows for hypothesis testing and the incorporation of error.

Our results show that the main driving factor behind the Hawaiian BBTV epidemic was most likely the transportation of contaminated plant material originated from Oahu to other islands in the archipelago. The total number of established interisland invasions observed in less than 20 years highlights the difficulty in controlling the movement of infectious hosts if those are cultivated plants that remain asymptomatic for many weeks after infection. Although molecular inference provided means to interpret this epidemic, those results had to be interpreted carefully as banana is a cultivated crop subject to disease control management practices, and quarantine and eradication efforts that may have introduced noise into phylogenetic signals.

Acknowledgements

We thank members of the Hawaii Department of Agriculture, Maui Invasive Species Committee and University of Hawaii at Manoa for providing some of the samples used in this study. We thank Matt Daugherty, Mark Wright, Cerruti Hooks and members of our laboratories for helpful discussions. Work was funded by grants from USDA-CSREES (WSARE and TSTAR), USDA-ARS (PBARC) and UCB-AES.

References

- Allen RN (1978) Spread of banana bunchy top disease in established banana plantations. *Australian Journal of Agricultural Research*, **29**, 1223–1233.
- Allen RN (1987) Further studies on epidemiological factors influencing control of banana bunchy top disease and evaluation of control measures by computer simulation. *Australian Journal of Agricultural Research*, **38**, 373–382.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, **19**, 535–544.
- Anhalt MD, Almeida RPP (2008) Effect of temperature, vector life stage and plant access period on transmission of *Banana bunchy top virus* to banana. *Phytopathology*, **98**, 743–748.
- Biek R, Drummond AJ, Poss M (2006) A virus reveals population structure and recent demographic history of its carnivore host. *Science*, **311**, 538–541.
- Biek R, Henderson JC, Waller LA, Rupprecht CE, Leal LA (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences, USA*, **104**, 7993–7998.
- Bromham L, Penny D (2003) The modern molecular clock. *Nature Reviews Genetics*, **4**, 216–224.
- Conant P (1992) Banana bunchy top disease, a new threat to banana cultivation in Hawaii. *Proceedings, Hawaiian Entomological Society*, **31**, 91–95.
- Cunningham AA, Daszak P, Rodríguez JP (2003) Pathogen pollution: defining a parasitological threat to biodiversity conservation. *Journal of Parasitology*, **89**, 78–83.
- Dale JL (1987) Banana bunchy top: an economically important tropical plant virus disease. *Advances in Virus Research*, **33**, 301–325.
- Drummond AJ, Rambaut A (2006) *BEAST Version 1.4*. Available from <http://beast.bio.ed.ac.uk/>.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**, 1307–1320.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Public Library of Science, Biology*, **4**, 699–710.
- Drummond AJ, Ho SYW, Rawlence N, Rambaut A (2007) A rough guide to BEAST 1.4. Accessed 26 April 2007 from <http://beast.bio.ed.ac.uk/Main_Page>.
- Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, **9**, 267–276.
- Girard JM, Wagner DM, Vogler AJ *et al.* (2004) Differential plague transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional, and global scales. *Proceedings of the National Academy of Sciences, USA*, **101**, 8408–8413.
- Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, **20**, 80–86.
- Hafner GJ, Harding RM, Dale JL (1995) Movement and transmission of *Banana bunchy top virus* DNA component one in bananas. *Journal of General Virology*, **76**, 2279–2285.
- Hedges SB, Kumar S (2004) Precision of molecular times estimates. *Trends in Genetics*, **20**, 242–247.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.
- Ho SYW (2007) Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology*, **38**, 404–414.
- Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics*, **22**, 79–83.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic over-estimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.
- Hu JS, Wang M, Sether D, Xie W, Leonhardt KW (1996) Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (*Pentalonia nigronervosa*). *Annals of Applied Biology*, **128**, 55–64.
- Janies D, Hill AW, Guralnick R, Habib F, Waltari E, Wheeler WC (2007) Genomic analysis and geographic visualization of the spread of avian influenza (H5N1). *Systematic Biology*, **56**, 321–329.
- Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**, 676–679.

- Lanave C, Preparatu G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rate. *Journal of Molecular Evolution*, **20**, 86–93.
- Lanciotti RS, Roehrig JT, Deubel V *et al.* (1999) Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science*, **286**, 2333–2337.
- Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925.
- Madden LV, Hughes G, van den Bosch F (2007) *The Study of Plant Disease Epidemics*. APS Press, St Paul, Minnesota.
- Maddison D, Maddison W (2002) *MacClade: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Massachusetts.
- Maddison WP, Maddison DR (2006) *Mesquite: Modular System for Evolutionary Analysis*, Version 1.11. Available at <http://mesquite-project.org>.
- Magee CJP (1940) Transmission studies on the banana bunchy-top virus. *Journal of the Australian Institute of Agricultural Sciences*, **6**, 109–110.
- Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology*, **14**, 2729–2738.
- Posada D (2006) ModelTest 8 Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Research*, **34**, W700–W703.
- Posada D, Crandall KA (1998) ModelTest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Pulquerio MJF, Nichols RA (2006) Dates from the molecular clock: how wrong can we be? *Trends in Ecology & Evolution*, **22**, 180–184.
- Robson JD, Wright MG, Almeida RPP (2007) Biology of *Pentalonia nigronervosa* (Hemiptera, Aphididae) on banana using different rearing methods. *Environmental Entomology*, **36**, 46–52.
- Roderick GK (1996) Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual Reviews of Entomology*, **41**, 263–290.
- Sanderson MJ (1997) A non-parametric approach to estimating divergence times in the absence of rate constancy. *Molecular Ecology and Evolution*, **14**, 1218–1231.
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Molecular Ecology and Evolution*, **19**, 101–109.
- Schluter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699–1711.
- Smith MC, Holt J, Kenyon L, Foot C (1998) Quantitative epidemiology of banana bunchy top virus disease and its control. *Plant Pathology*, **47**, 177–187.
- Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, **24**, 398–411.
- Swofford DL (2002) *PAUP* Phylogenetic analysis using parsimony (*and other methods)*, Version 4. Sinauer Associates, Sunderland, Massachusetts. <http://paup.csit.fsu.edu/>.
- Thorne JL, Kishino H (2002) Divergence time estimation and rate evolution with multilocus data sets. *Systematic Biology*, **51**, 689–702.
- Tsutsui ND, Suarez AV, Holoway DA, Case TJ (2001) Relationships among native and introduced populations of the Argentine ant *Linepithema humile* and the source of introduced populations. *Molecular Ecology*, **10**, 2151–2161.
- Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, **11**, 1842–1847.
- Yang Z (1996) Maximum-likelihood models for combined analyses of multiple sequence data. *Journal of Molecular Evolution*, **42**, 587–596.
- Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212–226.
- Yang Z, Yoder AD (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology*, **52**, 705–716.
- Zimmerman EC (1948) *Insects of Hawaii: A Manual of the Insects of the Hawaiian Islands, Including an Enumeration of the Species and Notes on Their Origin, Distribution, Hosts, Parasites, etc.* University of Hawaii Press, Honolulu, Hawaii.
- Zuckermandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: *Evolving Genes and Proteins* (Bryson V, Vogel HJ, eds), pp. 97–166. Academic Press, New York.

Rodrigo Almeida's research focuses on the ecology of insect-borne plant diseases, insect transmission of plant pathogens and insect-microbe interactions. Gordon Bennett is interested in the taxonomy and systematics of native Hawaiian leafhoppers and their coevolutionary interactions with their obligate endosymbionts. Mandy Anhalt is interested on insect transmission of plant pathogens. Chi-Wei Tsai is interested in the evolutionary relationships between virus and insect vector. Patrick O'Grady studies the ecology and evolutionary biology of Hawaiian *Drosophilidae*. His research has recently focused on taxonomy, molecular systematics, behavioural ecology and genomics.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Table of 64 included *Banana bunchy top virus* sequences. Taxon names correspond to those found in Fig. 1 (unless not shown in ancestral state reconstruction). Specific country of origin is listed included under locality, and was used for the ancestral state reconstructions of geographical distribution. When available (i.e. Hawaiian clade, gaps) location data was included. GenBank Accession numbers for all sequences are provided

Table S2 Oligonucleotide primer sequences used for amplification and sequencing of *Banana bunchy top virus* genome fragments, and annealing temperature used for amplification of each primer set

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