

# Recombination and Migration of *Cryphonectria hypovirus 1* as Inferred From Gene Genealogies and the Coalescent

Ignazio Carbone,\* Yir-Chung Liu,<sup>†1</sup> Bradley I. Hillman<sup>‡</sup> and Michael G. Milgroom<sup>†,2</sup>

\*Center for Integrated Fungal Research, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695,

<sup>†</sup>Department of Plant Pathology, Cornell University, Ithaca, New York 14853 and <sup>‡</sup>Department of Plant Pathology, Rutgers University, New Brunswick, New Jersey 08901

Manuscript received September 5, 2003

Accepted for publication January 6, 2004

## ABSTRACT

Genealogy-based methods were used to estimate migration of the fungal virus *Cryphonectria hypovirus 1* between vegetative compatibility types of the host fungus, *Cryphonectria parasitica*, as a means of estimating horizontal transmission within two host populations. Vegetative incompatibility is a self/non-self recognition system that inhibits virus transmission under laboratory conditions but its effect on transmission in nature has not been clearly demonstrated. Recombination within and among different loci in the virus genome restricted the genealogical analyses to haplotypes with common mutation and recombinational histories. The existence of recombination necessitated that we also use genealogical approaches that can take advantage of both the mutation and recombinational histories of the sample. Virus migration between populations was significantly restricted. In contrast, estimates of migration between vegetative compatibility types were relatively high within populations despite previous evidence that transmission in the laboratory was restricted. The discordance between laboratory estimates and migration estimates from natural populations highlights the challenges in estimating pathogen transmission rates. Genealogical analyses inferred migration patterns throughout the entire coalescent history of one viral region in natural populations and not just recent patterns of migration or laboratory transmission. This application of genealogical analyses provides markedly stronger inferences on overall transmission rates than laboratory estimates do.

GENEALOGY-BASED methods have been shown to enhance precision over  $F_{ST}$ -based estimators for estimating migration because they consider the temporal relationships of individuals in a population sample in addition to their frequency and spatial distributions (HUDSON *et al.* 1992b; TEMPLETON 1998). This temporal component is important for detecting and separating migration from other population diversifying forces, such as recombination (HUDSON *et al.* 1992b) and isolation (NIELSEN and WAKELEY 2001). Coalescent-based simulations of genealogical relationships further enhance the precision of migration estimates by allowing for unequal population sizes and asymmetries in migration rates in the reconstruction of ancestral migration patterns (BEERLI and FELSENSTEIN 1999, 2001; BAHLO and GRIFFITHS 2000; NIELSEN and WAKELEY 2001). Because most coalescent-based approaches assume neutrality and no recombination they are most powerful when used in conjunction with other genealogical meth-

ods that can (i) test the neutral mutation hypothesis and (ii) identify potential recombination events in the history of the sample (CARBONE and KOHN 2001a). Although these analytical methods have proven to be powerful for making inferences on migration and effective population sizes on simulated data sets (BEERLI and FELSENSTEIN 1999; NIELSEN and WAKELEY 2001; VITALIS and COUVET 2001; WAKELEY and ALIACAR 2001), relatively few empirical examples have been tested (CONGDON *et al.* 2000; BEERLI and FELSENSTEIN 2001; CARBONE and KOHN 2001a; HEATH *et al.* 2002; KINNISON *et al.* 2002).

One potential application of migration analyses is to track the movement of specific genotypes within and between populations. A specific case in point is the estimation of transmission of pathogens between host individuals (horizontal transmission). Horizontal transmission is a pivotal parameter affecting the evolution of virulence (BULL 1994; LEVIN 1996) and the potential for pathogens to invade host populations (ANDERSON and MAY 1986). Information is available about horizontal transmission of RNA viruses in some host systems, but relatively little detail is known about horizontal virus transmission in filamentous fungi. Horizontal transmission of viruses in filamentous fungi occurs when the cells of different individuals fuse, allowing viruses to move from the cytoplasm of one individual to the other. However, horizontal transmission is restricted by vegeta-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY125727–AY125813.

<sup>1</sup>Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544.

<sup>2</sup>Corresponding author: Department of Plant Pathology, 334 Plant Science Bldg., Cornell University, Ithaca, NY 14853.  
E-mail: mgm5@cornell.edu

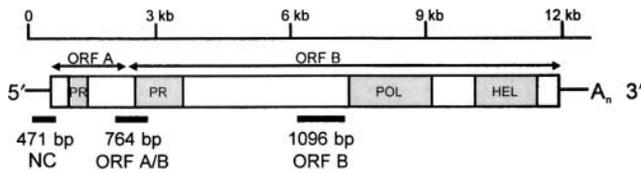


FIGURE 1.—Genome organization of CHV-1 (SHAPIRA *et al.* 1991) showing open reading frames (ORFs A and B) with proteases (PR), RNA polymerase (POL), and helicase (HEL) genes. Locations, sizes, and abbreviations for sequenced regions are also shown (see text for details).

tive or heterokaryon incompatibility (reviewed in CORTESI *et al.* 2001), a genetically controlled self/non-self recognition system that results in programmed cell death between incompatible genotypes (BIELLA *et al.* 2002; GLASS and KANEKO 2003). Cell death associated with vegetative incompatibility often, but not always, reduces horizontal transmission of viruses, plasmids, or other cytoplasmic elements (*e.g.*, CATEN 1972; HARTL *et al.* 1975; GRIFFITHS *et al.* 1990; CORTESI *et al.* 2001).

In the chestnut blight fungus, *Cryphonectria parasitica*, vegetative incompatibility has been hypothesized to be a major barrier to horizontal transmission and, therefore, a significant constraint on biological control of chestnut blight with viruses (ANAGNOSTAKIS *et al.* 1986; MILGROOM 1995; LIU *et al.* 2000). Vegetative incompatibility is controlled by alleles at six or more vegetative incompatibility (*vic*) loci in *C. parasitica*; multilocus *vic* genotypes are referred to as vegetative compatibility (*vc*) types (CORTESI and MILGROOM 1998). Detailed laboratory studies have shown that each *vic* allele has a different quantitative effect reducing the probability of virus transmission (CORTESI *et al.* 2001). Although it is possible to estimate the average transmission between *vic* genotypes in the laboratory, these estimates do not necessarily reflect actual transmission under natural conditions. In nature, transmission may be affected by the number of contacts between individuals, length of time

individuals are in contact, spatial heterogeneity of genotypes, etc. Therefore, estimating migration of viruses using genealogical methods offers an alternative means of estimating horizontal transmission between *vic* genotypes independent of laboratory estimates.

The primary objective in this research was to estimate migration of viruses between groups of vegetatively incompatible individuals (*i.e.*, *vic* genotypes) in two populations of the host fungus, *C. parasitica*. A gene genealogical approach to estimating migration requires that we test assumptions concerning recombination and population subdivision. Therefore, our subsidiary objectives were to test for recombination in the virus genome and estimate virus migration between populations.

## MATERIALS AND METHODS

*Cryphonectria hypovirus 1* (CHV-1) is a double-stranded RNA (dsRNA) virus in the family Hypoviridae that infects the chestnut blight fungus, *C. parasitica* (HILLMAN *et al.* 2000). Although classified as a dsRNA virus, the 12.7-kb RNA genome of CHV-1 (Figure 1) is more closely related to positive-sense RNA viruses and shares their essential replication properties (KOONIN *et al.* 1991; FAHIMA *et al.* 1993; HILLMAN *et al.* 2000). Interest in this virus derives from its potential as a biological control agent for chestnut blight (VAN ALFEN *et al.* 1975; NUSS 1992). CHV-1 has been found extensively in Europe (ALLEMANN *et al.* 1999); it has also been found in Japan and China (PEEVER *et al.* 1998; LIU *et al.* 2003) and, more recently, in Korea (J. LIM, B. CHA and M. G. MILGROOM, unpublished observations). CHV-1 was probably introduced into Europe with *C. parasitica*, which was introduced into North America and Europe from eastern Asia in the early 1900s (MILGROOM *et al.* 1996).

**Populations sampled:** We sampled CHV-1 from two populations of *C. parasitica*: Teano from Caserta Province in southern Italy and Bergamo from Bergamo Province in northern Italy. Populations of *C. parasitica* in Italy are dominated by a few *vic* genotypes (or *vc* types; CORTESI *et al.* 1996; MILGROOM and CORTESI 1999), making it possible to obtain multiple isolates of CHV-1 from the same *vc* types. In Teano ( $n = 194$ ), 96% of the fungal isolates were in *vc* types EU-10 (19%) and EU-12 (77%; LIU *et al.* 1996; MILGROOM and CORTESI 1999). In

TABLE 1  
Oligonucleotide primers used for amplifying parts of three regions of CHV-1

Region	Primer name	Primer sequence (5' → 3')	Nucleotide position <sup>a</sup>
5'-noncoding	747nc1	GTTGCTCGATATAGACCGCCATTA	32–55
	747nc2	CGTCGACGAAAGGATCCACTGTAG	554–578
ORF A/ORF B	747con8	AATCAACCACTTCCTG	2042–2057
	747con9	ATTGAAGTTCCGTTAG	2829–2844
	747con17	ATGTATAA(A+G)GAAGCCG	2364–2379
	747con4	AGGGGCAGTCCATGTTG	2422–2438
ORF B <sup>b</sup>	713con1	TACAATTGGTGGTGGCCG	6066–6083
	713con2	TAATCATCAAGGAAATTCCC	7269–7288

<sup>a</sup> Positions are relative to CHV-1/EP713 (GenBank accession no. M57938).

<sup>b</sup> Primers for ORF B were previously reported by LIU *et al.* (2003).

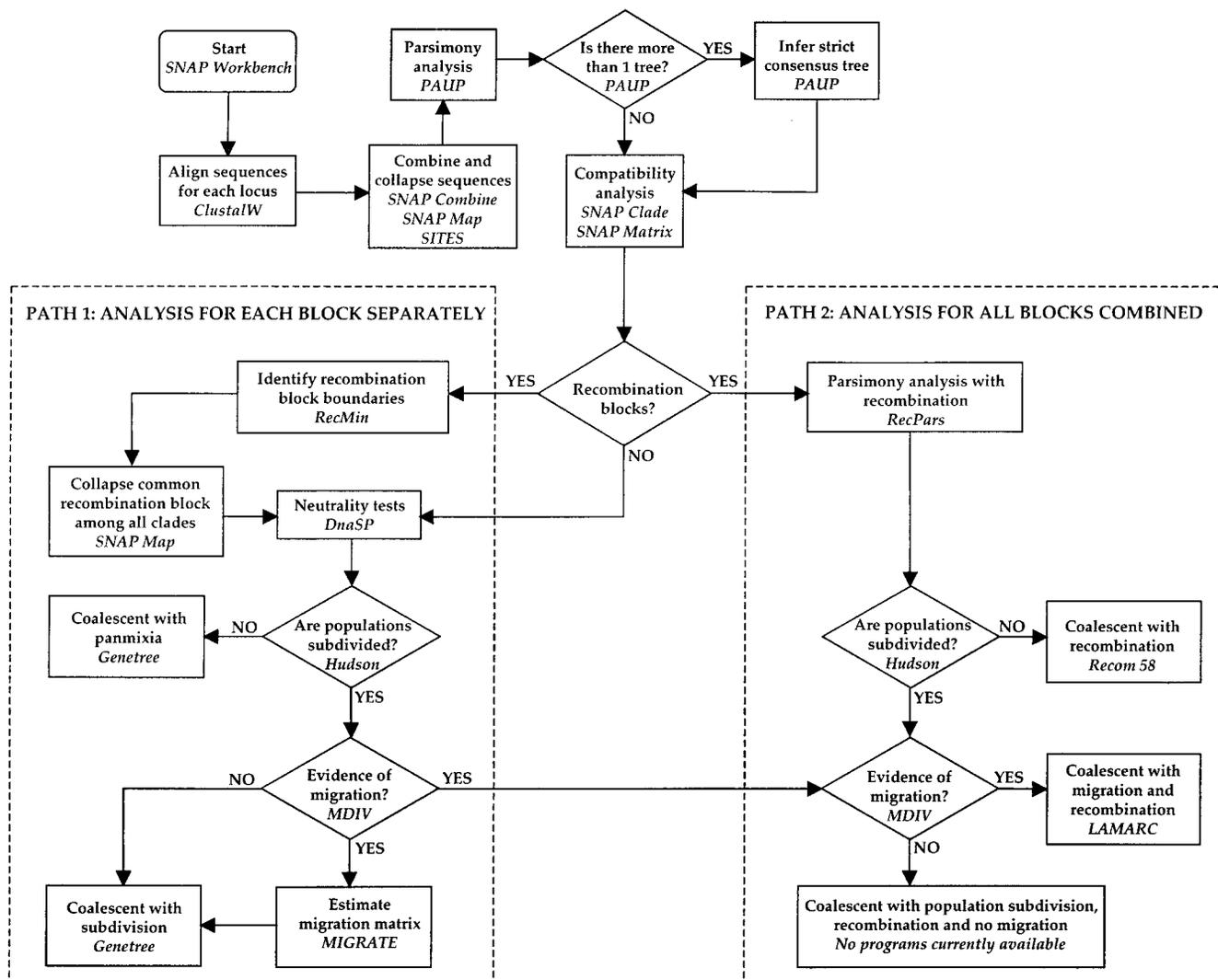


FIGURE 2.—Flowchart illustrating the sequence of steps and tools used in the analysis of the CHV-1. The analysis is guided by a series of questions shown in diamonds and, depending on the answers to these questions (yes or no), one or more analysis paths are possible. At the recombination block junction, two paths are possible depending on whether the analysis is performed for each recombination block separately (path 1) or for all blocks combined (path 2). The analyses for path 1 should be performed first so that inferences from MDIV, which assumes no recombination, can be used for path 2, which assumes recombination (see arrow connecting MDIV analyses in paths 1 and 2). The names of software tools that were used to perform specific analyses are italicized; *Hudson* refers to the programs Seqtomatrix and Permtest developed by R. Hudson. The presence of recombination increases the power of Hudson's tests to detect subdivision. This may be due to increased haplotype diversity in the total population or, more importantly, to recombination giving rise to unique haplotype blocks (see Figure 4) that distinguish the different subpopulations (clades). These tests may be more powerful in detecting population differences when certain assumptions of population processes are not satisfied or are unknown. The coalescent models implemented in MIGRATE, Rec58, and Genetree are positioned at the ends of the flowchart. These models make specific assumptions of the mutation process that must be validated *a priori* (see text for details). A coalescent model with population subdivision but no migration and with recombination has yet to be implemented.

Bergamo ( $n = 158$ ), 68% of the fungal isolates were in vc types EU-1 (33%) and EU-2 (35%; CORTESI *et al.* 1996). In contrast to fungal populations, we had no prior knowledge of diversity or population structure of CHV-1.

**Virus sampling:** To sample CHV-1, we screened *C. parasitica* isolates from Teano for dsRNA using a method described previously (MORRIS *et al.* 1983; LIU *et al.* 2003). *C. parasitica* isolates from Bergamo were screened for CHV-1 using an immunoblot method (PEEVER *et al.* 2000). For both samples, the presence of CHV-1 was confirmed by dsRNA isolation in

cellulose column chromatography (MORRIS and DODDS 1979; PEEVER *et al.* 1997). In addition to CHV-1 sampled from vc types EU-1 and EU-2 in Bergamo, we included one isolate from each of vc types EU-6, EU-15, and EU-17, which were isolated for a different study (PEEVER *et al.* 2000).

From each CHV-1 isolate, we obtained nucleotide sequences from three different regions of the genome (Figure 1): 471 bp from the 5'-noncoding region and 57 bp from the 5' end of open reading frame (ORF) A (referred to as NC); 283 bp from the 3' end of ORF A and the contiguous 481 bp from

TABLE 2

Reciprocal crosses with CHV-1-infected and virus-free isolates of *C. parasitica* to test for vertical transmission of viruses into ascospore progeny

Parents		No. of ascospore progeny assayed from	
CHV-1 infected (CHV-1 <sup>+</sup> )	Virus free (CHV-1 <sup>-</sup> )	CHV-1 <sup>+</sup> parent <sup>a</sup>	CHV-1 <sup>-</sup> parent
TE8	TE28	0 (0) <sup>b</sup>	68
TE8	TE72	0 (1)	ND
TE15	TE71	0 (0)	134
TE15	TE56	0 (0)	120
TE37	TE18	193 (3)	ND
TE37	TE34	261 (3)	ND
TE8 × TE37 <sup>c</sup>	—	175 (2)	—

ND, not done.

<sup>a</sup> Maternal parent from which ascospore progeny were obtained.

<sup>b</sup> Number of perithecia produced by CHV-1-infected maternal parent is shown in parentheses. Virus-free maternal isolates each produced >50 perithecia.

<sup>c</sup> Cross between two CHV-1-infected isolates.

the 5' end of ORF B (ORF A/B); and a 1096-bp segment approximately in the middle of ORF B. cDNA synthesis by reverse transcription was obtained using random hexamer primers as described (HILLMAN *et al.* 1992; CHUNG *et al.* 1994) and the GeneAmp RNA PCR kit (Perkin-Elmer-Cetus, Norwalk, CT). First-strand cDNA was synthesized from dsRNA primed with random hexamer primers. Oligonucleotide primers for amplifying three regions of CHV-1 by PCR (Table 1) anneal to conserved sequences and were designed to amplify a fragment from CHV-1 and the related hypovirus CHV-2 (HILLMAN *et al.* 2000). PCR was catalyzed with ExTaq polymerase (Mirus, Madison, WI) in the presence of 0.2 mM of each dNTP and 0.15 μM of the primers. cDNA template was denatured at 94° for 2 min. Thirty-three cycles of denaturation at 94° for 30 sec, annealing at 55° for 1.5 min, and polymerization at 72° for 1.5 min were conducted. Sequencing of both strands was done by the Cornell University Sequencing Facility or at Rutgers University.

**Virus cDNA sequence analysis:** The following steps in the analysis are outlined in the flowchart in Figure 2. The implementation of several programs in the flowchart was facilitated

using a Java program (SNAP Workbench) that manages and coordinates a series of other programs (PRICE and CARBONE 2003). SNAP Workbench was developed in the Carbone laboratory and is distributed over the Internet, <http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html>; additional programs in the SNAP (Suite of Nucleotide Analysis Programs) series are also accessible from this URL. An important functionality of our workbench software is the implementation of protocols consisting of a defined set of programs, assumptions, and parameter settings for following a particular path in the flowchart (*e.g.*, see paths 1 and 2 in Figure 2). SNAP Workbench provided a single graphical user interface for performing the analyses described below.

To begin the analysis, viral cDNA sequences corresponding to the coding strand of the dsRNA were aligned using CLUSTAL W version 1.7 (THOMPSON *et al.* 1994). Multiple DNA sequence alignments were generated for the three sequenced regions; alignments were accessioned in GenBank (AY125727–AY125813). Sequence data from multiple loci were combined into one alignment file using SNAP Combine (AYLOR and CARBONE 2003). Using SNAP Map (AYLOR and CARBONE 2003) and SITES version 1.1 (HEY and WAKELEY 1997) we collapsed sequences into unique haplotypes, recoding insertion or deletions (indels) and categorizing base substitutions as phylogenetically informative or uninformative, transitions *vs.* transversions, and replacement *vs.* synonymous amino acid changes in coding regions. A phylogenetic analysis with unweighted parsimony was performed using PAUP\* 4.0 (SWOFFORD 1998); if heuristic searches yielded two or more equally parsimonious trees, we inferred a strict consensus tree.

We generated site compatibility matrices for all subsets of haplotypes that share a common ancestor in the strict consensus tree using SNAP Clade (MARKWORDT *et al.* 2003). Compatibility methods examine the overall support or conflict among variable sites in a DNA sequence alignment and are useful for visual identification of partitions or blocks undergoing reticulate evolution (JAKOBSEN *et al.* 1997). In this article, we adapt compatibility methods for identifying recombination events among groups of two or more haplotypes or clades in a phylogeny. Compatibility matrices were graphically illustrated using SNAP Matrix (MARKWORDT *et al.* 2003) and examined for the existence of two or more sites showing an identical pattern of compatibility/incompatibility. Clusters of two or more identical sites in a matrix define a recombination block. If recombination blocks were evident we followed two possible analysis paths, as illustrated in the flowchart (Figure 2). In path 1, we identified recombination block boundaries and the minimum number of recombination events in a clade using RecMin (MYERS and GRIFFITHS 2003). We used the putative recombination boundaries identified within each recombining clade

t, transitions; v, transversions; i, phylogenetically informative sites; -, uninformative sites; r, nonsynonymous (*i.e.*, replacement) substitutions; s, synonymous substitutions.

<sup>a</sup> Inferred recombinant haplotype 12 (PC43) was excluded in the coalescent analysis of NC. The exclusion of haplotype 12 resulted in the elimination of variable site 13. The simple indel at site 2 was recoded as follows: 0 = T; 1 = -.

<sup>b</sup> Variation at site 2 violated the infinite sites model; *i.e.*, there were more than two segregating bases at this site. Because the coalescent analysis using Genetree assumes an infinite sites model (GRIFFITHS and TAVARÉ 1994), we had to exclude site 2 when performing the Genetree simulations. However, site 2 was included in the coalescent analysis when MIGRATE was used because this program assumes a finite sites model. Only variation at site 2 distinguished haplotypes 8 (HPC39) and 22 (PC43). The exclusion of site 2 in ORF A/B (Figure 6) resulted in the elimination of haplotype 22 from the sample and subsequent coalescence of PC43 to its ancestral haplotype 8. Inferred recombinant strain TE211 (haplotype 6) was excluded in the coalescent analysis using Genetree and MIGRATE because both programs assume no recombination in the history of the sample. The exclusion of haplotype 6 resulted in the elimination of variable site 38. Site numbers with an asterisk are compatible with each other in ORF A/B, but incompatible with other sites in the NC and ORF B regions (see Figure 4). Indel at site 25 was recoded as follows: 0 = TTTT; 1 = TTT-. Variation at sites 11 and 12 falls in the second and third codon positions.

<sup>c</sup> Haplotype number and frequency (*i.e.*, the number of occurrences of the haplotype in the sample) are shown in parentheses.



TABLE 4

Distribution of viral haplotypes, fungal host isolates, and vegetative compatibility types in the 5'-noncoding and ORF A/B regions of CHV-1 from Teano and Bergamo, Italy

Locus	Haplotype (frequency)	<i>C. parasitica</i> isolate <sup>a</sup>	vc type (EU-)
NC	1 (2)	TE151, TE195	10, 12
	2 (14)	TE15, PC88, TE211, HPC60, HPC27, HPC29, TE270, TE257, TE227, TE261, TE157, PC52, PC56, PC65	1, 2, 10, 12, 15
	3 (1)	PC29	1
	4 (1)	TE259	10
	5 (1)	HPC39	1
	6 (1)	PC75	6
	7 (2)	PC76, PC26	1, 2
	8 (1)	HPC24	2
	9 (2)	HPC4, HPC8	2, 17
	10 (1)	TE119	12
	11 (1)	TE165	10
	12 (1)	PC43	1
	13 (1)	PC89	2
ORF A/B	1 (2)	TE151, TE119	10, 12
	2 (1)	TE195	12
	3 (1)	TE15	12
	4 (1)	PC88	1
	5 (1)	PC29	1
	6 (1)	TE211	12
	7 (1)	TE259	10
	8 (1)	HPC39	1
	9 (2)	PC75, PC76	6, 1
	10 (1)	HPC60	1
	11 (1)	HPC27	1
	12 (1)	HPC24	2
	13 (1)	HPC29	15
14 (1)	TE270	12	
15 (1)	HPC4	2	
16 (1)	HPC8	17	
17 (1)	TE257	10	
18 (1)	TE165	10	
19 (1)	TE227	10	
20 (1)	TE261	10	
21 (1)	TE157	10	
22 (1)	PC43	1	
23 (2)	PC52, PC26	1, 2	
24 (1)	PC56	1	
25 (1)	PC65	2	
26 (1)	PC89	2	

vc, vegetative compatibility; NC, 5'-noncoding region.

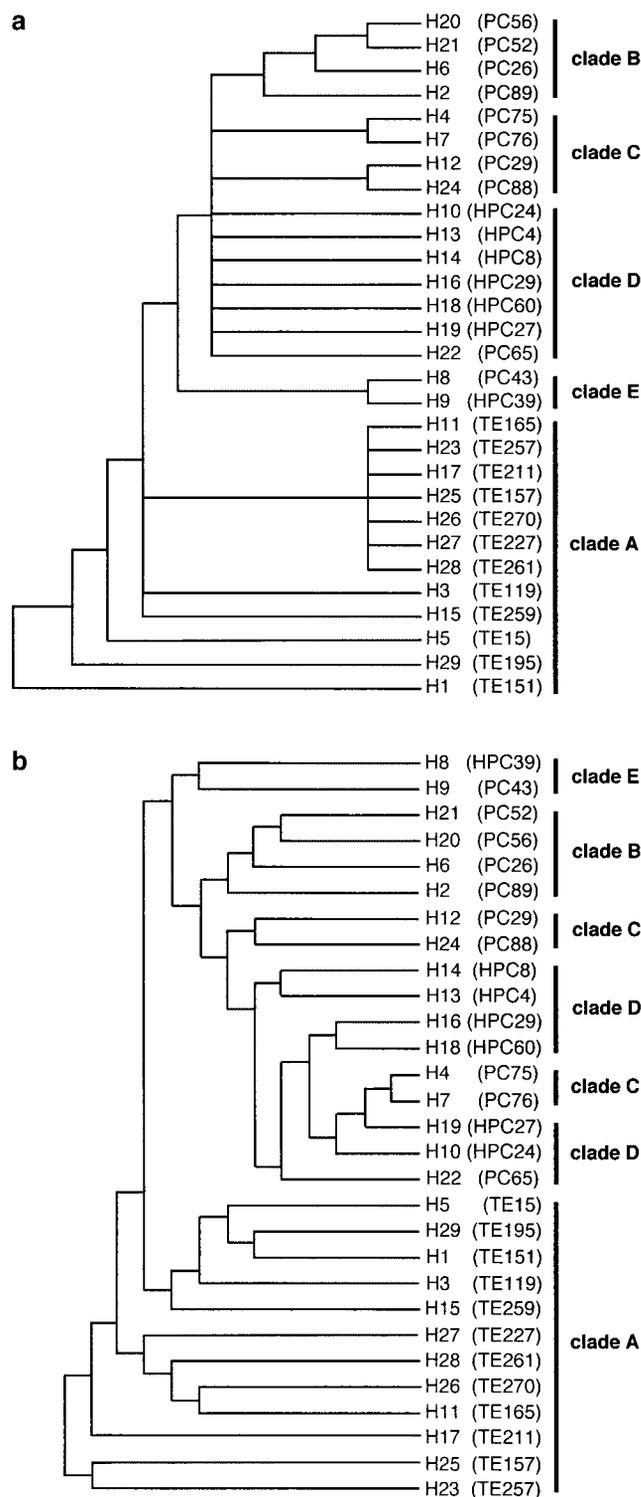
<sup>a</sup> *C. parasitica* isolate names beginning with TE are from Teano; those with PC or HPC are from Bergamo.

from RecMin as a guide to identify the common or shared recombination block boundaries among all clades. We then collapsed the common nonrecombining blocks among all clades into haplotypes using SNAP Map. In path 2, we used RecPars (HEIN 1990; <http://www.daimi.au.dk/~compbio/recpars/recpars.html>) to infer the number of recombinant segments (*i.e.*, recombination blocks) in the combined multilocus data set. The best phylogeny inferred using RecPars is the most parsimonious reconstruction for all phylogenetically informative sites within an inferred recombinant block and represents a minimal path through the ancestral recombination graph (SONG and HEIN 2003). By comparing the RecPars tree to the strict consensus tree we were able to recon-

struct phylogenetic relationships among haplotypes that were previously unresolved in the strict consensus tree and determine the relative timing of recombination events in the history of the sample.

**Virus tests of neutrality and population subdivision:** For each population and nonrecombining block we estimated the population mutation rate,  $\theta$ , from the number of segregating sites,  $s$  (WATTERSON 1975), and performed tests of neutrality using DNASP version 3.53 (ROZAS and ROZAS 1999). This provided *a priori* population parameter estimation and inferences on population growth patterns, *i.e.*, equilibrium *vs.* non-equilibrium, and evolutionary forces, *i.e.*, neutrality *vs.* selection. To test for population differentiation we used SNAP

Map to generate the appropriate sequence file, Seqtomatrix (HUDSON *et al.* 1992a) to convert the sequence file to a distance matrix, and Permtest (HUDSON *et al.* 1992a) to test for geographic subdivision in samples of DNA sequences from two or more localities; this latter test was done using a nonparametric permutation method based on Monte Carlo simulations. Hudson's test statistics ( $K_{ST}$ ,  $K_S$ , and  $K_T$ ) were calculated for each recombination block (path 1) and again for the entire data set (path 2). Testing the hypothesis of no genetic differentiation is an important first step and determines whether we assume a



panmictic or a subdivided population model for examining population divergence using Genetree (path 1) and recombination parameter estimation using Recom58 (path 2).

**Virus coalescent analysis:** If we found evidence for geographic subdivision, the observed patterns of genetic divergence were further explored to determine whether there was evidence for constant (*i.e.*, equilibrium) migration among populations or shared ancestral polymorphisms followed by population divergence but no migration. Two migration models were used for this analysis. First, MDIV was used to test for equilibrium migration *vs.* shared ancestral polymorphisms between two subdivided populations (NIELSEN and WAKELEY 2001). The approach of NIELSEN and WAKELEY (2001) uses either an infinite sites or a finite sites model without recombination and implements both likelihood and Bayesian methods using Markov chain Monte Carlo coalescent simulations for jointly estimating the population mean mutation rate ( $\theta$ ), divergence time ( $T$ ), migration rate ( $M$ ), and time since the most recent common ancestor (TMRCA) between two subdivided populations. This approach assumes that both populations descended from one panmictic population that may (or may not) have been followed by migration. Second, if MDIV showed evidence of equilibrium migration, MIGRATE was used to estimate migration rates assuming equilibrium migration rates (symmetrical or asymmetrical) in the ancestral history of populations (BEERLI and FELSENSTEIN 2001). Both MDIV and MIGRATE assume neutrality and no recombination and therefore these analyses must be restricted to neutral nonrecombining blocks in the data set (path 1). If we found evidence of equilibrium migration from our analyses in path 1, then we used this inference in path 2 (see Figure 2) to estimate migration and recombination rate parameters simultaneously using LAMARC ([http://evolution.genetics.washington.edu/lamarc/lamarc\\_about.html](http://evolution.genetics.washington.edu/lamarc/lamarc_about.html)). If we found evidence of population subdivision and no migration in path 1, then we reconstructed the ancestral history of the sample for each nonrecombining block using Genetree version 9.0 (<http://www.stats.ox.ac.uk/~griff/software.html>). There was no counterpart for the Genetree analysis in path 2 (see Figure 2). The genealogy with the highest root probability, the ages of mutations, and the TMRCA of the sample were estimated from coalescent simulations in path 1. We further identified the position (*i.e.*, coalescent time) of mutations giving rise to replacement and synonymous amino acid changes for substitutions in coding regions.

Finally, we estimated the recombination parameter,  $\rho$ , within each population using Recom58 (GRIFFITHS and MARJORAM 1996), which assumes no population subdivision. The magnitude of the recombination parameter is a measure of the likelihood of recombination over mutation at a specific

FIGURE 3.—Unrooted haplotype cladograms inferred from the combined NC, ORF A/B, and ORF B regions of CHV-1; indels and variable positions violating infinite sites were excluded. (a) The unrooted strict consensus haplotype tree inferred using parsimony alone. Each of the five clades has a distinct mutation and recombination history as inferred from site compatibility matrices (see Figure 4). No homoplasy was detected in clades B and E. Each haplotype is represented by a single CHV-1 isolate (shown in parentheses). Isolates with the prefix TE are from Teano; those with PC or HPC are from Bergamo. (b) The most parsimonious reconstruction of the same haplotypes in a assuming the existence of recombination in their evolutionary histories. Only clades B and E are monophyletic in both a and b.

site. Although a more efficient method for estimating the joint likelihood of  $\rho$  and  $\theta$  has been developed (FEARNHEAD and DONNELLY 2001), we were able to greatly reduce the computational time of *Recom58* by restricting our estimation of recombination rates within specific clades in the consensus tree. The novelty of our approach was in identifying clades with distinct

recombination blocks and testing for population subdivision before estimating recombination parameters (path 2). We were able to corroborate our inference of recombination block boundaries (path 1) by estimating the average number of recombination events per site in the consensus sequence for each clade (path 2). Specifically, the sites with the most events were compared to the position of putative recombination boundaries identified using the site compatibility analysis, as described above.

All coalescent analyses were greatly facilitated using our workbench software. *SNAP Workbench* executes multiple, computationally intensive programs simultaneously on a single machine, Linux cluster, or supercomputer. This functionality was particularly important for programs using coalescent simulations (*MIGRATE*, *LAMARC*, *MDIV*, *Recom58*, and *Genetree*) because many independent coalescent runs were necessary to ensure convergence.

**Vertical transmission of CHV-1 by sexual reproduction:** Inferences about horizontal transmission from estimates of virus migration are based on the assumption that viruses cannot move between *vc* types except by horizontal transmission. An alternative hypothesis is that viruses are vertically transmitted from virus-infected parents to offspring with recombinant *vic* genotypes during sexual reproduction. Vertical transmission of CHV-1 in *C. parasitica* occurs through asexual spores (RUSIN and SHAIN 1985; PEEVER *et al.* 2000) but these give rise to individuals with the same *vic* genotype as their parents. Vertical transmission reportedly does not occur through sexual spores (ascospores) of *C. parasitica* (ANAGNOSTAKIS 1982; ELLISTON 1985), which are often of recombinant *vic* genotypes, but documentation of this is poor. Therefore, to test this hypothesis we set up seven crosses, using three CHV-1-infected isolates and six virus-free isolates as parents (Table 2). Crosses were performed on autoclaved chestnut stems and analyzed as described previously (LIU and MILGROOM 1996). Four months after fertilization, we examined crosses for sexual fruiting bodies (perithecia) produced by each maternal parent. From each cross, we sampled 50–100 ascospore progeny from each of two or more perithecia. Progeny were sampled from all perithecia produced by CHV-1-infected maternal parents. Progeny isolates were assayed for CHV-1 by colony morphology by culturing on potato dextrose agar (PDA; Difco, Detroit); CHV-1-infected isolates of *C. parasitica* have characteristically less pigmentation when grown on PDA with low to moderate light (HILLMAN *et al.* 1990). Ascospore isolates sampled from a cross between two CHV-1-infected parents

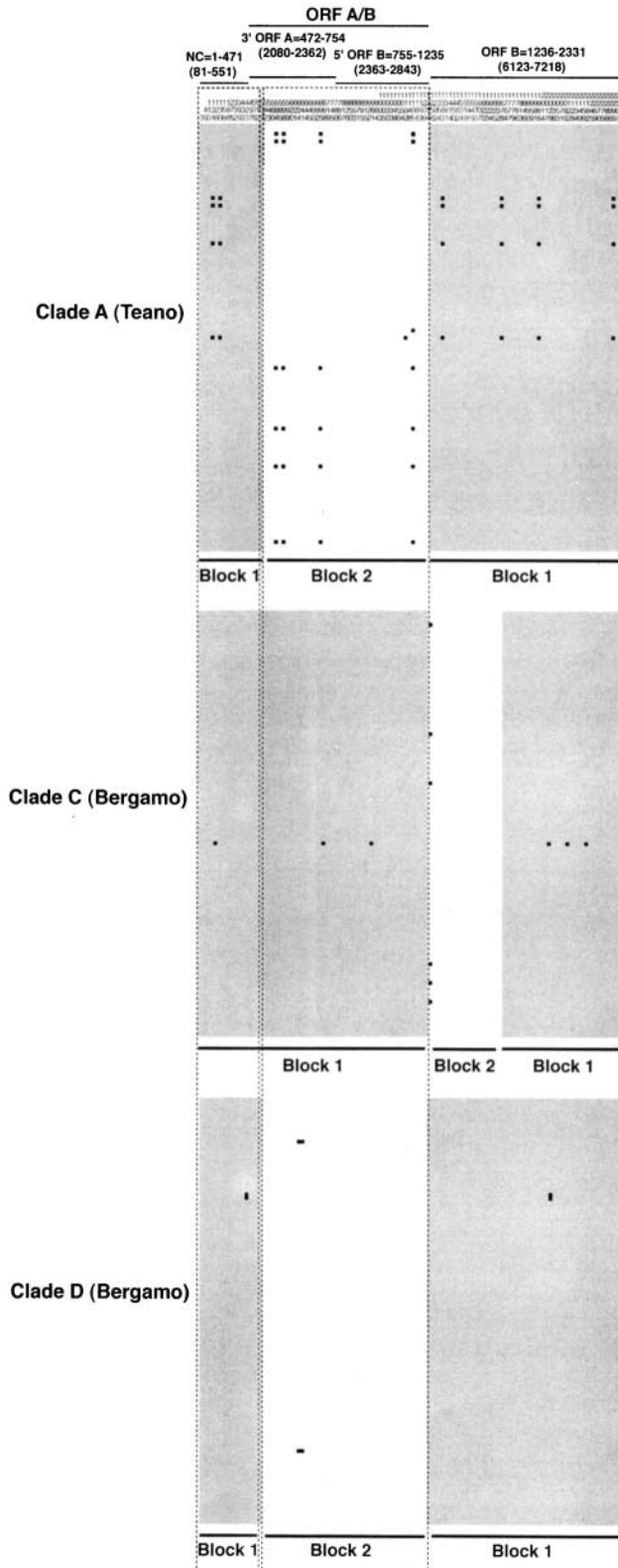


FIGURE 4.—Site compatibility matrices for homoplasious clades A, C, and D in the combined analysis of portions of the 5'-noncoding, ORF A/B, and ORF B regions in Teano and Bergamo, Italy. The numbers along the top of the matrix indicate variable positions in the multiple DNA sequence alignment of the combined data set; the second line of numbers in parentheses is nucleotide positions relative to CHV-1/Euro7 (GenBank accession no. AF082191). Incompatible sites are indicated by solid squares; all other sites in the matrix are compatible. The shaded and unshaded rectangles indicate the two inferred recombination blocks in each clade, designated numerically as blocks 1 and 2. The actual boundaries were determined by examining blocks among all clades. The dashed rectangles indicate the two regions (NC and ORF A/B) with common blocks spanning all clades. These regions were considered separately in coalescent analyses using *Genetree* (see Figures 7 and 8). Putative boundaries fall within the range of possible solutions as specified using *RecMin* (data not shown).

**TABLE 5**  
**Population statistics, neutrality tests, and migration estimates based on variation in NC, the 3' ORF A and 5' ORF B segments of ORF A/B, and the entire ORF A/B region of CHV-1 from Teano and Bergamo**

Population	Locus	Population statistics				Tests of neutrality <sup>b</sup>				2N <sub>c</sub> m estimates <sup>d</sup> (x = recipient)		
		l	n	s	k	Tajima's D statistic	Fu and Li's D* statistic	Fu and Li's Fu and Li's F** statistic	Fu's F <sub>3</sub> statistic	1, x	2, x	
1: Teano	NC	470	12	4	0.94	-1.023 (NS)	-0.459 (NS)	-0.681 (NS)	-0.753 (NS)	0.003 (1.429)	—	0.00
2: Bergamo		471	16	8	1.42	-1.487 (NS)	-1.281 (NS)	-1.536 (NS)	-4.216**	0.013 (6.198)	13.64	—
1: Teano	3' ORF A	283	11	9	2.07	-1.368 (NS)	-1.101 (NS)	-1.321 (NS)	-0.313 (NS)	0.007 (1.987)	—	1.21
2: Bergamo		283	17	15	3.35	-0.947 (NS)	-1.053 (NS)	-1.181 (NS)	-5.783***	0.037 (10.346)	11.71	—
1: Teano	5' ORF B	481	11	12	2.55	-1.645*	-2.127**	-2.268*	-7.663****	0.012 (5.532)	—	0.58
2: Bergamo		481	17	12	1.82	-1.825**	-1.637 (NS)	-1.949 (NS)	-3.004**	0.009 (4.555)	0.00	—
1: Teano	ORF A/B	764	11	21	4.62	-1.617*	-1.800 (NS)	-1.987 (NS)	-4.842***	0.008 (6.219)	—	1.08
2: Bergamo		763	17	27	5.17	-1.428 (NS)	-1.431 (NS)	-1.655 (NS)	-8.693****	0.019 (14.192)	3.41	—

Parameter estimates and neutrality tests were performed prior to coalescent-based inference using Genetree. *l*, consensus sequence length excluding sites with gaps; *n*, sample size; *s*, segregating nucleotide sites; *k*, average number of pairwise nucleotide differences. The average number (*k*) of pairwise differences between the *i*th and *j*th DNA sequences can be given by  $k = \sum \sum k_{ij} / \binom{n}{2}$ , where  $\binom{n}{2} = n(n-1)/2$ .  $\theta$ , population mean mutation rate, or Watterson's theta estimator, is defined as  $\theta = s / (1 + 1/2 + 1/3 + \dots + 1/n - 1)$ , where *s* is the number of segregating sites and *n* is the sample size (WATTERSON 1975). Watterson's estimate of theta is a simple maximum likelihood estimate of  $\theta$ ; it depends only on the number of segregating sites and the sample size and is not a function of the observed frequency of haplotypes; a dash indicates an undefined value. NS, not significant ( $P > 0.1$ ); \*0.05 <  $P < 0.1$ ; \*\*0.01 <  $P < 0.05$ ; \*\*\*0.001 <  $P < 0.01$ ; \*\*\*\* $P < 0.001$ .

<sup>a</sup> Watterson's estimate of theta,  $\theta = 2N_c\mu$ , per locus (WATTERSON 1975). The per-site population mean mutation rate is shown in parentheses.  
<sup>b</sup> Tajima's D [defined as  $D = X/\sqrt{\text{var}(x)}$ ], where  $X = k - s / (1 + 1/2 + 1/3 + \dots + 1/n - 1)$ , Fu and Li's D\* and F\*\* measure the departure from neutrality for departures from neutrality in a DNA region (Tajima 1989; Fu and Li 1993). These tests assume constant population size, no recombination, and no migration. Fu's F<sub>3</sub> tests whether the departure from neutrality is the result of population growth or background selection (Fu 1997).

<sup>c</sup> Coalescent-based estimate of theta,  $\theta = 2N_c\mu$ , per site (BEERLI and FELSENSTEIN 1999). The per locus population mean mutation rate is shown in parentheses.  
<sup>d</sup> The migration parameter, 2N<sub>c</sub>m, estimated using MIGRATE is defined as the effective number of migrants exchanged between two populations each generation, where N<sub>c</sub> is the effective population size and *m* is the migration rate per generation. The following is an example of how to read the migration parameter. For population 2 (Bergamo) in NC, 1, x means that immigration from population 1 (Teano) into Bergamo is 2N<sub>c</sub>m = 13.64 (a high level of migration). All migration estimates were checked by running the coalescent process using 10 different starting random number seeds; all runs yielded similar results.

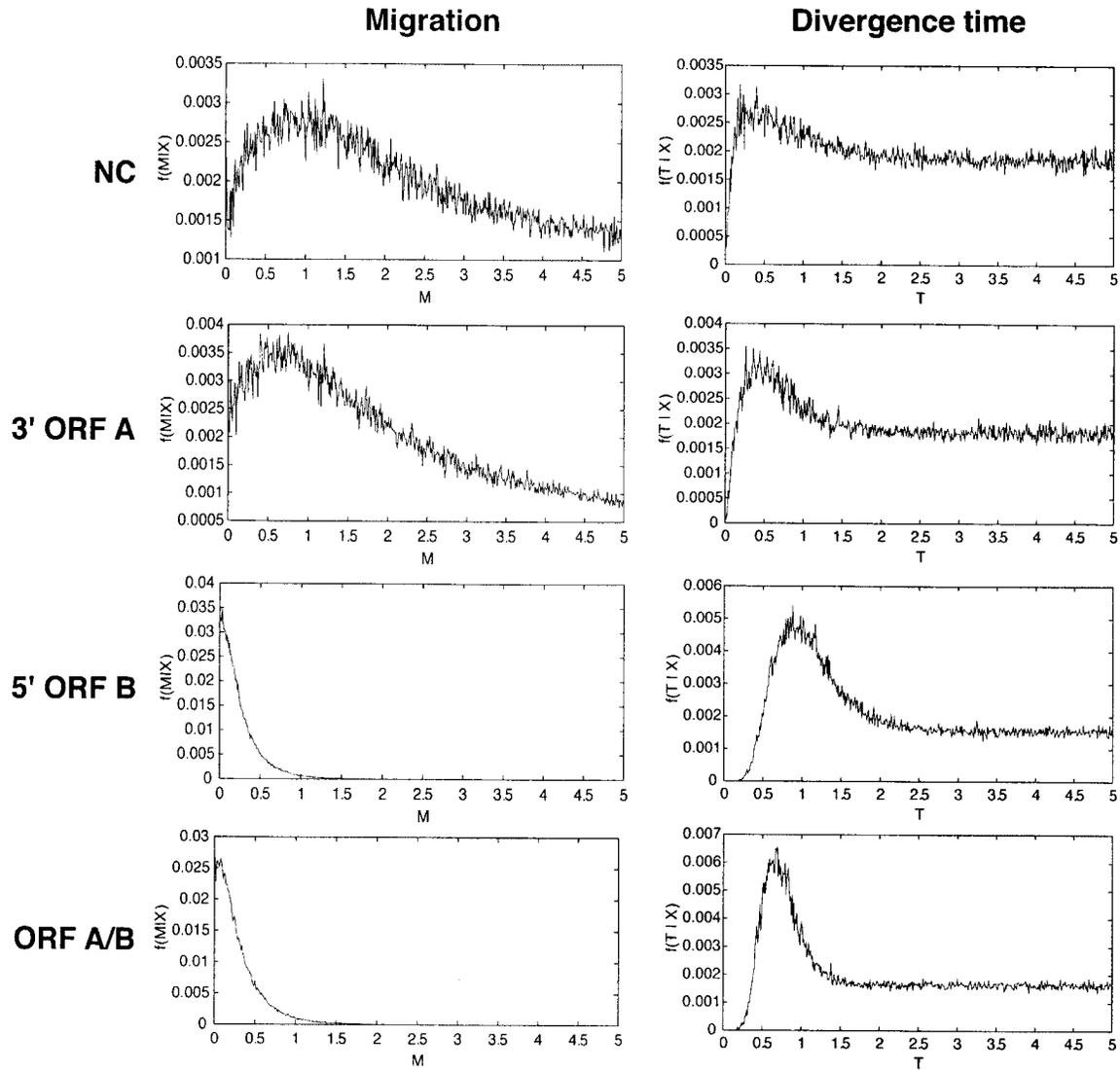


FIGURE 5.—The migration and divergence time posterior probability distributions between Teano and Bergamo estimated for each nonrecombining locus using MDIV. The data were simulated assuming an infinite sites model, 2,000,000 steps in the chain for estimating the posterior distribution and an initial 500,000 steps to ensure that enough genealogies were simulated before approximating the posterior distribution (NIELSEN and WAKELEY 2001). Ten independent replicates using different starting random number seeds were simulated under the same model and parameters; results were similar between runs. Data points from MDIV were plotted using gnuplot version 3.7 available from <http://www.gnuplot.info/>.

were also screened for dsRNA as described above (MORRIS *et al.* 1983; LIU *et al.* 2003).

## RESULTS

**Virus cDNA sequence analysis:** A total of 2331 nucleotides were sequenced from each of 12 and 17 isolates of CHV-1 in Teano and Bergamo, respectively. There were 14 variable sites in NC (Table 3), 51 in ORF A/B (Table 3), and 53 in ORF B (data not shown); CHV-1 isolates within each haplotype are shown in Table 4. Figure 3 shows unrooted cladograms of the combined NC, ORF A/B, and ORF B regions of CHV-1. A parsimony analysis with heuristic searching in PAUP (as-

sumes no recombination) resulted in 10 equally parsimonious trees each with a consistency index of 0.8321. The unrooted strict consensus tree is shown in Figure 3a. The most parsimonious reconstruction of the same haplotypes using RecPars (assumes recombination) is shown in Figure 3b. RecPars found two recombinant segments; however, the boundaries of these regions did not coincide with the boundaries inferred using site compatibility matrices (Figure 4). Since RecPars does not allow for indels and SNAP Clade does not allow for variable positions that violate an infinite sites model, both indels and variable positions violating infinite sites were excluded in parsimony analyses. This was necessary for direct comparison of parsimony trees inferred with

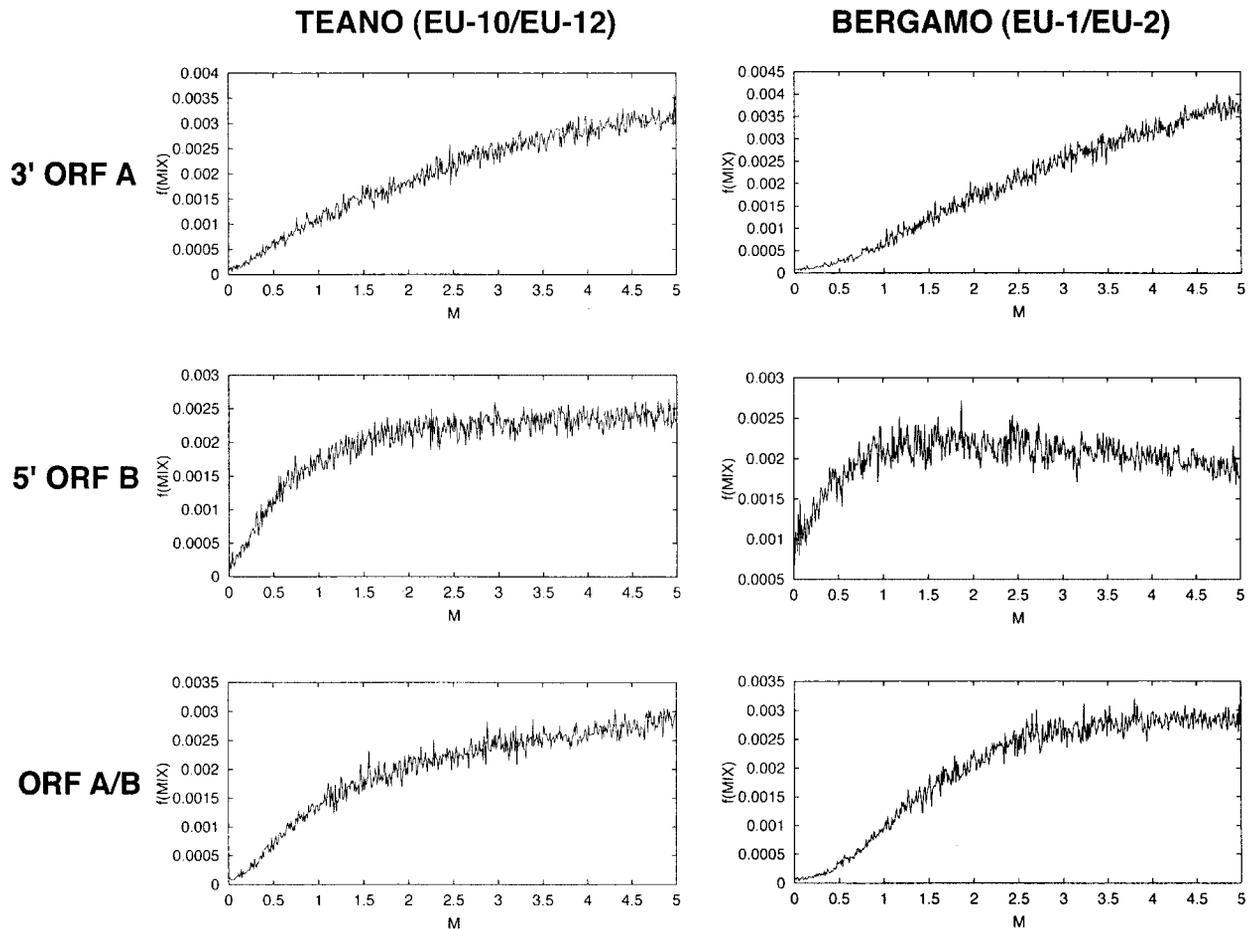


FIGURE 6.—The migration posterior probability distributions between the dominant vc types in Teano (EU-10 and EU-12) and Bergamo (EU-1 and EU-2) estimated for each nonrecombining locus using MDIV. The data were simulated assuming an infinite sites model, 2,000,000 steps in the chain with an initial 500,000 steps before estimating the posterior distribution. Ten independent replicates using different starting random number seeds were simulated under the same model and parameters; results were similar between runs. Data points from MDIV were plotted using gnuplot version 3.7.

and without recombination (Figure 3) and site compatibility matrices inferred using SNAP Clade (Figure 4). Clades were inferred using SNAP Clade and are defined as groups of two or more haplotypes with a distinct pattern of mutation and recombination in their history as illustrated in the site compatibility matrices in Figure 4. Distinct recombination blocks were detected in clades A, C, and D (Figure 4) in the strict consensus. All isolates from Teano (clade A) share a unique history of recombination and the two recombination blocks are well supported, with 6 sites (positions 121, 130, 1363, 1678, 1996, and 2299) in one block, spanning the NC and ORF B regions, that are incompatible with 4 sites (584, 598, 695, and 1165) in the other block, spanning ORF A/B (Figure 4). There was no evidence of recombination within each of these blocks in the entire data set. The actual sizes of these recombination blocks cannot be determined without sequencing the intervening regions between the NC, ORF A/B, and ORF B regions (Figure 4). In contrast to Teano, at least two different recombination histories (four distinct recombination blocks)

were inferred for the Bergamo isolates, with two recombination blocks in each of clades C and D (Figure 4). No homoplasy (*i.e.*, no incompatibility) was detected in clades B and E and only these clades were monophyletic in both parsimony trees (Figure 3).

The uncertainty in the branching patterns in the unresolved polytomies (clades A and D) in the strict consensus tree (Figure 3a) arises from an inherent limitation of the parsimony method itself, which assumes that there is no recombination in the history of the sequences. Therefore, we used RecPars (HEIN 1990) to find the most parsimonious reconstruction of our set of haplotypes using only base substitutions and inferred recombination events; RecPars cannot handle insertions or deletions. This analysis yielded a partial rooting of unrooted trees, which provided inference on ancestral recombination events in the sample. For example, the two largest unresolved polytomies observed in the strict consensus (clade D and a subset of the haplotypes in clade A) were further resolved when we assume recombination in the ancestral history of the sample (Fig-

TABLE 6

Viral migration estimates based on variation in the 3' ORF A and 5' ORF B segments of ORF A/B and the entire ORF A/B region of CHV-1 from the dominant vc types in Teano and Bergamo

Teano $2N_c m$ estimates <sup>a</sup> ( $x$ = recipient)					Bergamo $2N_c m$ estimates ( $x$ = recipient)					
vc type	Locus	$N$	$\theta$ ( $=2N_c\mu$ )	1, $x$	2, $x$	vc type	$n$	$\theta$ ( $=2N_c\mu$ )	1, $x$	2, $x$
1: EU-10	3' ORF A	7	$1.31 \times 10^{10}$	—	$9.12 \times 10^{14}$	1: EU-1	12	0.024	—	49.18
2: EU-12		4	0.011	$4.75 \times 10^{-8}$	—	2: EU-2	5	0.019	10.39	—
1: EU-10	5' ORF B	7	0.007	—	$4.09 \times 10^{-8}$	1: EU-1	12	0.024	—	21.71
2: EU-12		4	$3.31 \times 10^{10}$	$2.59 \times 10^{13}$	—	2: EU-2	5	0.004	0.00	—
1: EU-10	ORF A/B	7	0.003	—	$2.07$ (0.02) <sup>b</sup>	1: EU-1	12	0.015	—	$33.13$ (0.15)
2: EU-12		4	0.010	$1.13$ (0.60)	—	2: EU-2	5	0.015	$7.46$ (0.31)	—

A dash indicates an undefined value.

<sup>a</sup> All migration estimates were checked by running the coalescent process using 10 different starting random number seeds; all runs yielded similar results.

<sup>b</sup> Numbers in parentheses are the probabilities for transmission between the vc types in Teano and Bergamo based on laboratory estimates (CORTESI *et al.* 2001).

ure 3b); clade A, composed of all isolates from Teano, was the inferred root of the sample. The best parsimony tree inferred by RecPars shows how different clades sharing common recombination blocks are related to each other. For example, when we assume recombination (Figure 3b) clades C and D are split up into two different histories. This is consistent with the compatibility analysis, which revealed two distinct blocks in the ORF B region in clade C (Figure 4).

The results from the SNAP Clade and RecMin were used to determine the nonrecombining sequence blocks for coalescent analysis (see Figure 4). Specifically, we identified NC or ORF A/B as the regions that showed no intragenic recombination among all clades. In contrast, the putative recombination event falling in ORF B in clade C resulted in significant homoplasy. A phylogenetic analysis of ORF B resulted in 348 equally parsimonious trees. Even a small amount of homoplasy (C.I. = 0.857) yielded a significant amount of phylogenetic uncertainty; in this case, the strict consensus of all 348 trees was predominantly fan shaped, *i.e.*, with a topology similar to clade D in Figure 3a. If we examined only nonrecombining blocks with ORF B, there would be too few mutations within each block to yield any useful genealogy. Therefore, the sequences from ORF B were excluded from further analysis. From the remaining data we inferred 13 distinct haplotypes for NC and 26 for ORF A/B (Table 3). The distribution and characterization of base substitutions and indels in these regions are summarized in Table 3. In ORF A/B the ratio of nonsynonymous ( $r$ ) to synonymous ( $s$ ) substitutions was different within the two contiguous segments. In the 3' ORF A segment  $r/s = 11/13$ , whereas in the 5' ORF B segment  $r/s = 8/18$ .

#### Virus tests of neutrality and population subdivision:

*A priori* inferences of population processes were based on tests of neutrality (Tajima 1989; Fu and Li 1993; Fu 1997) that assume a constant population size, no

recombination, and no migration. A significant test result may suggest a departure from neutrality, the presence of population subdivision (SIMONSEN *et al.* 1995; Fu 1996), population growth, or background selection (Fu 1997). We distinguished among population growth and background selection by comparing Tajima's  $D$ , Fu and Li's  $D^*$  and  $F^*$ , and Fu's  $F_s$  tests. If only Fu and Li's  $D^*$  and  $F^*$  are significant this suggests background selection (Fu 1997). If only Fu's  $F_s$  is significant this suggests population growth or genetic hitchhiking. Estimates of  $D$  for NC and 3' ORF A were not significantly different from 0 ( $P > 0.1$ ; Table 5) within Teano and Bergamo; we cannot reject the hypothesis of selective neutrality for DNA sequence variation in these regions. The  $D$  values for 5' ORF B are only marginally significant ( $0.01 < P < 0.10$ ) and the  $F_s$  values are moderately to highly significant ( $0.001 < P < 0.05$ ). This was evidence in support of weak (negative) selection and population growth. In the presence of population subdivision the power of these tests decreases (Fu 1996) and alternative tests for population subdivision as proposed by HUDSON *et al.* (1992a) should be used to avoid making false positives. A significant  $P$  value ( $P < 0.0001$ ) for Hudson's tests ( $K_{ST} = 0.1021$ ,  $K_S = 11.5003$ ,  $K_T = 12.8079$ ) indicated that Teano and Bergamo were genetically differentiated (HUDSON *et al.* 1992a).

**Virus coalescent analysis:** The next objective was to determine whether there was any evidence of migration between the two localities. MDIV showed some evidence for intermediate levels of gene flow between Teano and Bergamo in NC ( $M = 1$ ) and 3' ORF A ( $M = 0.7$ ); however, the hypothesis of  $M = 0$  could not be rejected given the shape of the posterior distribution (Figure 5). The decrease in power may be the result of insufficient sequence variation and shorter sequences in NC ( $l = 470$ ,  $s = 12$ ,  $M = 1$ ,  $T = 0.3$ ,  $\theta = 2.3$ ) and 3' ORF B ( $l = 283$ ,  $s = 24$ ,  $M = 0.7$ ,  $T = 0.5$ ,  $\theta = 4.4$ ) *vs.* 5' ORF B ( $l = 481$ ,  $s = 24$ ,  $M = 0$ ,  $T = 1.0$ ,  $\theta = 5$ ) and ORF

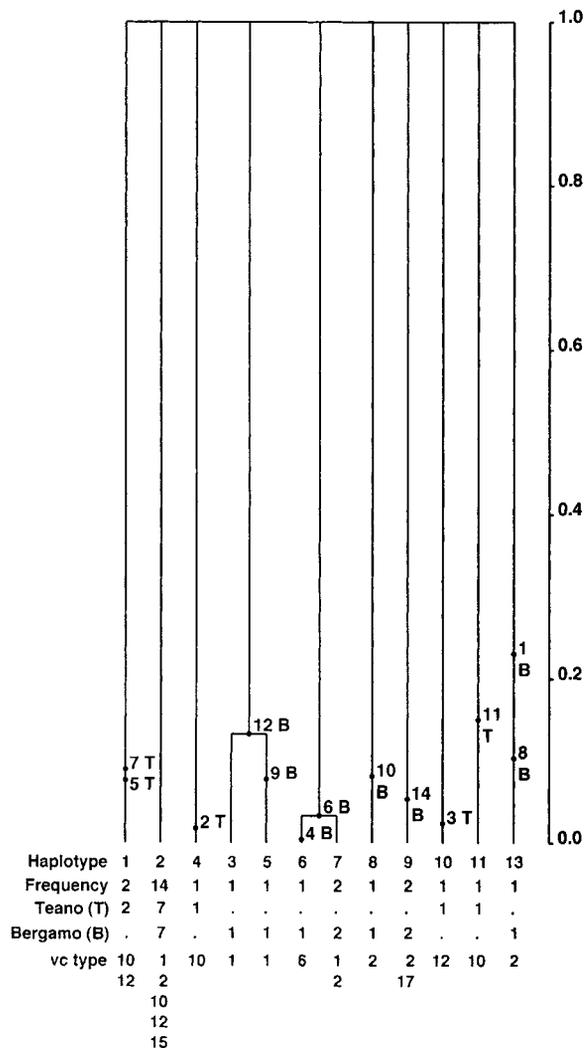


FIGURE 7.—Coalescent-based gene genealogy with the highest root probability (likelihood is  $6.8590 \times 10^{-15}$ , SD =  $2.9723 \times 10^{-13}$ ) showing the distribution of mutations in Teano (T) and Bergamo (B) for the NC region of CHV-1. The inferred genealogy is based on 10 different starting random number seeds, using 5 million simulations of the coalescent for each run, a Watterson's estimate of  $\theta = 2.0$ , and constant population sizes and growth rates. All runs converged to the same genealogy. The time scale is in coalescent units of effective population size. In the gene genealogy, the direction of divergence is from the top of the genealogy (oldest, *i.e.*, the past) to the bottom (youngest, *i.e.*, the present); coalescence is from the bottom (present) to the top (past). Since the gene genealogy is rooted, all the mutations and bifurcations are also time ordered from top to bottom. The numbers below the tree from top to bottom designate each distinct haplotype and its frequency (*i.e.*, the number of occurrences of the haplotype in the sample), the frequency of each haplotype in Teano and Bergamo, and the vc type associated with each haplotype (see Table 4).

A/B ( $l = 764$ ,  $s = 48$ ,  $M = 0$ ,  $T = 0.7$ ,  $\theta = 12$ ). The results from MIGRATE also showed no evidence of migration between Teano and Bergamo in the 5' ORF B region but some evidence of asymmetrical migration from

Teano to Bergamo in the NC, 3' ORF A, and the combined ORF A/B loci (Table 5). However, the inference of asymmetrical equilibrium migration (*i.e.*, migrants exchanged at a constant rate) from MIGRATE may not be valid because there is no evidence of migration between Teano and Bergamo (Figure 5). Collectively we interpret these results as evidence of shared ancestral polymorphisms between Teano and Bergamo but little or no detectable migration.

The hypothesis of no genetic differentiation of viruses between the dominant vc types in Teano (EU-10 and EU-12) and Bergamo (EU-1 and EU-2) could not be rejected using Hudson's tests ( $P = 0.524$  and  $P = 0.131$ , respectively). Although there was no evidence of subdivision we performed the analysis with MDIV (Figure 6) and MIGRATE (Table 6) to examine how these methods perform when there is no population subdivision. The migration posterior probability distributions estimated using MDIV did not reach a maximum (Figure 6). In Teano, parameter estimation using MIGRATE for 3' ORF A and 5' ORF B resulted in very high estimates for  $\theta$  and migration rates indicating that a maximum was not reached in the likelihood surface (Table 6). Our analyses of the combined ORF A/B region using MIGRATE suggested that migration between vc types was approximately equal in both directions in Teano and asymmetrical in Bergamo; however, these results are strongly influenced by the assumption of equilibrium migration in MIGRATE. The results from all analyses suggest unrestricted movement of CHV-1 between the dominant vc types in Teano and Bergamo.

In simulations using Genetree for NC (Figure 7) and ORF A/B (Figure 8) regions, we assumed a very low level of migration between Teano and Bergamo (close to 0) for the starting backward migration matrix. The coalescent-based genealogy inferred for NC did not have enough mutations (Figure 7) to resolve the Teano-Bergamo population splitting event. In contrast, the genealogy for ORF A/B was informative for inferring the mutational history of this region with respect to variation between populations and vc types within populations (Figure 8). On the basis of estimates of the mean ages of mutations from coalescent simulations, measured in units of  $N$  generations, the oldest mutations in Teano are 30, 1, and 17 with mean ages of mutations 0.726 (SD = 0.181), 0.421 (SD = 0.147), and 0.349 (SD = 0.141), respectively (Figure 8). These mutations are more likely to have arisen in Teano or in the progenitor of the founding population. The oldest mutation in Bergamo, 31, has a mean age of 0.569 (SD = 0.163). These results suggest two possible alternatives: (i) viral introduction and divergence started in the geographical area around Teano or (ii) viral divergence predates introduction into Teano; *i.e.*, the three ancestral haplotypes that existed at  $T = 0.8$  (see Figure 8) were immigrant haplotypes from a founding population in Asia. All mutations along branches in the genealogy are popu-

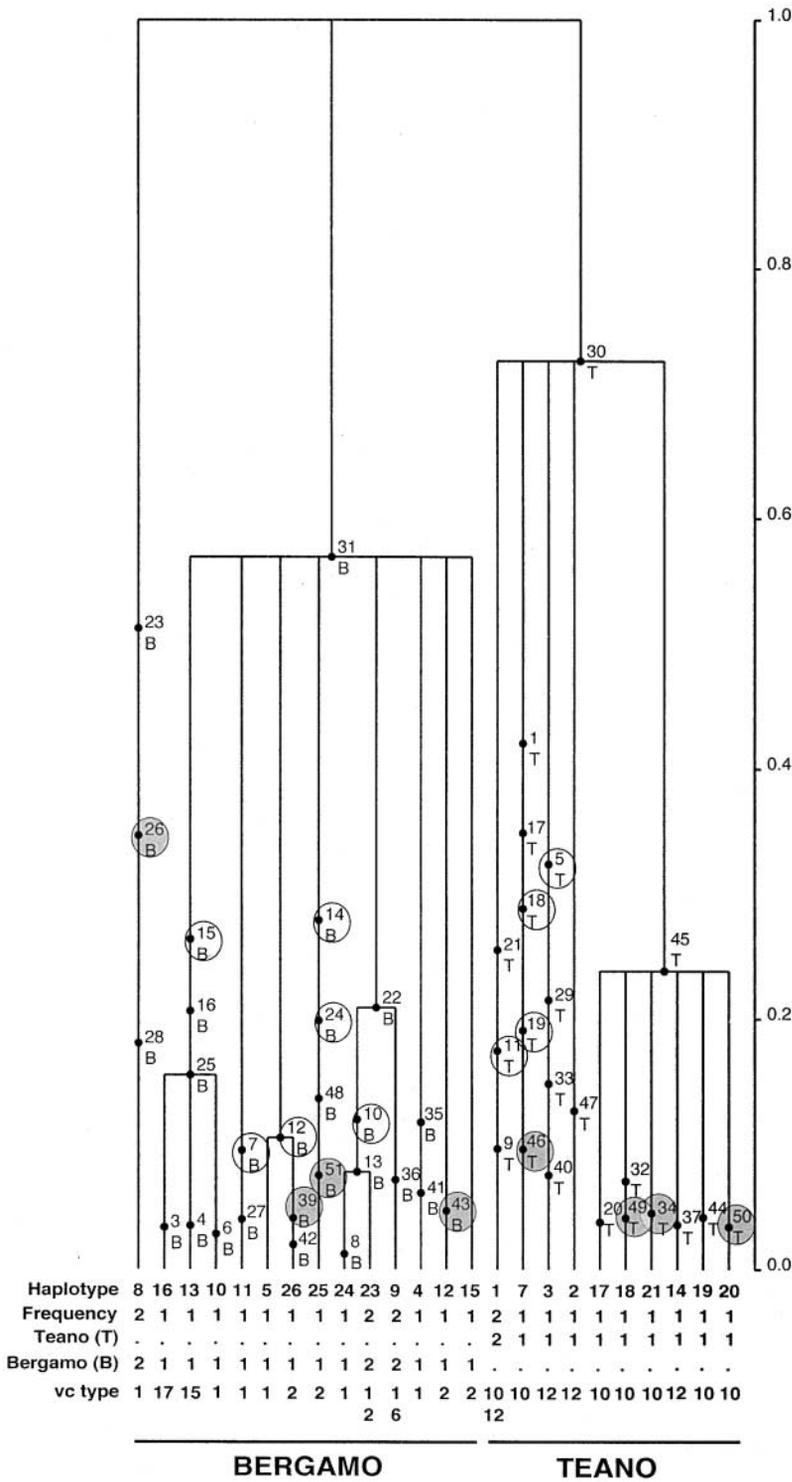


FIGURE 8.—Coalescent-based gene genealogy with the highest root probability (likelihood is  $1.9763 \times 10^{-35}$ , SD =  $3.5786 \times 10^{-33}$ ) showing the distribution of mutations and amino acid substitutions in Teano (T) and Bergamo (B) for the contiguous 3' ORF A and 5' ORF B loci (denoted collectively as ORF A/B) of CHV-1. The shaded circles designate replacement substitutions in 5' ORF B; the unshaded circles are for replacement substitutions in 3' ORF A. T and B in the circles denote mutations more likely to have arisen in Teano and Bergamo, respectively. Simulations were based on a Watterson's estimate of  $\theta = 7.0$ . See Figure 7 for further explanation.

lation specific and in some cases vc type specific as seen in Bergamo, *e.g.*, mutations 3 and 4, which are found in vc types EU-17 and EU-15, respectively. Finally, in simulations using Recom58, the recombination boundaries that we identified in Teano (positions 500 and 1200 in Figure 9) were found only in a subset of isolates from Bergamo (clade D; Figure 9). The existence of an

additional clade in Bergamo (clade C) with a recombination boundary around position 1700 (not found in Teano) suggests that viral evolution has been driven by local dynamics in Teano and Bergamo.

**Vertical transmission of CHV-1 by sexual reproduction:** We assayed a total of 951 ascospore progeny from seven reciprocal crosses to test for vertical transmission

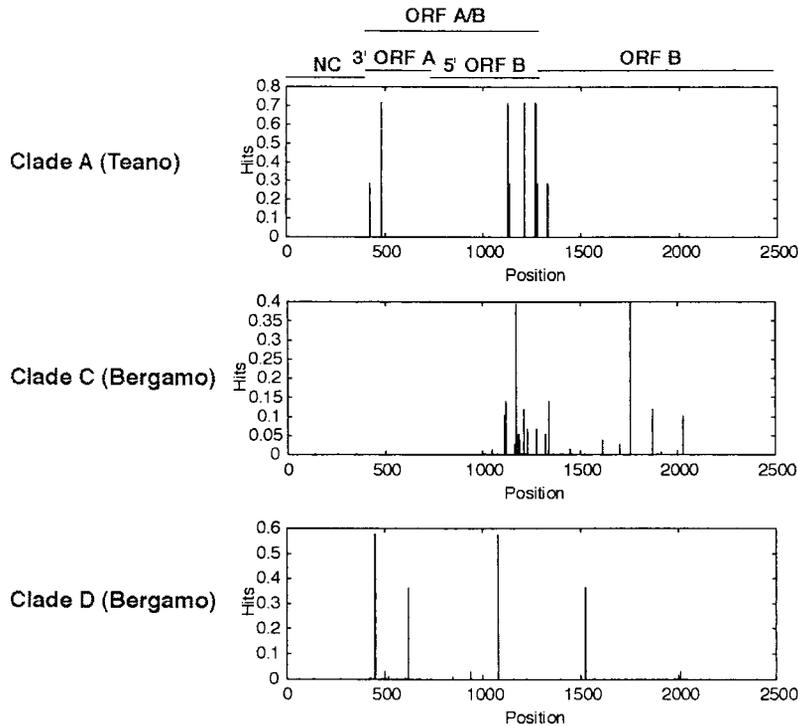


FIGURE 9.—The average number of recombination events (“hits”) per site in the consensus sequence for each clade based on coalescent analysis using *Recom58*, assuming recombination in the history of clades A, C, and D. The data were simulated assuming an infinite sites model with 1 million simulations of the coalescent,  $\theta = 7.0$ , a starting recombination parameter  $\rho = 0.5$ , and constant population sizes and growth rates. Ten different starting random number seeds were simulated under the same model and parameters; results were similar between runs. The simulated maximum likelihood estimates of  $\rho$  for clades A, C, and D were 1.0333, 0.8820, and 1.6385, respectively. The magnitude of the recombination parameter,  $\rho$ , is a measure of the likelihood of recombination over mutation at a specific site. In each of clades A, C, and D, sites with the most recombination hits correspond to putative recombination boundaries shown in Figure 4.

of CHV-1; 629 of these were from eight perithecia produced by CHV-1-infected maternal isolates (Table 2). None of the ascospore progeny contained CHV-1, supporting the hypothesis that CHV-1 is not vertically transmitted in sexual reproduction.

#### DISCUSSION

The main objective of this research was to estimate migration of viruses between host genotypes as a means of estimating horizontal virus transmission in nature. Overall, we found evidence for unrestricted migration of CHV-1 between dominant vc types in two populations. In Teano, results from Hudson’s tests and MDIV found no evidence of subdivision among vc types EU-10 and EU-12. This was corroborated using *MIGRATE*, which showed that migration was approximately equal in both directions, with slightly more migration from vc type EU-12 to vc type EU-10 than vice versa (Table 6). The ORF A/B gene genealogy (Figure 8) corroborated this result, showing a small asymmetry in migration in which vc type EU-10 was associated more frequently with the tip (*i.e.*, more recently derived) haplotypes (17, 18, 19, 20, and 21) than with interior (*i.e.*, ancestral) haplotypes (1 and 7). In contrast, EU-12 was associated more frequently with interior haplotypes (1, 2, and 3) than with tips (14). The gene genealogy based on the 5′-noncoding region was less informative because there were relatively fewer mutations in this region.

Similar to our analyses for Teano, results from Hudson’s tests and MDIV in Bergamo found no evidence

of subdivision among vc types EU-1 and EU-2. As above, this was corroborated using *MIGRATE*, which showed that migration between the dominant vc types was high and markedly asymmetric, with migration greater from EU-2 to EU-1 than vice versa (Table 6). Furthermore, the coalescent-based genealogy of ORF A/B (Figure 8) corroborates migration estimates from *MIGRATE*. The inferred ancestral haplotype 15, with no recently evolved mutations (Figure 8), is the interior-most haplotype in the gene genealogy and was isolated from vc type EU-2. Viral haplotypes associated with vc types EU-6, EU-15, and EU-17 are more recently evolved and share a common ancestor with viral haplotypes associated with EU-1. Consequently, the inferred order of haplotype divergence (for a representative of each vc type), from oldest to youngest, in Bergamo is 15, 9, (10, 13, 16); the haplotypes in parentheses diverged at approximately the same time.

To make inferences about horizontal transmission of viruses between vc types from migration analyses required that we test the assumption that viruses are not transmitted vertically to nonparental vc types during sexual reproduction; vertical transmission into recombinant *vic* genotypes would have the same effect on migration estimates as horizontal transmission. However, no viruses were transmitted into sexual offspring, even when both parents were virus infected (Table 2), supporting the few tests made previously (ANAGNOSTAKIS 1982; ELLISTON 1985). Therefore, migration of viruses between vc types must have occurred via horizontal transmission.

Approaching the study of virus transmission from the perspective of migration provides an alternative approach—resulting in different estimates—to laboratory studies on transmission. Migration estimates do not correlate well with the probability of transmission between *vc* types estimated in the laboratory (Table 6). For example, transmission between EU-10 and EU-12 in the laboratory is markedly asymmetric with greater transmission occurring from EU-10 to EU-12 (CORTESE *et al.* 2001). However, migration between these *vc* types is approximately equal in both directions (Table 6). In Bergamo, the asymmetry in migration is opposite to the asymmetry in virus transmission (Table 6). Furthermore, the average probability of transmission between dominant *vc* types (ignoring asymmetries) is greater in Teano than in Bergamo, but migration estimates are much larger within Bergamo than in Teano (Table 6). At least two different factors make these approaches fundamentally different. First, the coalescent-based migration analysis and laboratory studies provide estimates of transmission on different time scales. The coalescent estimates migration over the entire genealogical history of a locus and can assume either a constant or a varying population size. Laboratory studies, in contrast, are based on finite numbers of brief interactions between individuals to obtain estimates of probabilities of transmission per contact (*e.g.*, CORTESE *et al.* 2001). The second factor distinguishing these two approaches is that laboratory conditions are highly artificial, in which transmission may differ from natural conditions because of environmental factors affecting fungal physiology and cellular interactions. Coalescent-based methods integrate the transmission of viruses under natural conditions, over long periods of time, taking into account the interaction of numerous host individuals.

There was little or no detectable migration of CHV-1 between Teano and Bergamo as inferred from our analyses using MDIV and MIGRATE. The results from RecPars (Figure 3) suggest that the CHV-1 in Teano and Bergamo share a common ancestor that predates population divergence. The geographic distribution of mutations in the coalescent also supports the hypothesis that Teano is the older population. This suggests that the mutational structure of the founding viral population can be reconstructed more accurately from strains in Teano than from those in Bergamo. This does not, however, imply that Teano is the founding population. *C. parasitica* was first found in Europe (near Genoa, in northern Italy) in 1938 (BIRAGHI 1946). Assuming that CHV-1 was introduced with *C. parasitica* from Asia, these two populations were colonized within a relatively short period of time, either directly from Asia or from additional founder events within Italy. To test this hypothesis properly we would need to look at the same genomic regions in viral samples from Asia, the hypothesized centers of origin of the fungus and virus.

Genetic isolation and rare recombination in CHV-1

has the effect of creating distinct recombination blocks in evolutionarily older populations. Ancestral viral immigration to Bergamo initiated population subdivision such that the contemporary Teano and Bergamo populations are highly structured with restricted migration between them. Fungal subpopulations of *C. parasitica* in Teano and Bergamo are highly differentiated with respect to *vic* allele frequencies (MILGROOM and CORTESE 1999), with 44% of variation at these loci attributable to differences between populations (M. G. MILGROOM, unpublished data). Because fungal viruses are entirely dependent on their hosts for dispersal, restricted migration of *C. parasitica* likely restricts migration of CHV-1 also.

Homologous recombination is common in RNA viruses and has been extensively studied in the related Picornaviridae and in the plant-infecting Bromoviridae (BUJARSKI and NAGY 1996; AGOL 1997). Although recombination may occur anywhere along the viral genome, there may be preferred sites on the genome at which RNA recombination is more frequent (*e.g.*, BRUYERE *et al.* 2000). Current understanding of CHV-1 replication is poor, and there is no published information on preferred sites for homologous RNA recombination. While recombination in viruses is interesting *per se*, detection and localization of recombination events to specific clades in the phylogeny was an essential prerequisite for estimating migration using coalescence. Recombination blocks can extend over long stretches of nucleic acid and have been reported as a common phenomenon in the haplotype structure of fungi (CARBONE and KOHN 2001a,b) and the human genome (DALY *et al.* 2001; GABRIEL *et al.* 2002; STUMPF 2002). Recombination detected within ORF B of CHV-1 precluded its use for estimating migration. While it would have been possible to analyze smaller nonrecombining blocks within ORF B, these smaller partitions have too few characters for phylogenetic inference.

In addition to examining the distribution of migration, divergence, and mutation events in the ancestral histories of CHV-1 in Teano and Bergamo, the gene genealogies based on the coalescent allow us to examine patterns of divergence at the amino acid level in ORF A/B. Although the mutation rates are constant in 3' ORF A and 5' ORF B (see estimates of  $\theta$  and Tajima's *D* in Table 5), the ratio of nonsynonymous to synonymous substitutions is different between these two loci. In 3' ORF A,  $r/s = 11/13$ , suggesting selective neutrality, whereas in 5' ORF B  $r/s = 8/18$ , suggesting negative selection. Interestingly, the replacement substitutions in 5' ORF B (shaded circles in Figure 8) evolved more recently than replacement substitutions in 3' ORF A (unshaded circles in Figure 8). This suggests that negative selection may play an important role, in addition to drift, in shaping contemporary viral evolution in nature (also recently shown as an important diversifying mechanism in HIV-1 by KILS-HÜTTEN *et al.* 2001). Perhaps this

represents a mechanism that allows invading haplotypes to adapt to a new environment or resident haplotypes to adapt to a changing environment. These patterns can also be explained using a neutral mutation hypothesis whereby deleterious or beneficial mutations arise spontaneously and then either are purged or become fixed in the population.

Genealogical analysis has proven to be a powerful tool in this study for providing insight into the transmission dynamics of a virus in two host populations. An important advantage of estimating migration rates using the coalescent is that the accuracy of migration estimates is limited only by the number of informative characters for phylogenetic reconstruction. In this study, the coalescent provides a good approximation of the ancestral history even though the sample is small and may fail to capture the full range of genetic heterogeneity. Coalescent methods are powerful because they go beyond variation that is present in the sample to infer missing or ancestral genotypes in the population. At present, coalescent models have been developed that can accommodate the mutation process with or without migration (Genetree), recombination (Recom58), migration and recombination (LAMARC), or selection (NEUHAUSER and KRONE 1997) but not a combination of all these forces occurring simultaneously. The isolation of Teano and Bergamo coupled with viral recombination, purifying selection, and unrestricted migration among vc types within populations cannot be examined in a holistic fashion using the available coalescent models. LAMARC, for example, assumes equilibrium migration and recombination throughout the entire coalescent history of the sample of DNA sequences. An appropriate model would need to distinguish between shared ancestral polymorphisms, migration, recombination, and selection on several different temporal and spatial scales: among haplotypes, clades, and populations. A future challenge will be the development of multilocus coalescent models that can deal with recombination, migration, and selection, simultaneously. On the basis of the viral dynamics elucidated in this study, these models will be key in reconstructing the evolutionary dynamics of viral populations.

We thank Peter Beerli, Robert Griffiths, Linda Kohn, and Rasmus Nielsen for helpful comments and discussion on earlier drafts of this manuscript. We also thank Tobin Peever, Daniel Saha, and Alex Kay for help with screening for viruses in *C. parasitica* and Robert Foglia and Marcus Riedhammer for help with sequencing. This research was funded in part by the U.S. Department of Agriculture National Research Initiatives Competitive Grants Program grants 95-37303-1708 and 98-35303-6431 to M.G.M. and by McIntire-Stennis funds to B.I.H.

#### LITERATURE CITED

- AGOL, V. I., 1997 Recombination and other genomic rearrangements in picornaviruses. *Semin. Virol.* **8**: 77–84.
- ALLEMANN, C., P. HOEGGER, U. HEINIGER and D. RIGLING, 1999 Ge-

- netic variation of Cryphonectria hypoviruses (CHV1) in Europe, assessed using RFLP markers. *Mol. Ecol.* **8**: 843–854.
- ANAGNOSTAKIS, S. L., 1982 Biological control of chestnut blight. *Science* **215**: 466–471.
- ANAGNOSTAKIS, S. L., B. HAU and J. KRANZ, 1986 Diversity of vegetative compatibility groups of *Cryphonectria parasitica* in Connecticut and Europe. *Plant Dis.* **70**: 536–538.
- ANDERSON, R. M., and R. M. MAY, 1986 The invasion, persistence and spread of infectious diseases within animal and plant communities. *Philos. Trans. R. Soc. Lond. B* **314**: 533–570.
- AYLOR, D., and I. CARBONE, 2003 SNAP combine and map. Department of Plant Pathology, North Carolina State University, Raleigh, NC (<http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html>).
- BAHLO, M., and R. C. GRIFFITHS, 2000 Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* **57**: 79–95.
- BEERLI, P., and J. FELSENSTEIN, 1999 Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**: 763–773.
- BEERLI, P., and J. FELSENSTEIN, 2001 Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* **98**: 4563–4568.
- BIELLA, S., M. L. SMITH, J. R. AIST, P. CORTESI and M. G. MILGROOM, 2002 Programmed cell death correlates with virus transmission in a filamentous fungus. *Proc. R. Soc. Lond. Ser. B* **269**: 2269–2276.
- BIRAGHI, A., 1946 Il cancro del castagno causato da *Endothia parasitica*. *Ital. Agric.* **7**: 1–9.
- BRUYERE, A., M. WANTROBA, S. FLASINSKI, A. DZIANOTT and J. J. BUJARSKI, 2000 Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. *J. Virol.* **74**: 4214–4219.
- BUJARSKI, J. J., and P. D. NAGY, 1996 Different mechanisms of homologous and non-homologous recombination in brome mosaic virus: role of RNA sequences and replicase proteins. *Semin. Virol.* **7**: 363–372.
- BULL, J. J., 1994 Virulence. *Evolution* **48**: 1423–1437.
- CARBONE, I., and L. M. KOHN, 2001a A microbial population-species interface: nested cladistic and coalescent inference with multilocus data. *Mol. Ecol.* **10**: 947–964.
- CARBONE, I., and L. M. KOHN, 2001b Multilocus nested haplotype networks extended with DNA fingerprints show common origin and fine-scale, ongoing genetic divergence in a wild microbial metapopulation. *Mol. Ecol.* **10**: 2409–2422.
- CATEN, C. E., 1972 Vegetative incompatibility and cytoplasmic infection in fungi. *J. Gen. Microbiol.* **72**: 221–229.
- CHUNG, P. H., P. J. BEDKER and B. I. HILLMAN, 1994 Diversity of *Cryphonectria parasitica* hypovirulence-associated double-stranded RNAs within a chestnut population in New Jersey. *Phytopathology* **84**: 984–990.
- CONGDON, B. C., J. F. PIATT, K. MARTIN and V. L. FRIESEN, 2000 Mechanisms of population differentiation in marbled murrelets: historical versus contemporary processes. *Evolution* **54**: 974–986.
- CORTESI, P., and M. G. MILGROOM, 1998 Genetics of vegetative incompatibility in *Cryphonectria parasitica*. *Appl. Environ. Microbiol.* **64**: 2988–2994.
- CORTESI, P., M. G. MILGROOM and M. BISIACH, 1996 Distribution and diversity of vegetative compatibility types in subpopulations of *Cryphonectria parasitica* in Italy. *Mycol. Res.* **100**: 1087–1093.
- CORTESI, P., C. E. MCCULLOCH, H. SONG, H. LIN and M. G. MILGROOM, 2001 Genetic control of horizontal virus transmission in the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics* **159**: 107–118.
- DALY, M. J., J. D. RIOUX, S. F. SCHAFFNER, T. J. HUDSON and E. S. LANDER, 2001 High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**: 229–232.
- ELLISTON, J. E., 1985 Preliminary evidence for two debilitating cytoplasmic agents in a strain of *Endothia parasitica* from western Michigan. *Phytopathology* **75**: 170–173.
- FAHIMA, T., P. KAZMIERCZAK, D. R. HANSEN, P. PFEIFFER and N. K. VAN ALFEN, 1993 Membrane-associated replication of an unencapsidated double-strand RNA of the fungus, *Cryphonectria parasitica*. *Virology* **195**: 81–89.
- FEARNHEAD, P., and P. DONNELLY, 2001 Estimating recombination rates from population genetic data. *Genetics* **159**: 1299–1318.

- FU, Y.-X., 1996 New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**: 557–570.
- FU, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GABRIEL, S. B., S. F. SCHAFFNER, H. NGUYEN, J. M. MOORE, J. ROY *et al.*, 2002 The structure of haplotype blocks in the human genome. *Science* **296**: 2225–2229.
- GLASS, N. L., and I. KANEKO, 2003 Fatal attraction: nonself recognition and heterokaryon incompatibility in filamentous fungi. *Eukaryot. Cell* **2**: 1–8.
- GRIFFITHS, A. J. F., S. R. KRAUS, R. BARTON, D. A. COURT, C. J. MYERS *et al.*, 1990 Heterokaryotic transmission of senescence plasmid DNA in *Neurospora*. *Cult. Genet.* **17**: 139–145.
- GRIFFITHS, R., and P. MARJORAM, 1996 Ancestral inference from samples of DNA sequences with recombination. *J. Comp. Biol.* **3**: 479–502.
- GRIFFITHS, R. C., and S. TAVARÉ, 1994 Ancestral inference in population genetics. *Stat. Sci.* **9**: 307–319.
- HARTL, D. L., E. R. DEMPSTER and S. W. BROWN, 1975 Adaptive significance of vegetative incompatibility in *Neurospora crassa*. *Genetics* **81**: 553–569.
- HEATH, D. D., C. BUSCH, J. KELLY and D. Y. ATAGI, 2002 Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Mol. Ecol.* **11**: 197–214.
- HEIN, J., 1990 Reconstructing evolution of sequences subject to recombination using parsimony. *Math. Biosci.* **98**: 185–200.
- HEY, J., and J. WAKELEY, 1997 A coalescent estimator of the population recombination rate. *Genetics* **145**: 833–846.
- HILLMAN, B. I., R. SHAPIRA and D. L. NUSS, 1990 Hypovirulence-associated suppression of host functions in *Cryphonectria parasitica* can be partially relieved by high light intensity. *Phytopathology* **80**: 950–956.
- HILLMAN, B. I., Y. TIAN, P. J. BEDKER and M. P. BROWN, 1992 A North American hypovirulent isolate of the chestnut blight fungus with European isolate-related dsRNA. *J. Gen. Virol.* **73**: 681–686.
- HILLMAN, B. I., D. W. FULBRIGHT, D. L. NUSS and N. K. VAN ALFEN, 2000 *Hypoviridae*, pp. 515–520 in *Virus Taxonomy: Seventh Report of the International Committee for the Taxonomy of Viruses*, edited by M. H. V. v. REGENMORTEL, C. M. FAUQUET, D. H. L. BISHOP, E. B. CARSTENS, M. K. ESTES *et al.* Academic Press, San Diego.
- HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992a A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992b Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- JAKOBSEN, I. B., S. R. WILSON and S. EASTEAL, 1997 The partition matrix: exploring variable phylogenetic signals along nucleotide sequence alignments. *Mol. Biol. Evol.* **14**: 474–484.
- KILS-HÜTTEN, L., R. CHEYNIER, S. WAIN-HOBSON and A. MEYERHANS, 2001 Phylogenetic reconstruction of intrapatient evolution of human immunodeficiency virus type 1: predominance of drift and purifying selection. *J. Gen. Virol.* **82**: 1621–1627.
- KINNISON, M. T., P. BENTZEN, M. J. UNWIN and T. P. QUINN, 2002 Reconstructing recent divergence: evaluating nonequilibrium population structure in New Zealand chinook salmon. *Mol. Ecol.* **11**: 739–754.
- KOONIN, E. V., G. H. CHOI, D. L. NUSS, R. SHAPIRA and J. C. CARRINGTON, 1991 Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. *Proc. Natl. Acad. Sci. USA* **88**: 10647–10651.
- LEVIN, B. R., 1996 The evolution and maintenance of virulence in microparasites. *Emerg. Infect. Dis.* **2**: 93–102.
- LIU, Y.-C., and M. G. MILGROOM, 1996 Correlation between hypovirus transmission and the number of vegetative incompatibility (*vic*) genes different among isolates from a natural population of *Cryphonectria parasitica*. *Phytopathology* **86**: 79–86.
- LIU, Y.-C., P. CORTESI, M. L. DOUBLE, W. L. MACDONALD and M. G. MILGROOM, 1996 Diversity and multilocus genetic structure in populations of *Cryphonectria parasitica*. *Phytopathology* **86**: 1344–1351.
- LIU, Y.-C., R. DURRETT and M. G. MILGROOM, 2000 A spatially-structured stochastic model to simulate heterogeneous transmission of viruses in fungal populations. *Ecol. Model.* **127**: 291–308.
- LIU, Y.-C., B. I. HILLMAN, D. LINDER-BASSO, S. KANEKO and M. G. MILGROOM, 2003 Evidence for interspecies transmission of viruses in natural populations of filamentous fungi in the genus *Cryphonectria*. *Mol. Ecol.* **12**: 1619–1628.
- MARKWORDT, J., R. DOSHI and I. CARBONE, 2003 SNAP clade and matrix. Department of Plant Pathology, North Carolina State University, Raleigh, NC (<http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html>).
- MILGROOM, M. G., 1995 Population biology of the chestnut blight fungus, *Cryphonectria parasitica*. *Can. J. Bot.* **73** (Suppl. 1): S311–S319.
- MILGROOM, M. G., and P. CORTESI, 1999 Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proc. Natl. Acad. Sci. USA* **96**: 10518–10523.
- MILGROOM, M. G., K. WANG, Y. ZHOU, S. E. LIPARI and S. KANEKO, 1996 Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**: 179–190.
- MORRIS, T. J., and J. A. DODDS, 1979 Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* **69**: 854–858.
- MORRIS, T. J., J. A. DODDS, B. HILLMAN, R. L. JORDAN, S. A. LOMMEL *et al.*, 1983 Viral specific dsRNA: diagnostic value for plant virus disease identification. *Plant Mol. Biol. Rep.* **1**: 27–30.
- MYERS, S. R., and R. C. GRIFFITHS, 2003 Bounds on the minimum number of recombination events in a sample history. *Genetics* **163**: 375–394.
- NEUHAUSER, C., and S. M. KRONE, 1997 The genealogy of samples in models with selection. *Genetics* **145**: 519–534.
- NIELSEN, R., and J. WAKELEY, 2001 Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- NUSS, D. L., 1992 Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenesis. *Microbiol. Rev.* **56**: 561–576.
- PEEVER, T. L., Y.-C. LIU and M. G. MILGROOM, 1997 Diversity of hypoviruses and other double-stranded RNAs in *Cryphonectria parasitica* in North America. *Phytopathology* **87**: 1026–1033.
- PEEVER, T. L., Y.-C. LIU, K. WANG, B. I. HILLMAN, R. FOGLIA *et al.*, 1998 Incidence and diversity of double-stranded RNAs infecting the chestnut blight fungus, *Cryphonectria parasitica*, in China and Japan. *Phytopathology* **88**: 811–817.
- PEEVER, T. L., Y.-C. LIU, P. CORTESI and M. G. MILGROOM, 2000 Variation in tolerance and virulence in the chestnut blight fungus-hypovirus interaction. *Appl. Environ. Microbiol.* **66**: 4863–4869.
- PRICE, E. W., and I. CARBONE, 2003 SNAP workbench. Department of Plant Pathology, North Carolina State University, Raleigh, NC (<http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html>).
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- RUSSIN, J. S., and L. SHAIN, 1985 Disseminative fitness of *Endothia parasitica* containing different agents for cytoplasmic hypovirulence. *Can. J. Bot.* **65**: 54–57.
- SHAPIRA, R., G. H. CHOI and D. L. NUSS, 1991 Virus-like genetic organization and expression strategy for a double-stranded RNA genetic element associated with biological control of chestnut blight. *EMBO J.* **10**: 731–739.
- SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413–429.
- SONG, Y. S., and J. HEIN, 2003 Parsimonious reconstruction of sequence evolution and haplotype blocks: finding the minimum number of recombination events, pp. 287–302 in *Algorithms in Bioinformatics*, edited by G. BENSON and R. D. M. PAGE. Proceedings of the Third International Workshop, WABI, Budapest.
- STUMPF, M. P., 2002 Haplotype diversity and the block structure of linkage disequilibrium. *Trends Genet.* **18**: 226–228.
- SWOFFORD, D. L., 1998 *PAUP\**. *Phylogenetic Analysis Using Parsimony (\* and Other Methods)*, Version 4.0. Sinauer Associates, Sunderland, MA.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TEMPLETON, A. R., 1998 Nested clade analyses of phylogeographic

- data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**: 381–397.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- VAN ALFEN, N. K., R. A. JAYNES, S. L. ANAGNOSTAKIS and P. R. DAY, 1975 Chestnut blight: biological control by transmissible hypovirulence in *Endothia parasitica*. *Science* **189**: 890–891.
- VITALIS, R., and D. COUVET, 2001 Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* **157**: 911–925.
- WAKELEY, J., and N. ALIACAR, 2001 Gene genealogies in a metapopulation. *Genetics* **159**: 893–905.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* **7**: 256–276.

Communicating editor: S. TAVARÉ

