Phylogeography of *Ustilago maydis virus H1* in the USA and Mexico

Peter D. Voth,[1] Linah Mairura,[2] Ben E. Lockhart[3] and Georgiana May[2]

Plant Biological Sciences Graduate Program[1], Department of Ecology, Evolution, and Behaviour[2] and Department of Plant Pathology[3], University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, St Paul, MN 55108, USA

Correspondence
Peter D. Voth
voth0016@umn.edu

Ustilago maydis virus H1 (Umv-H1) is a mycovirus that infects *Ustilago maydis*, a fungal pathogen of maize. As *Zea mays* was domesticated, it carried with it many associated symbionts, such that the subsequent range expansion and cultivation of maize should have affected maize symbionts' evolutionary history dramatically. Because transmission of Umv-H1 takes place only through cytoplasmic fusion during mating of *U. maydis* individuals, the population dynamics of *U. maydis* and maize are expected to affect the population structure of the viral symbiont strongly. Here, the impact of changes in the evolutionary history of *U. maydis* on that of Umv-H1 was investigated. The high mutation rate of this virus allows inferences to be made about the evolution and divergence of Umv-H1 lineages as a result of the recent changes in *U. maydis* geographical and genetic structure. The phylogeographical history and genetic structure of Umv-H1 populations in the USA and Mexico were determined by using analyses of viral nucleotide sequence variation. Infection and recombination frequencies, genetic diversity and rates of neutral evolution were also assessed, to make inferences regarding evolutionary processes underlying the population genetic structure of ancestral and descendent populations. The results suggest that Mexico represents the ancestral population of Umv-H1, from which the virus has been carried with *U. maydis* populations into the USA. Thus, the population dynamics of one symbiont represent a major evolutionary force on the co-evolutionary dynamics of symbiotic partners.

Received 19 April 2006
Accepted 26 July 2006

INTRODUCTION

Ustilago maydis virus H1 (Umv-H1) is in the family *Totiviridae* and has a 6 kb, monopartite, double-stranded RNA (dsRNA) genome encapsidated in a non-enveloped, 40 nm, isometric particle (Martínez-Espinoza *et al.*, 2002; Wickner *et al.*, 2000). Like other mycoviruses, Umv-H1 is transmitted only during mating involving cytoplasmic fusion of host cells (Martínez-Espinoza *et al.*, 2002). The genome encodes three proteins, organized 5′→3′: a capsid, a putative protease and an RNA-dependent RNA polymerase (RdRp) (Kang *et al.*, 2001). The entire genome is expressed as a single polyprotein, which is then cleaved into these three proteins (Kang *et al.*, 2001). In addition to the genome, toxin-encoding satellite dsRNAs may also be encapsidated (Martínez-Espinoza *et al.*, 2002).

Recent theoretical and empirical results suggest that symbiotic interactions strongly affect the evolutionary trajectory of participating symbionts and are important in structuring communities (Bronstein, 1994; García-Arenal *et al.*, 2001; Schardl *et al.*, 1997; Schmitt & Breinig, 2002; Thompson, 1999; Vives *et al.*, 2002; Werren, 1997). Host population history can have dramatic effects on pathogen evolution, diversification and population structure (Abubakar *et al.*, 2003; García-Arenal *et al.*, 2001; Ohshima *et al.*, 2002; Roossinck *et al.*, 1999; Vives *et al.*, 2002). The agricultural system of maize, *Ustilago maydis* and Umv-H1 is well suited to the study of complex symbioses because the component organisms are experimentally tractable and the evolutionary history and biogeography for teosinte, the ancestor of maize (*Zea mays* subsp. *parviglumis*), and maize (*Z. mays* subsp. *mays*) have been well described (Harlan, 1992; Matsuoka *et al.*, 2002; Smith, 1995). *U. maydis* is only pathogenic on maize and teosinte (Fischer & Shaw, 1953) and Umv-H1 has not been observed in unrelated fungi. Moreover, the high mutation rate that we observe for Umv-H1, like that of other RNA viruses (Abubakar *et al.*, 2003; García-Arenal *et al.*, 2001; Holmes, 2004; Ohshima *et al.*, 2002; Roossinck *et al.*, 1999), and the tightly obligate nature of this plant–fungus–virus interaction allow us to make inferences about the evolution and divergence of Umv-H1 lineages as a result of the recent changes in *U. maydis* geographical and population genetic structure.

In this study, we determined the phylogeographical history and genetic structure of Umv-H1 populations in the USA

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are DQ494564–DQ494673.

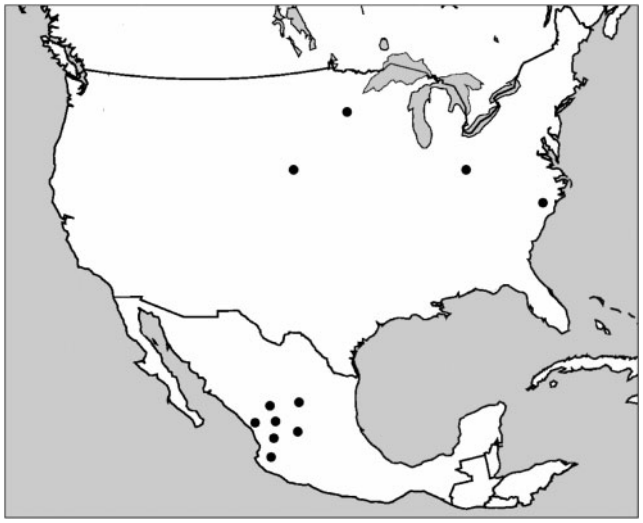

Fig. 1. Locations of Umv-H1 populations.

and Mexico by analysing nucleotide sequences of regions from the capsid and RdRp genes. We assessed infection and recombination frequencies, genetic diversity and rates of neutral evolution to make inferences regarding the evolutionary processes underlying population genetic structure of ancestral and descendent populations of Umv-H1 in the USA and Mexico. Based on the assumption that the historical biogeography of *U. maydis* is the same as the history of maize, we examined the effects of the population structure of the fungal host on its viral symbiont.

METHODS

Sample collection. Collections of *U. maydis* teliospores were taken from a single gall on infected maize and teosinte (*Z. mays* subsp. *parviglumis* and *Z. mays* subsp. *mexicana*) plants from locations in the USA and Mexico. In the USA, collections were made in Minnesota, Ohio, Nebraska and North Carolina. In Mexico, samples of *U. maydis* were collected from seven locations: Cuitzeo, Michoacan; Guachinango, Jalisco; Texcoco, Mexico; Mazatlan, Guerrero; El Estanco, Jalisco; La Cabacera, Jalisco; and Francisco Villa, Durango (Fig. 1; Table 1). Diploid teliospores were germinated to generate haploid sporidia, which grow as a yeast. Sporidial cultures were used to isolate the virus.

Infection frequencies. Infection frequencies were determined by examining total dsRNA extracts from sporidia for the presence of the 6 kb genomic dsRNA of Umv-H1. In other experiments, we demonstrated that the genomic fragment, almost 6 kb in size, sometimes runs as a 4–4·5 kb fragment against DNA size standards (data not shown). The total dsRNA extraction protocol from sporidial cultures was adapted from that described by Seroussi *et al.* (1989). Sporidia were grown for 48 h in potato/glucose broth (Difco) with streptomycin (0·05 mg ml^{-1}). Cultures were centrifuged and sporidia were resuspended in 2× SSC (Sambrook *et al.*, 1989), then boiled for 10 min. The cells were then centrifuged and the pellet was dried under vacuum. The pellet was then ground with 160 µl phenol : chloroform (1 : 1) and silica gel. Sixty microlitres of water was added, followed by vortexing, centrifugation and removal of 50 µl of the aqueous phase, which contained the dsRNA. After electrophoresis in 1 % agarose in TAE buffer, the dsRNAs were visualized by staining with ethidium bromide.

Immunocapture (IC) RT-PCR. Anti-Umv-H1 rabbit antibodies were produced by using virions purified from a mixture of Umv-H1 isolates from the USA, Mexico and Uruguay. Sporidia were grown in 1 l potato/glucose broth containing 0·05 mg streptomycin ml^{-1} for 3–4 days. The tissue was pelleted and ground with a mortar and pestle in 30 ml viral extraction buffer [500 mM sodium/potassium phosphate (pH 7·4) containing 1 M urea, 5 % (w/v) PVP and 0·5 % 2-mercaptoethanol], sand and carborundum. The fungal homogenate was filtered, 2 ml 30 % Triton X-100 was added and the mixture was stirred for 30 min. The filtrate was centrifuged at 20 000 ***g*** (maximum) for 10 min and the pellet was discarded. The supernatant was layered over 5 ml 40 % sucrose in 100 mM phosphate buffer, pH 7·4, and ultracentrifuged in a Beckman 50-2Ti rotor at

Table 1. Locations of *U. maydis* collections, abbreviations, infection frequencies of *U. maydis* by Umv-H1 host plant and number of Umv-H1 sequences obtained from *U. maydis* populations

Location	Abbreviation	Infection frequency of *U. maydis* (%)	Host plant	No. Umv-H1 sequences
USA:				
Owatonna, Minnesota	MN	16·5	*Z. mays* subsp. *mays*	8
Lincoln, Nebraska	NB	19·1	*Z. mays* subsp. *mays*	7
Raleigh, North Carolina	NC	55·6	*Z. mays* subsp. *mays*	6
Wooster, Ohio	OH	45·5	*Z. mays* subsp. *mays*	7
Mexico:				
Cuitzeo, Michoacan	CM	100·0	*Z. mays* subsp. *mays*	3
	CT		*Z. mays* subsp. *mexicana*	5
El Estanco, Jalisco	EE	100·0	*Z. mays* subsp. *mays*	4
Guachinango, Jalisco	GM	100·0	*Z. mays* subsp. *mays*	2
	GT		*Z. mays* subsp. *parviglumis*	2
Texcoco, Mexico	INI	100·0	*Z. mays* subsp. *mays*	4
Mazatlan, Guerrero	M	100·0	*Z. mays* subsp. *mays*	4
La Cabacera, Jalisco	LC	100·0	*Z. mays* subsp. *mays*	2
Francisco Villa, Durango	F	100·0	*Z. mays* subsp. *mays*	1

148 000 ***g*** (max) for 90 min to isolate the Umv-H1 virions. The supernatant was discarded and the pelleted virions were resuspended in PBS.

A rabbit was initially immunized by intravenous and subcutaneous injection of purified virions emulsified in Titre Max Gold adjuvant (Sigma). When the antibody titre in the rabbit reached its maximum, the animal was bled out. IgG was prepared from whole serum by ammonium sulfate and DEAE/cellulose fractionation. IgG was eluted in PBS and adjusted to a final concentration of 1 mg ml^{-1}.

Using the anti-Umv-H1 rabbit antibodies, an IC RT-PCR technique was developed. PCR tubes were initially coated with purified IgG (0·01 mg ml^{-1}) diluted in 60 mM carbonate buffer (pH 9·6). Then, a crude extract of ground fungal tissue was added. The Umv-H1 virions were trapped by the antibodies and provided the template for subsequent RT-PCR. Primers were designed to amplify regions of the capsid and RdRp genes on the viral genome. The capsid primers were, 5′→3′, TGGCGACCTGGGCTAAAG and GCGTCAAAGTAGTA-CTGCCA, positioned 849 and 1423 bp into the genome, respectively. Primers for the RdRp region were, 5′→3′, TACCCTGACTT-CAATTCAATGC and CCATGATCCTTAGGTACTCGTG, positioned at 4574 and 5061 bp, respectively.

Amplification of these gene regions was carried out by using a one-step RT-PCR kit [Promega Access RT-PCR: Tf1 DNA polymerase, avian myeloblastosis virus (AMV) reverse transcriptase, 5× Tf1/AMV buffer, 40 μM dNTPs and nuclease-free water]. The RT-PCR mix was made without the enzymes and was added to the PCR tubes. The template dsRNA was denatured at 100 °C for 90 s, quenched in an ice bath (Attoui *et al.*, 2000) and the enzymes were added. The RT-PCR conditions used to amplify the capsid region were: (i) 48 °C for 1 h; (ii) 94 °C for 2 min; (iii) 94 °C for 30 s; (iv) 61 °C for 30 s; (v) 68 °C for 1 min, 68 °C for 10 min (steps iii–v were repeated for 50 cycles). The PCR conditions to amplify the RdRp region were: (i) 48 °C for 1 h; (ii) 94 °C for 2 min; (iii) 94 °C for 30 s; (iv) 56 °C for 30 s; (v) 68 °C for 1 min; (vi) 68 °C for 10 min (steps iii–v were repeated for 50 cycles). Some Umv-H1 isolates required lowering of the annealing temperature to 50 °C.

DNA sequencing. RT-PCR products were fractionated on 1 % agarose gels in TAE buffer and appropriately sized fragments were purified by using a Qiagen gel extraction kit. Subsequently, DNA sequence was obtained by using an ABI 3100 genetic analyser at the Advanced Genetic Analysis Center at the University of Minnesota. Sequences were edited manually by using SEQUENCHER (Gene Codes Corp.).

Phylogenetic analyses. Edited sequences were aligned by using CLUSTAL_X v. 1.82 (Thompson *et al.*, 1997), using default settings, and verified manually by using MacClade 4.0 (Maddison & Maddison, 2000). Three sequences (CM91, GT42 and MM321) were removed because they demonstrated recombination events (as described below) and thus would introduce unnecessary homoplasy in phylogenetic analyses. Phylogenetic analyses were performed by using neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML) algorithms implemented in PAUP* v. 4.0b10 (Swofford, 1999) on DNA sequence data from the capsid gene, the RdRp gene and the genes combined. The substitution model used for ML analyses was TrNef+I+Γ (Tamura & Nei, 1993), as determined from MODELTEST v3 (Posada & Crandall, 1998). To generate trees based on amino acid sequence, DNA sequence alignments were translated to amino acid sequence by using MacClade 4.0 (Maddison & Maddison, 2000).

Estimates of genetic diversity. Estimates of Wright's *F* statistic (Wright, 1921, 1931) were performed by using ARLEQUIN v. 2.0 (Schneider *et al.*, 2000) based on DNA sequence data from both gene sequences (capsid and RdRp). Sequences were grouped according to subpopulations and those with fewer than four individuals were omitted from the analyses. Pairwise nucleotide diversity with Jukes–Cantor corrections (Jukes & Cantor, 1969) was determined for phylogenetic clades of Umv-H1 in the USA and Mexico by using MEGA2 (Kumar *et al.*, 2001).

Substitution rates and divergence dates. Synonymous and non-synonymous substitutions were analysed by using MEGA2 (Kumar *et al.*, 2001), implementing the Pamilo–Bianchi–Li method (Li, 1993; Pamilo & Bianchi, 1993) to correct for multiple substitutions at a single site. The number of synonymous substitutions per synonymous site (dS) was calculated for each clade representing viral sequences from the USA. Synonymous substitution rates were calculated as synonymous substitutions per site per year and calibrated by using a date of 2000 years before present (ybp) for all Umv-H1 sequences from the USA represented in the tree. The rates of evolution for the USA clades were averaged and the standard deviation (SD) of the three evolutionary rates was determined. This substitution rate and SD were used to determine divergence dates and SD for each clade. The SD of dS for each clade was divided by the mean synonymous substitution rate SD to determine the divergence date error in ybp.

Recombination analyses. Analysis of recombination between Umv-H1 genomes was carried out by using the program GENECONV (Sawyer, 1999). Following evaluation for recombination, the dataset was analysed by using the SplitsTree program (Huson, 1998) to display visually the results from GENECONV (Sawyer, 1999) and to estimate the time of recombination events.

RESULTS

Frequency of infection by Umv-H1 in *U. maydis*

Total dsRNA extracts from haploid *U. maydis* cultures were assessed for the presence of Umv-H1 as indicated by the presence of a 6 kb genomic dsRNA (see Methods). We found that 66 of 232 *U. maydis* individuals surveyed harboured Umv-H1 (34·2 % infection) in the USA populations and all of 44 *U. maydis* individuals surveyed harboured Umv-H1 (100 % infection) in the Mexican populations. The infection frequencies for subpopulations within the USA varied greatly, ranging from 16·5 % (Minnesota) to 55·5 % (North Carolina) (Table 2).

Phylogeographical analysis of Umv-H1 populations

The suitability of combining capsid and RdRp DNA sequence datasets for phylogenetic analysis was tested by conducting NJ, MP and ML analyses of the two genes separately and comparing the resulting trees. None of the MP, ML or NJ trees for the two genes yielded significantly different topologies based on the Kishino–Hasegawa test ($P<0·05$) (Kishino & Hasegawa, 1989; Swofford, 1999) implemented in PAUP* v. 4.0b10. Similarity of trees resulting from the two genes' datasets was further tested by using the Templeton test implemented in PAUP* v. 4.0b10, with the result that the null hypothesis could not be rejected ($P<0·05$), demonstrating no significant differences between the NJ, MP and ML trees. MP and NJ phylogenetic analyses performed on the predicted

Table 2. Pairwise F_{st} values between Umv-H1 populations in the USA and Mexico

Values above the diagonal are distances in km and those below the diagonal are F_{st} values.

	MN	NB	NC	OH	Cuitzeo	El Estanco	Guachinango	Inifap	Mazatlan
MN		468	1542	996	2768	2866	2798	2777	3013
NB	0·309		1676	1253	2341	2419	2347	2370	2604
NC	0·005	0·501		624	2810	3049	3025	2685	2892
OH	0·136	0·537	0·400		2940	3131	3088	2863	3089
Cuitzeo	0·418	0·596	0·542	0·423		309	357	241	340
El Estanco	0·290	0·538	0·439	0·359	0·148		96	545	564
Guachinango	0·269	0·523	0·420	0·337	0·114	−0·070		598	644
Inifap	0·306	0·587	0·488	0·374	0·010	0·051	0·001		236
Mazatlan	0·205	0·460	0·343	0·342	0·425	0·132	0·127	0·295	

amino acid sequence alignments of the capsid and RdRp, separately, also did not yield trees that were significantly different based on the Kishino–Hasegawa test ($P<0·05$) (Kishino & Hasegawa, 1989). ML analysis of the combined nucleotide sequences gave the unrooted tree shown in Fig. 2.

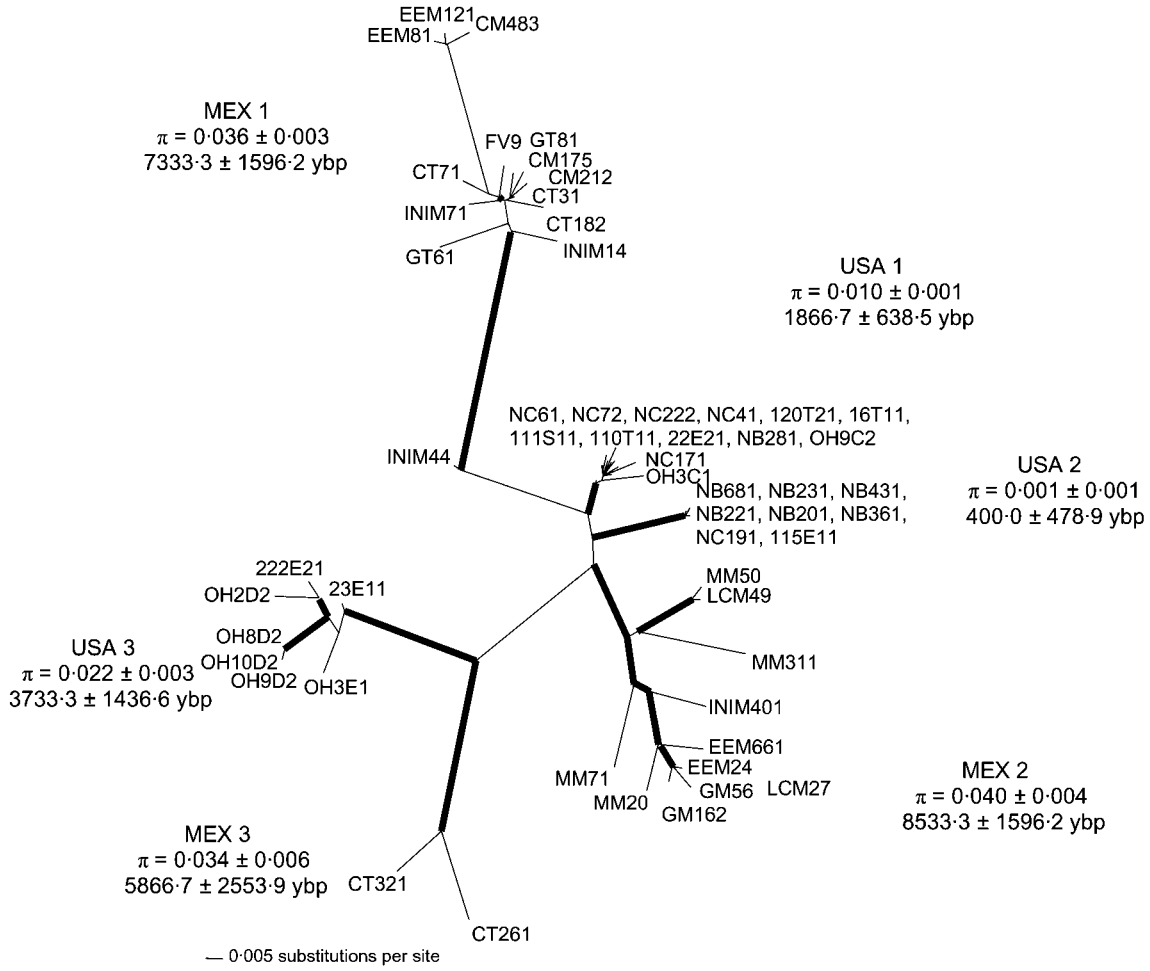

Fig. 2. Unrooted ML phylogram of 55 Umv-H1 sequences from the USA and Mexico based on the combined nucleotide sequences from regions of the capsid and RdRp genes. The individual clades are labelled and show values of genetic diversity (π) and dates of divergence. Branches in bold represent bootstrap values >70 %.

Analysis of DNA and protein sequences yielded phylogenetic trees with six distinct clades (Fig. 2). Parsimony bootstrap (1000 replicates) analyses determined >90 % support for all nodes supporting these six major clades. Three individual clades represented all sequences from subpopulations from the USA and three clades represented all sequences from Mexican subpopulations. Within the six major clades, closely related Umv-H1 sequences had often been collected from the same localities, i.e. most Nebraska sequences clustered together. However, we also observed 'mixed' clades made up of sequences from Umv-H1 collected from different locations; for example, clade USA 1 includes sequences collected from Minnesota, North Carolina, Nebraska and Ohio. Similarly, two of the three Mexican clades, MEX 1 and MEX 2, contained closely related sequences from different geographical locations. To quantify the genetic diversity of each clade, pairwise nucleotide diversity (π) was determined for capsid and RdRp DNA sequence data by using MEGA2 (Kumar *et al.*, 2001). We observed that clades containing sequences from Mexican Umv-H1 collections exhibited greater nucleotide diversity (as π) and longer branch lengths than did most clades with Umv-H1 sequences from the USA (Fig. 2).

To evaluate critically the level of geographical structure in USA and Mexican Umv-H1 populations, we used the method of Slatkin & Maddison (1989) to estimate levels of migration at different spatial scales. In this approach, geographical location of collections (i.e. Mexico or the USA) was used as a character and mapped onto the phylogeny. The lowest (most parsimonious) number of steps required to explain the observed character-state changes (changes in geographical location) is an estimate of migration and compared through simulations with the number of steps required under randomized character-state assignments (i.e. no geographical structure in populations or panmixis). In the analysis using USA and Mexico as character states, only three migration steps were required to explain empirical observations and this result was significantly different from that expected under panmixis (1000 randomized trees; *t*-test, $P<0{\cdot}001$). When this analysis was performed on the populations from within the USA or Mexico separately, with individual subpopulation locations as character states, there was no significant difference between the number of migrational steps required to explain the distribution of subpopulation location and that expected under panmixis ($P<0{\cdot}05$).

Geographical structure of genetic diversity

Pairwise nucleotide diversity was determined for subpopulations of Umv-H1 in the USA and Mexico by using MEGA2 (Kumar *et al.*, 2001). The mean value for π was $0{\cdot}058\pm0{\cdot}005$ across all Umv-H1 collections, $0{\cdot}044\pm0{\cdot}005$ for the USA collections alone and $0{\cdot}063\pm0{\cdot}006$ for Mexican collections alone. The π values for the different subpopulations within the USA ranged from $0{\cdot}013\pm0{\cdot}002$ (NB) to $0{\cdot}053\pm0{\cdot}005$ (MN). For subpopulations in Mexico, π values ranged from $0{\cdot}049\pm0{\cdot}005$ (INI) to $0{\cdot}099\pm0{\cdot}009$ (EE). Based on these data, we observe not only that subpopulations in Mexico exhibit higher genetic diversities than most subpopulations from the USA, but also that the Mexican Umv-H1 population as a whole exhibits somewhat higher nucleotide diversity than does the USA Umv-H1 population.

Overall, pairwise comparisons between all populations in the USA and those in Mexico yielded a mean pairwise F_{st} value of 0·264 (Table 2) and demonstrated a significant level of differentiation between populations of Umv-H1 in these two large geographical areas. Pairwise values of F_{st} between subpopulations within the USA ranged from 0·005 (NC vs MN) to 0·537 (OH vs NB). Similarly, in Mexico, F_{st} values between subpopulations ranged from $-0{\cdot}070$ (G vs EE) to 0·425 (C vs M). Pairwise F_{st} values for comparisons between subpopulations in the USA and Mexico were generally high and ranged from 0·205 (MN vs M) to 0·596 (NB vs C).

Populations at equilibrium between migration, drift and mutation will demonstrate increased levels of population differentiation as geographical distance between populations increases, or isolation by distance (IBD). To test for IBD, all combinations of pairwise genetic distances [$F_{st}/(1-F_{st})$] were plotted against the geographical distance between each of the subpopulations within the USA and Mexico. We assessed IBD for populations between the USA and Mexico and observed no correlation between genetic and geographical distance, i.e. no IBD (Fig. 3a). There does not appear to be a high level of Umv-H1 migration between populations in the USA and Mexico. Within the USA and Mexico, we also observed a wide range of F_{st} values corresponding to all ranges of geographical distance and, similar to the larger-scale results, this analysis demonstrated no IBD (Fig. 3b). Our results suggest that subpopulations within and between the USA and Mexico are not at equilibrium of drift and migration, perhaps a surprising result given the apparent age of Mexican populations or the high migration rates that might be expected of an agricultural pathogen such as *U. maydis*.

Substitution rates and divergence dates

We calibrated the mean rate of synonymous substitutions by setting 2000 ybp as the earliest date of establishment of USA populations. *U. maydis* populations could not have been present in the USA prior to the introduction of maize at 2000 ybp because no known alternative hosts exist outside Mexico and Central America (Harlan, 1992; Matsuoka *et al.*, 2002; Smith, 1995). Substitution rates were not calibrated directly for populations in Mexico because *U. maydis* infects teosintes, the progenitors of maize, which existed in Mexico long before the domestication of maize (8000–10 000 ybp) (Harlan, 1992; Matsuoka *et al.*, 2002; Smith, 1995). The resulting estimated rate of evolution was $7{\cdot}50\times10^{-6}\pm6{\cdot}26\times10^{-6}$ synonymous substitutions per synonymous site per year. The dS values for each clade used to calculate divergence dates are: USA 1, $0{\cdot}014\pm0{\cdot}004$; USA 2, $0{\cdot}003\pm0{\cdot}003$; USA 3, $0{\cdot}028\pm0{\cdot}009$; MEX 1, $0{\cdot}055\pm0{\cdot}010$; MEX 2, $0{\cdot}064\pm0{\cdot}010$; MEX 3,

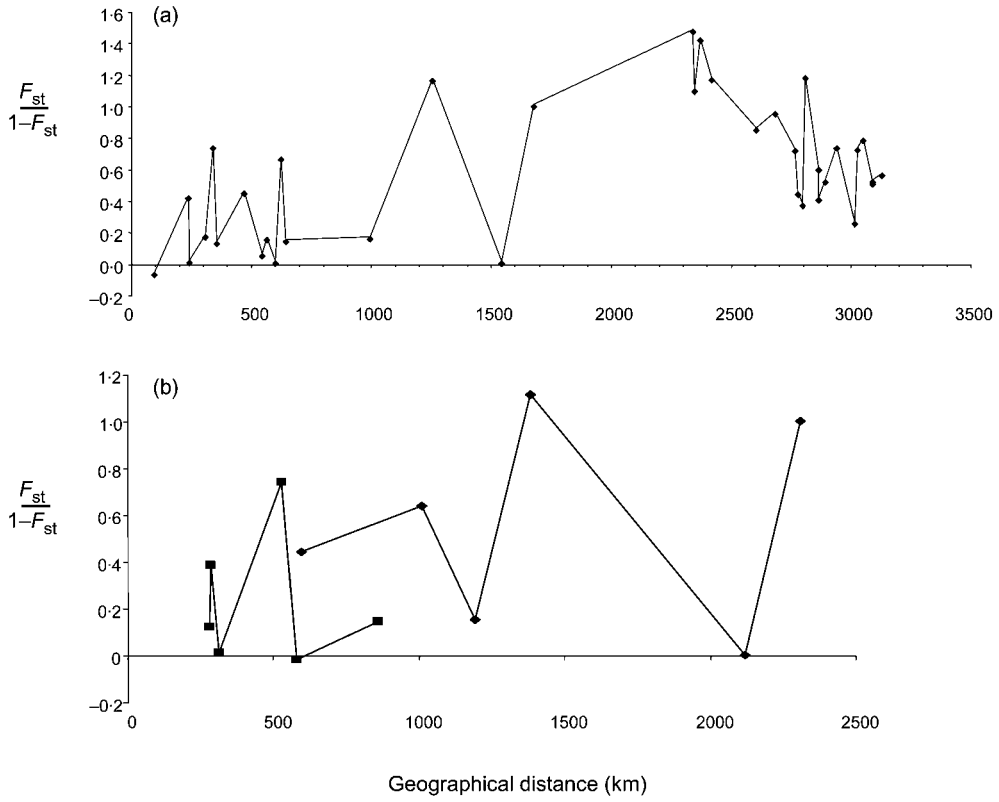

Fig. 3. Pairwise genetic distance versus geographical distance between all populations (a) and between subpopulations within the USA (◆) and Mexico (■) (b).

$0{\cdot}044 \pm 0{\cdot}016$. The divergence dates and errors for each clade are shown in Fig. 2.

Selection analysis

We calculated the number of non-synonymous substitutions per non-synonymous site (dN) and number of synonymous substitution per synonymous site (dS) to obtain the dN/dS ratio for capsid and RdRp sequences across all populations. We found that the mean dN/dS ratio across both genes is 0·033; that for capsid sequence is 0·046 and that for the RdRp sequence is 0·017. These results show that, despite the large number of polymorphisms between all of the Umv-H1 sequences analysed, most of the substitutions were silent, encoding the same amino acid, and very few replacement substitutions were observed, results demonstrating strong purifying selection.

Recombination analysis

We evaluated incidence of recombination events within and between sequences from the capsid (5′) and RdRp (3′) regions of the genome by using GENECONV (mismatch = 0, default settings) (Sawyer, 1999). Recombination tracts were identified in seven Umv-H1 sequences. However, a number of these seven individual Umv-H1 genomes shared the same recombination tract and further analysis identified only four independent, ancestral recombination events, all of which involved sequences from Mexican collections. One ancestral recombination event occurred within the capsid sequence, one within the RdRp sequence and the two other recombination tracts that we identified must have had a breakpoint within the non-sequenced region between the two genes. By using the more conservative estimate provided by Bonferroni-corrected P values ($P > 0{\cdot}05$), the total number of independent recombination events detected was three, not including the tract within the RdRp discussed above.

These recombination results were corroborated by analyses using SplitsTree (Huson, 1998), which resolves ambiguities in a phylogenetic tree by forming loops linking potentially recombinant sequences. The SplitsTree results demonstrated loops connecting the same recombinant Umv-H1 sequences as were identified in GENECONV analyses (data not shown). Altogether, the GENECONV and SplitsTree analyses demonstrated that recombination does occur, but at a very low frequency.

DISCUSSION

Our results suggest strongly that Umv-H1 populations in *U. maydis* originated in Mexico, dating back to at least the time

that maize was domesticated from teosinte (approx. 8000–10 000 ybp). Maize was brought only relatively recently into regions of the present-day USA (1000–2000 ybp; Harlan, 1992; Matsuoka *et al.*, 2002; Smith, 1995) and an assemblage of microbes, including *U. maydis* and Umv-H1, travelled with their host. Here, we describe new aspects of the biology of Umv-H1 as its rates of substitution and recombination, and then describe an evolutionary history for Umv-H1. That history largely parallels that of its fungal host and host plant, except that Umv-H1 has lost ground in the USA populations of *U. maydis*, existing there at much lower frequencies than in Mexico.

The compactness of the viral genome, which encodes just three essential proteins, and the employment of error-prone reverse transcriptase represent a juxtaposition of conflicting evolutionary pressures on the viral genome. The low dN/dS ratios that we estimated for the Umv-H1 genome are similar to those of other viruses (Abubakar *et al.*, 2003; Kurath *et al.*, 2003; Moury *et al.*, 2002; Powers *et al.*, 2000) and suggest intense purifying selection against amino acid replacements (Li & Roossinck, 2004). At the same time, we estimated high rates of substitution and found high levels of nucleotide diversity. Whilst elimination of a deleterious substitution pulls with it all associated substitutions due to linkage (Li & Roossinck, 2004) and thus limits standing variation, the rate of evolution for this dsRNA virus is high enough to allow study of the temporal and spatial dynamics of populations (Holmes, 2004).

Our results demonstrate for the first time recombination in Umv-H1 populations, although very infrequent, which will lessen the impact of selective sweeps described above on standing variation in the virus. Biologically, the molecular recombination observed here must result from co-infection of *U. maydis* by more than one Umv-H1 genome. Instead, previous studies found that progeny resulting from co-infection of *U. maydis* carried only one of two co-infecting viral genotypes, a process attributed to an 'exclusion phenomenon' (Day, 1981; Koltin & Kandel, 1978). The very low rate at which we observe recombination implies that the evolutionary trajectory of Umv-H1 populations is affected more strongly by mutation and selective sweeps than by the effect of recombination and selection on individual substitutions, but that, infrequently, new genotypes arise by recombination.

The results of population genetic and phylogenetic analyses demonstrate significant geographical structuring between viral populations in the USA and Mexico, a pattern that could only have been set up over the last 2000 years. Our data support a Mexican origin of Umv-H1 populations, where we observed higher levels of infection and greater levels of genetic diversity than in the USA populations. Subsequent to the introduction of maize into the USA and establishment of *U. maydis* and Umv-H1 populations there, apparently little migration between these major regions has occurred. These results, together with the historical records for maize cultivation, suggest strongly that the patterns of biogeographical diversity that we observe in Umv-H1 are the direct result of historical founder events as *U. maydis* and Umv-H1 followed maize from Mexico into the USA. Our estimations of very low ongoing migration and relatively few founder events, coupled with observations of strong differentiation between major geographical regions, suggest strongly that the extant populations of Umv-H1 that we observe in the USA today are the result of ancestral founding events, largely unaffected by current trade in maize.

Nonetheless, regional trade in maize and even *U. maydis* has had an apparent effect. Since domestication approximately 8000–10 000 years ago, maize has been traded extensively among indigenous people inhabiting the Americas. *U. maydis* is corn smut or huitlacoche, and has long been considered a delicacy in Mexico and thus received much attention (Ruiz-Herrera & Martínez-Espinoza, 1998). The direct trade in huitlacoche may provide an increased rate of local migration of the fungal host and its viral symbiont. Accordingly, viral sequences recovered from Mexico cluster into clades composed of sequences recovered from disparate geographical locations. Whilst it might be expected that the movement of Umv-H1 with *U. maydis* may have prevented the establishment of geographical substructuring within the Mexican population or between teosinte and maize hosts, it is surprising to find little evidence of IBD. Because IBD is a signature of populations at equilibrium between migration and drift, we infer that Mexican populations and those in the USA are not at equilibrium and look to further research to determine underlying mechanisms.

Results of the phylogenetic analyses show that the diversity of Umv-H1 sequences is divided neatly into three well-supported clades of Mexican isolates and three of USA isolates. Whilst we might have expected the USA clades to have been derived more recently from within the Mexican clades, rates of substitution are high enough to obscure basal relationships. Currently, we are sampling Mexican *U. maydis* collections and Umv-H1 sequences further to improve resolution at the deeper nodes and obtain better inferences for source populations in Mexico. In agreement with a Mexican origin, the Mexican clades exhibit higher levels of genetic diversity and older dates of origin compared with the three clades representing viral populations from the USA. Most interestingly, the dates for Mexican clades correspond well with the domestication of maize in Mexico from teosintes and suggest that these populations date to domestication, rather than to older populations inhabiting teosintes in Mexico.

The results of population and phylogenetic analyses demonstrate, as expected, a much more recent history for USA Umv-H1 populations. Several of the USA clades exhibit star phylogenies, signatures of recent establishment and expansion of clonal lineages (Roossinck *et al.*, 1999) and, because the host fungus, *U. maydis*, is an outcrossing, obligately sexual organism (Barnes *et al.*, 2004; Zambino *et al.*, 1997), we infer that clonal patterns result from the recent expansion of a cytoplasmically inherited virus. In fact,

divergence times of the viral lineages in the USA are remarkably concordant with that which we might expect from anthropological data for the very recent expansion of maize cultivation into the present-day USA. We estimated establishment of the most geographically diverse clade, USA 1, at ~1900 ybp, whereas the USA 2 clade, mostly comprising sequences from Nebraska, is dated at only ~400 ybp. Discordant with dates of movement of maize, the USA 3 clade is dated at almost 4000 ybp, well before maize made it out of Mexico. Collections of *U. maydis* comprising USA 3 Umv-H1 isolates were made from an agricultural experiment station involved historically in the development of maize resistance to smut, and perhaps such a distinct date is the result of use of *U. maydis* collections from other regions of the world or selection. As in Mexico, lack of IBD suggests that populations in the USA are not at equilibrium of drift and migration. Indeed, evidence for recent, infrequent and long-distance migration events was found. For example, a North Carolina isolate was very similar to many Nebraska isolates, suggesting long-distance migration from the recently founded Nebraska location to the south-eastern USA. With the paucity of migration between Mexico and the USA, recently established Umv-H1 populations in the USA must have originated from a few, original founding events and subsequent migration from other populations within the USA.

The results for geographical structure and the dates of origin for viral populations at the time of maize domestication, 8000–10 000 ybp, along with similarly high genetic differentiation between geographically distant populations of *U. maydis* (Barnes *et al.*, 2004) highlight the profound impact of host population dynamics on the evolution of associated, interacting symbionts (Abubakar *et al.*, 2003; Bousalem *et al.*, 2000; García-Arenal *et al.*, 2001; Holmes, 2004; Kurath *et al.*, 2003; Powers *et al.*, 2000; Roossinck *et al.*, 1999; Vives *et al.*, 2002). Still, although the population structure of Umv-H1 is linked inherently to the biogeography and migration of *U. maydis* and more deeply to the plant host of the fungal pathogen, the evolutionary relationship of *U. maydis* and Umv-H1 may have undergone a not too subtle shift in the USA descendent populations. There, we found that *U. maydis* is not infected as uniformly by Umv-H1 as the Mexican *U. maydis* populations. The most parsimonious inference is of a widespread loss of Umv-H1 from *U. maydis* in the USA. One hypothesis is that there is a cost to the fungal host of supporting the virus, especially if Umv-H1 carries satellite dsRNAs encoding one of three proteinaceous toxins secreted extracellularly. All three of the secreted toxins kill susceptible *U. maydis* individuals at the time of mating and cell fusion. Alternatively, severe bottlenecks involved in the founding of *U. maydis* populations in the USA could lead to the loss of the virus by drift.

Together, our results suggest strongly that Umv-H1 in *U. maydis* originated in Mexico and that, much more recently, Umv-H1 was introduced into the USA region along with *U. maydis* and maize. Umv-H1 tracks its fungal host, but not uniformly, as a substantial fraction of *U. maydis* in the USA has lost the viral symbiont. The population dynamics of one symbiont represent a major force in the evolution of population structure of the other symbiont, which affects the evolution of such complex symbioses as a whole.

ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation (DMS-0083468 to G. M. and DEB-0508264 to G. M. and P. D. V.), the University of Minnesota Center for Community Genetics and the University of Minnesota Plant Molecular Genetics Institute. We thank Jean Pan, Andrew Munkacsi and Brett Couch for their valuable input and insight in reviewing this manuscript and are grateful to Andrew Munkacsi, Jesus Sanchez, Nick Lauter and Peter Tiffin for their help with sample collections.

REFERENCES

Abubakar, Z., Ali, F., Pinel, A., Traoré, O., N'Guessan, P., Notteghem, J.-L., Kimmins, F., Konaté, G. & Fargette, D. (2003). Phylogeography of *Rice yellow mottle virus* in Africa. *J Gen Virol* **84**, 733–743.

Attoui, H., Billoir, F., Cantaloube, J. F., Biagini, P., de Micco, P. & de Lamballerie, X. (2000). Strategies for the sequence determination of viral dsRNA genomes. *J Virol Methods* **89**, 147–158.

Barnes, C. W., Szabo, L. J., May, G. & Groth, J. V. (2004). Inbreeding levels of two *Ustilago maydis* populations. *Mycologia* **96**, 1236–1244.

Bousalem, M., Douzery, E. J. P. & Fargette, D. (2000). High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution. *J Gen Virol* **81**, 243–255.

Bronstein, J. L. (1994). Conditional outcomes in mutualistic interactions. *Trends Ecol Evol* **9**, 214–217.

Day, P. R. (1981). Fungal virus populations in corn smut from Connecticut. *Mycologia* **73**, 379–391.

Fischer, G. W. & Shaw, C. G. (1953). A proposed species concept in the smut fungi, with application to North American species. *Phytopathology* **43**, 181–188.

García-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157–186.

Harlan, J. R. (1992). *Crops and Man*, 2nd edn. Madison, WI: American Society of Agronomy.

Holmes, E. C. (2004). The phylogeography of human viruses. *Mol Ecol* **13**, 745–756.

Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kang, J.-G., Wu, J.-C., Bruenn, J. A. & Park, C.-M. (2001). The H1 double-stranded RNA genome of *Ustilago maydis virus-H1* encodes a polyprotein that contains structural motifs for capsid polypeptide, papain-like protease, and RNA-dependent RNA polymerase. *Virus Res* **76**, 183–189.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA

sequence data, and the branching order in hominoidea. *J Mol Evol* **29**, 170–179.

Koltin, Y. & Kandel, J. S. (1978). Killer phenomenon in *Ustilago maydis*: the organization of the viral genome. *Genetics* **88**, 267–276.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetic analysis software. *Bioinformatics* **17**, 1244–1245.

Kurath, G., Garver, K. A., Troyer, R. M., Emmenegger, E. J., Einer-Jensen, K. & Anderson, E. D. (2003). Phylogeography of infectious haematopoietic necrosis virus in North America. *J Gen Virol* **84**, 803–814.

Li, W.-H. (1993). Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* **36**, 96–99.

Li, H. & Roossinck, M. J. (2004). Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol* **78**, 10582–10587.

Maddison, D. & Maddison, W. (2000). MacClade 4: analysis of phylogeny and character evolution. Sunderland, MA: Sinauer Associates.

Martínez-Espinoza, A. D., García-Pedrajas, M. D. & Gold, S. E. (2002). The Ustilaginales as plant pests and model systems. *Fungal Genet Biol* **35**, 1–20.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J. G., Buckler, E. & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci U S A* **99**, 6080–6084.

Moury, B., Morel, C., Johansen, E. & Jacquemond, M. (2002). Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. *J Gen Virol* **83**, 2563–2573.

Ohshima, K., Yamaguchi, Y., Hirota, R. & 10 other authors (2002). Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. *J Gen Virol* **83**, 1511–1521.

Pamilo, P. & Bianchi, N. O. (1993). Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Mol Biol Evol* **10**, 271–281.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Powers, A. M., Brault, A. C., Tesh, R. B. & Weaver, S. C. (2000). Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J Gen Virol* **81**, 471–479.

Roossinck, M. J., Zhang, L. & Hellwald, K.-H. (1999). Rearrangements in the 5′ nontranslated region and phylogenetic analyses of cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752–6758.

Ruiz-Herrera, J. & Martínez-Espinoza, A. D. (1998). The fungus *Ustilago maydis*, from the aztec cuisine to the research laboratory. *Int Microbiol* **1**, 149–158.

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

Sawyer, S. A. (1999). GENECONV: a computer package for the statistical detection of gene conversion. St Louis, MO: Department of Mathematics, Washington University in St Louis.

Schardl, C. L., Leuchtmann, A., Chung, K.-R., Penny, D. & Siegel, M. R. (1997). Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts. *Mol Biol Evol* **14**, 133–143.

Schmitt, M. J. & Breinig, F. (2002). The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol Rev* **26**, 257–276.

Schneider, S., Roessli, D. & Excoffier, L. (2000). ARLEQUIN: a software for population genetics data analysis, ver. 2.000. Geneva: Genetics and Biometry Laboratory, Dept of Anthropology, University of Geneva.

Seroussi, E., Peery, T., Ginzberg, I. & Koltin, Y. (1989). Detection of killer-independent dsRNA plasmids in *Ustilago maydis* by a simple and rapid method of extraction of dsRNA. *Plasmid* **21**, 216–225.

Slatkin, M. & Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**, 603–613.

Smith, B. D. (1995). *The Emergence of Agriculture*. New York: Scientific American Library.

Swofford, D. L. (1999). PAUP*: phylogenetic analysis using parsimony (*and other methods), ver 4.0b10. Sunderland, MA: Sinauer Associates.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–526.

Thompson, J. N. (1999). The evolution of species interactions. *Science* **284**, 2116–2118.

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

Vives, M. C., Rubio, L., Galipienso, L., Navarro, L., Moreno, P. & Guerri, J. (2002). Low genetic variation between isolates of *Citrus leaf blotch virus* from different host species and of different geographical origins. *J Gen Virol* **83**, 2587–2591.

Werren, J. H. (1997). Biology of *Wolbachia*. *Annu Rev Entomol* **42**, 587–609.

Wickner, R. B., Ghabrial, S. A., Bruenn, J. A., Buck, K., Patterson, J. L., Stuart, K. D. & Wang, C. C. (2000). Family *Totiviridae*. In *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 491–501. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Wright, S. (1921). Systems of mating. *Genetics* **6**, 111–178.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Zambino, P., Groth, J. V., Lukens, L., Garton, J. R. & May, G. (1997). Variation at the *b* mating type locus of *Ustilago maydis*. *Phytopathology* **87**, 1233–1239.