Phylodynamic evidence of the migration of turnip mosaic potyvirus from Europe to Australia and New Zealand

Ryosuke Yasaka[1,2], Kiho Ohba[1], Mark W. Schwinghamer[3], John Fletcher[4], Francisco M. Ochoa Corona[5,*], John E. Thomas[6], Simon Y. W. Ho[7], Adrian J. Gibbs[8] and Kazusato Ohshima[1,2,†]

[1] Laboratory of Plant Virology, Department of Applied Biological Sciences, Faculty of Agriculture, Saga University, 1-banchi, Honjo-machi, Saga 840-8502, Japan

[2] The United Graduate School of Agricultural Sciences, Kagoshima University, 1-21-24, Kagoshima 890-0065, Japan

[3] New South Wales Department of Primary Industries, Tamworth Agricultural Institute, 4 Marsden Park Road, Tamworth, NSW 2340, Australia

[4] The New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch, New Zealand

[5] National Institute for Microbial Forensics & Food and Agricultural Biosecurity, Department of Entomology & Plant Pathology, Oklahoma State University,127 Noble Research Center, Stillwater, OK 74078, USA

[6] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Ecosciences Precinct, GPO Box 267, Brisbane, QLD 4001, Australia

[7] School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia

[8] Emeritus Faculty, Australian National University, ACT 2601, Australia

---

*Past Address; Investigation & Diagnostic Centre (IDC), Plant Health & Environment Laboratory (PHEL), Biosecurity New Zealand, Ministry of Agriculture & Forestry, 231 Morrin Road, St Johns, Auckland 1140, New Zealand

†Corresponding author. Tel: +81-952-28-8730, Fax: +81-952-28-8709, E-mail address: ohshimak@cc.saga-u.ac.jp (K. Ohshima)

Key words: turnip mosaic potyvirus; evolution; time; migration; Australia; New Zealand

Short title: Migration of turnip mosaic virus from Europe to Australia and New Zealand

The word count; summary 247 (max=250), main text 5,091 (text)+406 (figure legends) (max=5,500)

Total numbers of figures and tables; four and two (max=8).

Two supplementary tables and seven supplementary figures are available with the online version of this paper.

ABSTRACT

Thirty-two isolates of turnip mosaic virus (TuMV) were collected mostly from *Brassicaceae* plants in Australia and in New Zealand during 1994-2011. Host reaction studies showed that most of the isolates belonged to *Brassica* (B) host-infecting type. We mostly performed sequence-based phylogenetic and population genetic analyses of the complete genomic sequences and of three non-recombinogenic regions of those sequences (protein encoding regions of the partial helper-component proteinase protein, protein 3 and nuclear inclusion b protein). The substitution rates, divergence times, and phylogeographic patterns of the virus populations were estimated. Six inter- and five intralineage recombination type patterns were found in the genomes of the Australian and New Zealand isolates, and all were novel. Only one recombination type pattern has been found in both countries. Australian and New Zealand populations were genetically different, and were different from the European and Asian populations. Our Bayesian coalescent analyses, based on a combination of novel and published sequence data from three non-recombinogenic protein-encoding regions, showed that TuMV probably started to migrate from Europe to Australia and New Zealand more than 80 years ago, and distinct populations arose as a result of evolutionary drivers such as recombination. The basal-B2 subpopulation in Australia and New Zealand seems to be older than those of the world-B2 and B3 populations. Ours is the first population genetic study of TuMV in the Antipodes. It shows that the time of migration of TuMV correlates well with the establishment of agriculture and migration of Europeans to these countries.

INTRODUCTION

Studies of the genetic structure of populations of plant viruses are important for understanding virus evolution and emergence (García-Arenal *et al.*, 2001; Gibbs *et al.*, 2008; Gibbs & Ohshima, 2010), especially of those viruses that evolve at measurable rates and adapt rapidly to new or resistant hosts (Ohshima *et al.*, 2010). These not only include plant viruses with RNA genomes but also those with single- and double-stranded DNA genomes, such as begomoviruses and mastreviruses in the family *Geminiviridae* (Lefeuvre *et al.*, 2010, Rocha *et al.*, 2013) and cauliflower mosaic virus in the family *Caulimoviridae* (Yasaka *et al.*, 2014). All of these reports have shown that the evolution of virus populations is shaped by founder effects, selection and recombination.

Potyviruses are RNA viruses and are among the most important pathogens of crops. They have spread throughout much of the subtropical and temperate zones of the world (Gibbs & Ohshima, 2010; King *et al.*, 2012). Potato virus Y (Ogawa *et al.*, 2008; Visser *et al.*, 2012), turnip mosaic virus (TuMV) (Nguyen *et al.*, 2013b; Ohshima *et al.*, 2002), soybean mosaic virus (Seo *et al.*, 2009) and zucchini yellow mosaic virus (Lecoq *et al.*, 2009) are important potyviruses with worldwide distributions. Nevertheless, there remains a poor understanding of how and when they dispersed, and of what factors controlled that spread.

TuMV infects a wide range of plant species, most from the family *Brassicaceae* (Walsh & Jenner, 2002). TuMV, like other potyviruses, is transmitted by aphids in the non-persistent manner. TuMV has flexuous filamentous particles 700-750 nm long, each of which contains a single copy of the genome, which is a single-stranded, positive-sense RNA molecule of about 9,833 nucleotides. This is translated into one large polyprotein which hydrolyzes itself into at least 10 proteins (King *et al.*, 2012). Furthermore, an overlapping 'pretty interesting *Potyviridae* ORF' (PIPO) exists in the +2 reading frame within the protein 3 (P3) encoding region (Chung *et al.*, 2008).

Earlier studies have shown that TuMV originated from wild orchids in Europe and then emerged to spread among species of *Brassicaceae* in the Mediterranean region, including Southeast Europe, Asia Minor and mid-Eurasia (Nguyen *et al.*, 2013b; Ohshima *et al.*, 2002; Tomimura *et al.*, 2004). Crops of the *Brassicaceae* that are most commonly cultivated in Europe are *Brassica* species, whereas both *Brassica* and *Raphanus crops* are important in Asia Minor and Asian countries (Nguyen *et al.*, 2013a; Tomimura *et al.*, 2003; Tomitaka & Ohshima, 2006; Tomitaka *et al.*, 2007). TuMV isolates are of five host-infecting types. OM-host type isolates infect some plants of Brassicaceae but not brassica plants, [(B)]-host type isolates infect *Brassica* plants latently and occasionally, but do not infect *Raphanus* plants, [B]-host type isolates infect most *Brassica* species systemically, causing mosaic of the systemically infected leaves, but do not infect *Raphanus* plants. [B(R)]-host type isolates infect most *Brassica* systemically, causing mosaic of systemically infected leaves, and infect *Raphanus* plants latently and occasionally. [BR]-host type isolates infect both *Brassica* and *Raphanus* plants systemically, causing mosaic of the systemically infected leaves. The basal-B cluster of (B) or B-host type isolates was most variable, was paraphyletic to the other lineages, and was isolated from both non- and *Brassicaceae* plants. The world-B cluster is the most variable and widespread cluster; most European isolates did not infect *Raphanus*, whereas Asian isolates infect both *Brassica* and *Raphanus*.

Turnip-mosaic was first reported in Australia and New Zealand in the 1930s (Chamberlain, 1936; Samuel, 1931) and characterized by symptoms, by host range and by sap- and aphid-transmission. More recent reports from Australia (Gibbs *et al.*, 2006; Schwinghamer *et al.*, 2014) and New Zealand (Fletcher *et al.*, 2010; Ochoa Corona *et al.*, 2007) characterized the virus by molecular techniques and showed that two isolates from Australia were closely related to isolates from Europe. Here, we report an in-depth analysis of the populations of TuMV in Australia and New Zealand, mostly from *Brassicaceae* hosts, together with the full genomic sequences of 32 of the isolates. We use data from full and partial genomic sequences for evolutionary analyses, including recombination and phylogenetic analyses, and for the estimation of subpopulation

differentiation, relationships and divergence between their populations and those in Europe and Asia. We make phylodynamic comparisons using the genomic sequences of 229 isolates collected worldwide, and discuss what they reveal about the changes that have occurred during continent-wide evolution and migration of populations. Our analyses provide a preliminary definition of the present geographical structure of TuMV populations in Australia and New Zealand, and indicate that they reflect recent human immigration patterns and the agricultural history of the two countries.

RESULTS

Biological and molecular characteristics

A total of 32 TuMV isolates collected in Australia and New Zealand was examined in this study; 16 from eastern Australia, and one from the North Island and 15 from the South Island of New Zealand (Fig. 1, Table S1). Thus, isolates were collected from those parts of the two countries in which brassica crops are cultivated. One Australian isolate was from *Cicer arietinum*, a legume, and 15 from brassicas. The New Zealand samples included one from a crocus and 15 from brassicas and from the closely related plant, *Nasturtium officinale*. The viruses were found in commercial fields as well as in home gardens.

All the Australian and New Zealand isolates infected *Brassica juncea* cv. Hakarashina and *B. rapa* cv. Hakatasuwari plants, but occasionally infected *B. oleracea* var. capitata cvs. Ryozan 2-go and Shinsei. They did not infect Japanese radish (*Raphanus sativus* cvs. Akimasari, Taibyo-soubutori and Houryou). Therefore, we concluded that the Australian and New Zealand isolates are *Brassica* (B) host infecting type. Interestingly, Australian and New Zealand isolates showed local lesions on the inoculated leaves and then systemic symptoms in *Chenopodium quinoa*, whereas most Asian and European isolates showed local lesions only on the inoculated leaves (data not shown).

We analysed the 32 full genomes sequenced in this study, as well as 197 full genomic TuMV sequences obtained from public DNA sequence databases. The 197 full genomic sequences contain the two published sequences from Australia and New Zealand. The genomes of 29 Australian and New Zealand isolates were 9,798 nucleotides long excluding 5' end 35 nt primer sequences, whereas three New Zealand isolates (NZ403, NZ403B and NZ415) were two nucleotides shorter in the 3' non-coding region (NCR; 207 nt long). The regions encoding the protein 1 (P1), helper-component proteinase protein (HC-Pro), P3, PIPO, 6 kDa 1 protein (6K1), cylindrical inclusion protein (CI), 6 kDa 2 protein (6K2), genome linked viral protein (VPg), nuclear inclusion a-proteinase protein (NIa-Pro), nuclear inclusion b protein (NIb) and coat protein (CP) encoding regions were 1086, 1374, 1065, 177, 156, 1932, 159, 576, 729, 1551 and 864 nucleotides long, respectively. All of the motifs reported for different potyvirus-encoded proteins were found. The new genomic sequences determined in this study have been deposited in DDBJ/EMBL/GenBank databases with Accession Codes AB989628–AB989659.

Genetic recombination in Australia and New Zealand

The genomic sequences of 32 Australian and New Zealand isolates and 197 published sequences were assessed for evidence of recombination. Each of the identified sites was examined individually and a phylogenetic approach was used to verify the parent/donor assignments made using the RDP4 package (Martin *et al.*, 2010). Having examined all sites with an associated *P*-value of $<10^{-6}$ (i.e., the most likely recombination sites), we retained the intralineage recombinants (parents from the same major group lineage) and removed the interlineage recombinants (i.e., those with parents from different major lineages) by treating the identified recombination sites as missing data in subsequent analyses.

Twenty-one unequivocal recombination sites were found in the genomes of 34 Australian and New Zealand isolates (Fig. 2, Table S2). In the Australian population, one isolate (AUST21) was identified to be a non-recombinant of the world-B3 subgroup, whereas one isolate (BRS1) of the basal-B and two isolates (AUST19 and AUST23) of the world-B group were identified to be intralineage recombinants. The other 13

isolates were interlineage recombinants between world-B and basal-B parents. In the New Zealand population, no non- and interlineage recombinants were found. All were intralineage recombinants of basal-B or world-B parents. Fourteen New Zealand isolates were single, double or triple intralineage recombinants of world-B parents. Triple intralineage recombinants of world-B parents (AUST19, AUST23 and NZ402) were present both in Australian and New Zealand populations but were not dominant in either country. Twenty recombination sites, except one at nt 6132 in the VPg encoding region, had not been found in other TuMV populations (Nguyen *et al*., 2013b; Ohshima *et al*., 2007), indicating that the Australian and New Zealand populations were distinct.

Phylogenetic relationships

A phylogenetic network was inferred using Neighbor-Net from the concatenated 5' NCR, main ORF and 3' NCR sequences (Fig. S1). Three isolates (AUST10, AUST13 and BRS1) from Australia, three isolates (NZ403, NZ403B and NZ415) from New Zealand, and many German, Italian and Spanish isolates of Europe recombinants and non-recombinants fell into the 'basal-B and recombinants' group. Furthermore, many Australian and New Zealand isolates with worldwide isolates fell into the 'world-B and recombinants' group and clustered with European isolates. None of Australian and New Zealand isolates (either full genomes or parts of them) grouped with the Orchis, 'basal-BR and recombinants', or 'Asian-BR and recombinants' groups.

Because only one of the Australian and New Zealand isolates was not a recombinant, the relationships of the isolates were investigated using the three regions of the genomic sequences that gave the least evidence of recombination: Region A (nt 1460-3472, numbers corresponding to the positions in original UK 1 genome) covered part of the HC-Pro and P3 regions; Region B (nt 3812-6016) included part of the CI and VPg regions; and Region C (nt 6479-8068) included part of the NIa and NIb regions (see Fig. 2). Trees were calculated using 225, 214 and 226 non- and intralineage recombinant sequences, respectively. The relationships of isolates were investigated by maximum likelihood (ML) implemented in PhyML (Guindon & Gascuel, 2003) (Fig. S2). These partitioned most of the sequences into the same five major consistent genetic groups, as reported previously (Nguyen *et al*. 2013b): Orchis, basal-B, basal-BR, Asian-BR and world-B. The basal-B group further split into basal-B1 and B2 subgroups and the world-B group split into the world-B1, B2 and B3 subgroups.

In the Region A tree (Fig. S2a), 13 Australian isolates and three New Zealand isolates fell into the basal-B2 subgroup, two Australian isolates and 13 New Zealand isolates fell into the world-B2 group, and only one Australian and one New Zealand isolate fell into the world-B3 group. In the Region B tree (Fig. S2b), two Australian isolates and three New Zealand isolates fell into the basal-B2 subgroup, three Australian isolates and 13 New Zealand isolates fell into the world-B2 group, and 12 Australian isolate fell into the world-B3 subgroup. In the Region C tree (Fig. S2c), two Australian and three New Zealand isolates fell into the basal-B2 subgroup, no Australian and nine New Zealand isolates fell into the world-B2 subgroup, and 14 Australian and five New Zealand isolates fell into the world-B3 group. These trees confirmed that none of the Australian or New Zealand isolates had regions from Orchis, basal-BR or Asian-BR parental lineages. For further sequence analyses, we used the parts of the HC-Pro, P3 and NIb encoding regions that contained no recombination cross-over points, in any sequence. We called these the HC-Pro*, P3* and NIb* regions (see Methods).

Genetic population structure

The haplotype and nucleotide diversities of the Australian and New Zealand populations in the TuMV

phylogenetic groups were compared (Table S3). In most cases, haplotype diversity values were large and nucleotide diversity values were small (i.e., few single-nucleotide polymorphisms, but most of them unique). The nucleotide diversities of Australian isolates in most phylogenetic groups were greater than those of New Zealand isolates in the HC-Pro* and P3* regions; the two regions were similarly variable, but the NIb* region less variable. Overall, the combination of high haplotype diversity and overall lack of nucleotide diversity within individual geographical groups indicate that there has been a recent population expansion. This was confirmed by the Bayesian molecular-clock analyses described below.

Evolutionary rates and timescales

A Bayesian phylogenetic method (Drummond *et al*., 2012) was used to estimate the evolutionary rates and timescales for the HC-Pro*, P3* and NIb* regions. The best-supported demographic models were of constant size for all protein-encoding regions (Table 1). A relaxed-clock model provided a better fit than the strict-clock model, indicating the presence of rate variation among lineages. The presence of an adequate temporal signal in the data was confirmed using a date-randomization test (Fig. S3), in which the calculated rate estimate was compared with estimates from date-randomized replicates. We note, however, that the date-randomization test involves the assumption of random phylogenetic and temporal sampling, which is unlikely to be met by our data set. The impact of non-random sampling on the test is unknown.

The mean estimated substitution rates were 1.47×10^{-3} subs/site/year for HC-Pro*, 1.35×10^{-3} subs/site/year for P3* and 1.30×10^{-3} subs/site/year for NIb* regions (Table 1). Mean estimates of the age of the root of all the TuMVs were 610 years for HC-Pro*, 806 years for P3*, and 679 years for NIb* regions (Table 1, Fig. 3, Fig. S4). The relationships between Australian and New Zealand isolates and European country isolates were also confirmed using ML trees (Fig. S5). These estimates are potentially inflated by the inclusion of transient polymorphisms that would normally be removed by purifying selection over longer timeframes (Duchêne et al., 2014; Gibbs et al., 2010; Wertheim & Kosakovsky Pond, 2011).

Dating of recombination events

We estimated the ages of recombination events (Table 2, Fig. S6) using the method described by Visser *et al*. (2012). Recombinant sequences were split into their separate regions and realigned using gaps. When a sequence is a recombinant with two 'parents', it is split into two regions and the empty sites are filled with gaps. In this way, one recombinant sequence becomes two non-recombinant sequences, each with missing data. The analysis of the split sequences indicated that the interlineage recombination sites of Australian isolates with basal-B2 and world-B3 parents at nt 818 and nt 3475 of Australian isolates occurred 50-10 and 51-22 years ago, and the intralineage recombination sites of basal-B2 parents nt 6019 of New Zealand isolates occurred 75-19 years ago. The intralineage recombination sites of world-B parents at nt 5602 and nt 5665 of New Zealand isolates occurred 49-20 and 22-11 years ago. Therefore, the ages of basal-B parent related recombination events were older.

Plausible routes of TuMV migraton into Australia and New Zealand

We investigated the likely routes of TuMV migration into Australia and New Zealand using a Bayesian phylogeographic analysis (Lemey *et al*., 2009) of the HC-Pro*, P3* and NIb* datasets of non-recombinant isolates. Isolates were tagged with their countries of provenance. Our results indicate that TuMV migrated

between European countries and from European countries to Australia and to New Zealand. Therefore, we further investigated the routes of migration for each phylogenetic subgroup; focusing on the basal-B2, world-B2 and world-B3 subgroups because Australian and New Zealand isolates were only from these three subgroups (Fig. 4). For instance, migrations from Germany to Australia and to New Zealand were supported by results from the HC-Pro* and P3* regions [Bayes factor (BF)=54 and BF=22 for HC-Pro*, BF=129 and BF=63 for P3*, respectively] and from Germany to Australia was supported by NIb* region (BF=55) for basal-B2 subgroup (Fig. 4a). The estimated ages of migrations were about 83-70 [95% highest posterior density interval (HPD): 159-23] years ago for Australia and 45-32 (95% HPD: 72-16) years ago for New Zealand. The migration from Australia to New Zealand was supported only by NIb* region (BF=68) and it was 33 (95% HPD: 54-20) years ago. In contrast, the migrations of the world-B2 and world-B3 subgroups occurred more recently and these were within 42-20 (95%HPD; 63-12) years before present (Fig. 4b, c and d). There was also significant support (BF≥100) for migration between the neighbouring countries within Europe and East Asia (Fig. S7).

This analysis was confirmed by ML trees of the HC-Pro*, P3* and NIb* regions Fig S5. These show which isolates were closest to the Australian and New Zealand clusters, and were therefore likely to be from the populations that provided the migrants. For example, the ML analysis found that all the basal-B2 gene populations from Australia and New Zealand probably came from Germany (Fig. 4, Fig. S5), and it can be seen in the trees that the closest isolates were, for Australian populations, two from Germany, DEU 7 and AllA, and one from the USA (PV134) , and for the New Zealand isolates, two from Germany, TIGA and TIGD, with many European isolates in sister clusters. Similarly the more recent world-B2 and -B3 migrants were, on average, closest to UK isolates (Fig. S5).

DISCUSSION

We present here a preliminary assessment of the genetic structure of the TuMV populations of Australia and New Zealand. Previously published data, including five CP gene sequences and one genomic sequence, indicated that there were at least two distinct lineages of TuMV in Australia (Gibbs *et al.*, 2006) and one in New Zealand, and that all were closely related to TuMVs found in Europe. Here we have reported the genomic sequences of 34 isolates from Australia and New Zealand, allowing us to assess their position in the world TuMV population in both space and time. We used both Bayesian and ML methods to analyse dated isolates of known provenance. The two approaches gave closely similar and internally consistent results, which we have used to determine when, and from where, the present Australian and New Zealand populations of TuMV came. This also allowed us to evaluate whether they arrived by natural means or with human assistance.

Many of the isolates from Australia and New Zealand are recombinants, but all of the recombination cross-over points found in these isolates are in genomic positions that are clearly different from those in all isolates (195 isolates) from other parts of the world and known to us. This fact, together with estimates of the dates for the recombination events and for the divergence of the Australian and New Zealand isolates from their nearest relatives, shows that migration probably preceded recombination.

The Australian and New Zealand populations of TuMV are closely related to viruses found in Germany and the UK, which, in turn, are related to older and more diverse European populations. This indicates that the Australian populations are recent migrants from Germany and the UK. Fig. 4 summarizes the sources and likely migration dates of three genomic regions of the lineages that have been found. The clearest evidence is from the basal-B2 populations. These were the first to arrive in the Australia (from Germany, about 70 years ago) and New Zealand (from Germany about 35 years ago, its NIb gene via Australia). The isolates of the other two TuMV taxonomic groups, world-B2 and world-B3, arrived from the UK about 35 years ago. Thus, there is evidence that

a minimum of three TuMV isolates evaded quarantine and entered Australia, and four entered New Zealand. Turnip mosaic disease was first recorded in both countries in the 1930s (Chamberlain, 1936; Samuel, 1931). This is consistent with the dates that we have obtained in our analyses of the Australian population, but may indicate that the earlier New Zealand population might not have been sampled, or might have not survived.

Our evidence indicates that TuMV is most likely to have migrated to Australia and New Zealand in plant materials imported from Europe (16-18,000 km), rather than from South East Asia (5-9,000 km) in imports, or by natural means such as flying aphids,. This is because the known TuMV populations of the south and east of Asia, Japan, Vietnam and China, are genetically linked but distinct from those of Europe (Nguyen *et al.*, 2013ab; Tomimura *et al.*, 2003). Australia and New Zealand were first populated around 50,000 and 800 years ago, respectively, but regular trade to these countries, and the development of agriculture, did not start until about 600 years ago and was mostly derived from Europe. These incursions have grown in volume ever since, especially since World War II (Wace, 1985). Although human migration to Australia and New Zealand was, until recently, dominated by people from the UK, an equal number came from mainland Europe. Hence, the UK and Europe have been important sources of crop seeds and weeds (Wace 1985, Zubareva et al., 2013). As a consequence, it is not surprising that the TuMV populations of Australia and New Zealand are most closely related to those of Germany and the UK.

Interestingly the basal-B2 isolates from Australia and New Zealand are closest to German isolates from *Alliaria officinalis* (AllA), *Lactuca sativa* (DEU7) and *Tigridia* spp. (TIGA and TIGD). All of these have horticultural, rather than agricultural, links suggesting that the first TuMV migrants to the Antipodes were in horticultural materials. In contrast, the isolates closest to the world-B2 and world-B3 migrants were from brassicas.

Our analysis of TuMV provides a significant snapshot of the evolution and emergence of a highly pathogenic virus in association with human immigration and agriculture history. It will also be important in the future to study the effect of large changes in TuMV host populations in these countries, such as those that may be caused by the recent widespread increase in area of crops of canola throughout Australia, and the emergence of this plant as one of the commonest roadside weeds in the regions where it is being grown. There is also a need to investigate TuMV in the brassica crops of Western Australia and the many native species of *Brassicaeae* in all of Oceania.

METHODS

Virus isolates. The brassica crop-producing areas of Australia and New Zealand were surveyed during the growing seasons of 1994–2011. All collected samples were tested by direct double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) using TuMV antiserum (Clark & Adams, 1977). Details of the isolates, their place of origin, original host plant, year of isolation, and host type are shown in Table S1, together with details of the isolates used in the analyses and for which complete genomic sequences have already been reported.

It is essential to clone the viral isolates being studied before they are sequenced in studies of plant virus evolution because of the high frequency of mixed infections in the field, not only with other viruses but also with other isolates of the same virus (Ohshima *et al.*, 2002; Tomitaka & Ohshima, 2006). In the earlier studies, TuMV isolates were usually cloned by single lesion isolations. Consequently, we found very few mismatches in the sequences between the overlapping RT-PCR products (Nguyen *et al.*, 2013a). Moreover, cloning is required when attempting to analyse recombinational events and the genetic structure of populations. In present study,

all the isolates were inoculated to *C. quinoa* or *Nicotiana tabacum* cv. Samson and serially cloned through single lesions at least three times. They were propagated in *B. rapa* cv. Hakatasuwari or *N. benthamiana* plants. Plants infected systemically with each of the TuMV isolates were homogenized in 0.01 M potassium phosphate buffer (pH 7.0), and the isolates were mechanically inoculated to young plants of *Brassicaeae* plants (Nguyen *et al.*, 2013b). Inoculated plants were kept for at least four weeks in a glasshouse at 25°C.

Viral RNA and sequencing. We determined the genomic sequences of TuMV collected in different areas of Australia and New Zealand. The viral RNAs were extracted from TuMV-infected *B. rapa* cv. Hakatasuwari or *N. benthamiana* leaves using Isogen (Nippon Gene, Japan). The RNAs were reverse transcribed by PrimeScript® *Moloney murine leukemia virus* (MMLV) reverse transcriptase (TakaraBio, Japan) and amplified using high-fidelity Platinum™ *Pfx* DNA polymerase (Invitrogen, USA). The products obtained by reverse transcription and polymerase chain reaction (RT-PCR) were separated by electrophoresis in agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen K. K., Japan). Sequences from each isolate were determined using four to five overlapping independent RT-PCR products to cover the complete genome. To ensure that they were from the same genome and were not from different components of a genome mixture, the sequences of the RT-PCR products of adjacent regions of the genome overlapped by around 200-350 nt. Each RT-PCR product was sequenced by primer walking in both directions using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (Life Technologies, USA) and an Applied Biosystems 310 and 3130X Genetic Analyzer. Sequence data were assembled using BioEdit version 5.0.9 (Hall, 1999).

Alignment of sequences. The genomic sequences of 229 isolates (Table S1) were used for phylogenetic and recombination analysis. Two sequences of Japanese yam mosaic virus (JYMV) (Fuji & Nakamae, 1999; 2000), one of scallion mosaic virus (ScaMV) (Chen *et al.*, 2002), one of narcissus yellow stripe virus (NYSV) (Chen *et al.*, 2006) and two of narcissus late season yellows virus (NLSYV) (Lin *et al.*, 2012; Wylie *et al.*, 2014) were used as outgroup taxa because these are the other members of the TuMV phylogenetic group. The amino acid sequences of the polyproteins were aligned with the outgroup sequences using CLUSTAL_X2 (Larkin *et al.*, 2007) with TRANSALIGN (kindly supplied by Georg Weiller) to maintain the degapped alignment of the encoded amino acids. The aligned subsequences were then reassembled to form complete polyprotein sequences 8,922 nt long. The polyprotein sequences were then joined with aligned 5' and 3' NCR sequences of each isolate. This produced sequences 9,087 nt long, excluding the 35 nucleotides that were used to design the primer for RT-PCR amplification.

Recombination analyses. Putative recombination breakpoints in all sequences were identified using RDP (Martin & Rybicki, 2000), GENECONV (Sawyer, 1999), BOOTSCAN (Salminen *et al.*, 1995), MAXCHI (Maynard-Smith, 1992), CHIMAERA (Posada & Crandall, 2001) and SISCAN programs (Gibbs *et al.*, 2000) implemented in the RDP4 package (Martin *et al.*, 2010) and also the original PHYLPRO (Weiller, 1998) and SISCAN version 2 (Gibbs *et al.*, 2000) programs. First, we checked for incongruent relationships using the programs implemented in RDP4. These analyses were done using default settings for the different detection programs and a Bonferroni-corrected *P*-value cut-off of 0.05 or 0.01. All isolates that had been identified as likely recombinants by the programs in RDP4, supported by three different methods with an associated *P*-value of $>1.0\times10^{-6}$, were re-checked using the original PHYLPRO version 1 and SISCAN version 2. We checked 100- and 50-nt slices of all sequences for evidence of recombination using these programs. These analyses also determined which non-recombinant sequences had regions that were closest to the regions of the recombinant sequences and hence indicated the likely lineages that provided those regions of the recombinant genomes. For convenience, we called these the 'parental isolates' of recombinants. Finally, TuMV sequences were also aligned

without outgroup sequences, producing sequences 9,710 nt long excluding the 35 nucleotides. We checked these for evidence of recombination using the programs described above.

Phylogenetic analyses. The phylogenetic relationships of the aligned full and partial genomic sequences were inferred using the Neighbor-Net method in SPLITSTREE version 4.11.3 (Huson *et al.*, 2006) and using ML in PhyML version 3 (Guindon & Gascuel, 2003). For the ML analysis, we used the general time-reversible (GTR) model of nucleotide substitution, with rate variation among sites modelled using a gamma distribution and a proportion of invariable sites (GTR+ I+r_4). This model was selected in R (Schliep, 2011) using the Bayesian information criterion, which has been shown to perform well in a variety of scenarios (Luo *et al.*, 2010). Branch support was evaluated by bootstrap analysis based on 1000 pseudoreplicates. The inferred trees were displayed by TREEVIEW (Page, 1996) and FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Nucleotide and amino acid similarities were estimated using the Kimura two-parameter method (Kimura, 1980) and the Dayhoff PAM 001 matrix (Schwarz & Dayhoff, 1979), and the within-population diversities were assessed using MEGA version 6 (Tamura *et al*, 2013).

Estimation of substitution rates and divergence times. Bayesian phylogenetic analyses were performed in BEAST version 1.8.0 (Drummond *et al.*, 2012) to estimate the evolutionary rate and timescale of TuMV populations. Analyses were based on partial protein-encoding regions of HC-Pro (nt 1460-2494, corresponding to the positions in original UK 1 genome), P3 (nt 2591-3463) and NIb (nt 7208-8068). Recombinant sequences and some nucleotides from recombination ends were discarded from the three regions (see Fig. 2). We call these regions HC-Pro*, P3* and NIb*, respectively. The sampling times of the sequences were used to calibrate the molecular clock. BFs were used to select the best-fitting molecular-clock model and coalescent priors for the tree topology and node times. Strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clocks (Drummond *et al.*, 2006) were compared, and five demographic models (constant population size, expansion growth, exponential growth, logistic growth, and the Bayesian skyline plot) were also compared.

Posterior distributions of parameters, including the tree, were estimated by Markov Chain Monte Carlo (MCMC) sampling. Samples were drawn every 10^4 MCMC steps over a total of 10^8 steps, with the first 10% of samples discarded as burn-in. Acceptable sampling from the posterior and convergence to the stationary distribution were checked using the diagnostic software Tracer version 1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Tree files were generated with software included in the BEAST package and Bayesian maximum-clade-credibility (MCC) trees were displayed by FigTree version 1.4.2.

Sampling times need to have a sufficient spread in relation to the substitution rate to allow reliable estimation substitution rates and divergence times from heterochronous sequence data (Drummond *et al.*, 2003). The temporal signal in our data sets was checked by comparing our rate estimates with those from ten date-randomized replicates. A data set was considered to have sufficient temporal structure when the mean rate estimate from the original data set was not contained in any of the 95% credibility intervals of the rates estimated from the date-randomized replicates. This follows the approach taken in previous studies of viruses (Duchêne *et al.*, 2014; Ramsden *et al.*, 2009).

The spatial population dynamics of TuMV through time were inferred in BEAST using a diffusion model with discrete location states (Lemey *et al.*, 2009). This approach uses a model that describes the spatial migration of TuMV lineages throughout their demographic history. The most important pairwise diffusions can be

identified using BFs (Suchard *et al.*, 2001). We produced a graphical animation of the estimated spatio-temporal movements of TuMV lineages using SPREAD version 1.0.6 (Bielejec *et al.*, 2011) and Google Earth (http://www.google.com/earth).

Demographic analyses. Haplotype and nucleotide diversities were estimated using DnaSP version 5.0 (Librado & Rozas, 2009). Haplotype diversity refers to the frequency and number of haplotypes (i.e. unique combinations of nucleotide polymorphisms) in the population. Nucleotide diversity estimates the average pairwise differences among sequences. Non-synonymous (dN) and synonymous (dS) substitution (dN/dS) ratios were calculated for each protein-encoding region using the Pamilo-Bianchi-Li method in MEGA version 6 (Tamura *et al.*, 2013).

ACKNOWLEDGEMENTS

We thank Yasuhiro Tomitaka, Haruka Imamura, Takuya Noguchi (Laboratory of Plant Virology, Saga University), and Yukio Nagano (Analytical Research Center for Experimental Sciences, Saga University) for their careful technical assistance. This work was in part funded by Saga University and supported by JSPS KAKENHI Grant Number 18405022 and 24405026. Isolates analysed in the present study were officially imported to Japan with permission from the Japanese Plant Protection Station, Ministry of Agriculture, Forestry and Fisheries Japan.

REFERENCES

Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27, 2910–2912.

Chamberlain, E. E. (1936). Turnip-mosaic. A virus disease of crucifers. *New Zeal Jour Agr* 53, 321–330.

Chen, J., Zheng, H. Y., Chen, J. P. & Adams, M. J. (2002). Characterisation of a potyvirus and a potexvirus from Chinese scallion. *Arch Virol* 147, 683–693.

Chen, J., Lu, Y.-W., Shi, Y.-H., Adams, M. J. & Chen J.-P. (2006). Complete nucleotide sequence of the genomic RNA of Narcissus yellow stripe virus from Chinese narcissus in Zhangzhou city, China. *Arch Virol* 151, 1673–1677.

Chung, B. Y.-W., Miller, W. A., Atkins, J. F. & Firth, A. E. (2008). An overlapping essential gene in the *Potyviridae*. *Proc Natl Acad Sci U S A* 105, 5897–5902.

Clark, M. F. & Adams, A. N. (1977). Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J Gen Virol* 34, 475–483.

Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. G. (2003). Measurably evolving populations. *Trends Ecol Evol* 18, 481–488.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4, e88.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29, 1969–1973.

Duchêne, S., Holmes, E. C. & Ho, S. Y. W. (2014). Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc Roy Soc B* 281, 20140732.

Fletcher, J. D., Lister, R. A., Bulman, S. R. & Heenan, P. B. (2010). First record of *Turnip mosaic virus* in *Pachycladon* spp. (Brassicaceae): an endangered native plant species in New Zealand. *Australas Plant Dis Notes* 5, 9–10.

Fuji, S. & Nakamae, H. (1999). Complete nucleotide sequence of the genomic RNA of a Japanese yam mosaic virus, a new potyvirus in Japan. *Arch Virol* 144, 231–240.

Fuji, S. & Nakamae, H. (2000). Complete nucleotide sequence of the genomic RNA of a mild strain of Japanese yam mosaic potyvirus. *Arch Virol* 145, 635–640.

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

Gibbs, A. J., Fargette, D., Garcia-Arenal, F. & Gibbs, M. J. (2010). Time – the emerging dimension of plant virus studies. *J Gen Virol* 91, 13–22.

Gibbs, A. J., Mackenzie, A. M. Wei, K.-J. & Gibbs, M. J. (2006). The potyviruses of Australia. *Arch Virol* 153, 1411–1420.

Gibbs, A. J., Ohshima, K., Gibbs, M. & García-Arenal, F. (2008). More about plant virus evolution; past, present and future. In *Origin and Evolution of Viruses*, 2nd ed, pp. 229–249. Edited by E. Domingo, C. R. Parrish & J. J. Holland. San Diego: Elsevier Academic Press.

Gibbs, A. J. & Ohshima, K. (2010). Potyviruses and the digital revolution. *Annu Rev Phytopathol* **48**, 205–223.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582.

Guindon, S. & Gascuel, O. (2003). A simple fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52, 696–704.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41, 95–98.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23, 254–267.

Jenner, C. E., Sanchez, F., Nettleship, S. B., Foster, G. D., Ponz, F. & Walsh, J. A. (2000). The cylindrical inclusion gene of *Turnip mosaic virus* encodes a pathogenic determinant to the brassica resistance gene *TuRB01*. *Mol Plant Microbe Interact* 13, 1102–1108.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16, 111–120.

King, A. M. Q., Adams, M. J, Carstens, E. B. & Lefkowitz, E. J. (2012). Virus Taxonomy: Ninth report of the international committee on taxonomy of viruses. In *Family Potyviridae*, pp. 1069–1077. San Diego: Elsevier Academic Press.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics* 23, 2947–2948.

Lecoq, H., Wipf-Scheibel, C., Chandeysson, C., Lê Van, A., Fabre, F. & Desbiez, C. (2009). Molecular epidemiology of *Zucchini yellow mosaic virus* in France: an historical overview. *Virus Res* 141, 190–200.

Lefeuvre, P., Martin, D. P., Harkins, G., Lemey, P., Gray, A. J. A., Meredith, S., Lakay, F., Monjane, A., Lett, J.-M. & other authors (2010). The spread of tomato yellow leaf curl virus from the middle east to the world. *PLoS Pathog* 6, e1001164.

Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5, e1000520.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.

Lin, S.-Q., Shen, J.-G., Gao, F.-L., Cai, W., Huang, Z., Xie, L.-Y. & Wu, Z.-J. (2012). Complete genome sequence of narcissus latent season yellows virus infecting Chinese narcissus in China. *Arch Virol* 157, 1821–1824.

Luo, A., Qiao, H., Zhang, Y., Shi, W., Ho, S. Y. W., Xu, W., Zhang, A. & Zhu, C. (2010). Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evol Biol* 10, 242.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463.

Maynard-Smith, J. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* 34, 126–129.

Nguyen, H. D., Tran, H. T. N. & Ohshima, K. (2013a). Genetic variation of the *Turnip mosaic virus* population of Vietnam: A case study of founder, regional and local influences. *Virus Res* 171, 138–149.

Nguyen, H. D., Tomitaka, Y., Ho, S. Y. W., Duchêne, S., Vetten, H.-J., Lesemann, D., Walsh, J. A., Gibbs, A. J. & Ohshima, K. (2013b). Turnip mosaic potyvirus probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. *PLoS ONE* 8, e55336.

Ochoa Corona, F. M., Lebas, B. S. M., Elliott, D. R., Tang, J. Z. & Alexander, B. J. R. (2007). New host records and new host family range for *Turnip mosaic virus* in New Zealand. *Australas Plant Dis Notes* 2, 127–130.

Ogawa, T., Tomitaka, Y., Nakagawa, A. & Ohshima, K. (2008). Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North America and European populations. *Virus Res* 131, 199–212.

Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., Sano, T., Azuhata, F., Walsh, J. A., Fletcher, J., Chen, J., Gera, A. & Gibbs, A. J. (2002). Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. *J Gen Virol* 83, 1511–1521.

Ohshima, K., Tomitaka, Y., Wood, J. T., Minematsu, Y., Kajiyama, H., Tomimura, K. & Gibbs, A. J. (2007). Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *J Gen Virol* 88, 298–315.

Ohshima, K., Akaishi, S., Kajiyama, H., Koga, R. & Gibbs, A. J. (2010). Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. *J Gen Virol* 91, 788–801.

Page, R. D. M. (1996). Treeview: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12, 357–358.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98, 13757–13762.

Ramsden, C., Holmes, E. C. & Charleston, M. A. (2009). Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol Biol Evol* 26, 143–153.

Rocha, C. S., Castillo-Urquiza, G. P., Lima, A. T. M., Silva, F. N., Xavier, C. A. D., Hora-Júnior, B. T., Beserra-Júnior, J. E. A., Malta, A. W. O., Martin, D. P. & other authors (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J Virol* 87, 5784–5799.

Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning. *AIDS Res Hum Retroviruses* 11, 1423–1425.

Samuel, G. (1931). Summary of plant disease records in south Australia for the two years ending June 30, 1930. *So Aust Dept Agr Jour* 34, 746.

Sawyer, S. A. (1999). GENECONV: a computer package for the statistical detection of gene conversion, Distributed by the Author. Department of Mathematics. Washington University, St Louis, available at http://www.math.wustl.edu/~sawyer.

Schliep, K. P. (2011). Phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593.

Schwarz, R. & Dayhoff, M. (1979). Matrices for detecting distant relationships. In *Atlas of protein sequences*, pp. 353–358. Edited by M. Dayhoff. Washington, DC: National Biomedical Research Foundation.

Schwinghamer, M. W., Schilg, M. A., Walsh, J. A., Bambach, R. W., Cossu, R. M., Bambridge, J. M., Hind-Lanoiselet, T. L., McCorkell, B. E. & Cross, P. (2014). *Turnip mosaic virus*: potential for crop losses in the grain belt of New South Wales, Australia. *Australas Plant Pathol* Published online 2014 July 2 DOI: 10.1007/s13313–014–0304–9.

Seo, J. K., Ohshima, K., Lee, H. G., So, M., Choi, H. S., Lee, S. H., Sohn, S. H. & Kim, K. H. (2009). Molecular variability and genetic structure of the population of *Soybean mosaic virus* based on the analysis of complete genome sequences. *Virology* 393, 91–103.

Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. (2001). Bayesian selection of continuous- time Markov chain evolutionary models. *Mol Biol Evol* 18, 1001–1013.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30, 2725–2729.

Tomimura, K., Gibbs, A. J., Jenner, C. E., Walsh, J. A. & Ohshima, K. (2003). The phylogeny of *Turnip mosaic virus*; comparisons of 38 genomic sequences reveal a Eurasian origin and a recent 'emergence' in east Asia. *Mol Ecol* 12, 2099–2111.

Tomimura, K., Spăk, J., Katis, N., Jenner, C. E., Walsh, J. A., Gibbs, A. J. & Ohshima, K. (2004). Comparisons of the genetic structure of populations of *Turnip mosaic virus* in West and East Eurasia. *Virology* 330, 408–423.

Tomitaka, Y. & Ohshima, K. (2006). A phylogeographic study of the *Turnip mosaic virus* population in East Asia reveals an ‘emergent’ lineage in Japan. *Mol Ecol* 15, 4437–4457.

Tomitaka, Y., Yamashita, T. & Ohshima, K. (2007). The genetic structure of populations of *Turnip mosaic virus* in Kyushu and central Honshu, Japan. *J Gen Plant Pathol* 73, 197–208.

Visser, J. C., Bellstedt, D. U. & Pirie, M. D. (2012). The recent recombinant evolution of a major crop pathogen, *Potato virus Y*. *PLoS ONE* 7, e50631.

Wace, N. (1985). The isolated continent. In *Pests and parasites as migrants*, pp. 3–22. Edited by A. Gibbs & R. Meischke. Cambridge: Cambridge University Press.

Walsh, J. A. & Jenner, C. E. (2002). *Turnip mosaic virus* and the quest for durable resistance. *Mol Plant Pathol* 3,

289–300.
Weiller, G. F. (1998). Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Mol Biol Evol* 15, 326–355.
Wertheim, J. O. & Kosakovsky Pond, S. L. (2011). Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol* 28, 3355-3365.
Wylie, S. J., Li, H., Sivasithamparam, K. & Jones, M. G. K. (2014). Complete genome analysis of three isolates of narcissus late season yellows virus and two of narcissus yellow stripe virus: three species or one? *Arch Virol* 159, 1521–1525.
Yasaka, R., Nguyen, H. D., Ho, S. Y. W., Duchêne, S., Korkmaz, S., Katis, N., Takahashi, H., Gibbs, A. J. & Ohshima, K. (2014). The temporal evolution and global spread of *Cauliflower mosaic virus*, a plant pararetrovirus. *PLoS ONE* 9, e85641.
Zubareva, I. A, Vinogradova, S. V., Gribova, T. N., Monakhos, S. G., Skryabin, K. G. & Ignatov, A. N. (2013). Genetic diversity of *Turnip mosaic virus* and the mechanism of its transmission by Brassica seeds. *Doklady Biochem. Biophy* 450, 119-122.

Figure legends

Fig. 1. Maps showing the provenance of the turnip mosaic virus isolates from Australia and New Zealand that were studied. Dots on the map correspond to the isolates listed in Table S1.

Fig. 2. Recombination maps of turnip mosaic virus genomes of Australian and New Zealand isolates. The estimated nucleotide positions of the recombination sites and those in parentheses are shown relative to the 5' end of the genome using the numbering of the aligned sequences used in the present study and the UK 1 isolate (Jenner *et al.*, 2000), respectively. Vertical arrows and lines show estimated recombination sites (listed in Table S2). The grey and dot boxes indicate basal-B and world-B parents, respectively. The horizontal arrows show the regions (A, B and C) used to infer trees from non- and intralineage recombinant sequences (shown in Supplementary Fig. S2). The recombination sites newly identified in the present study (black) or those identified in earlier studies (red) are separately listed.

Fig. 3. Bayesian maximum-clade-credibility chronogram inferred from the protein-encoding region of turnip mosaic virus. The tree was calculated from the 186 non-recombinant isolates of partial protein 3 (P3*) (nt 2591-3463, corresponding to the positions in original UK 1 genome) sequences. The non-Australian and non-New Zealand (sub)groups of basal-B1, basal-BR, Asian-BR and world-B1 are collapsed. Horizontal blue bars represent the 95% highest posterior density interval (HPD) intervals of estimates of node ages. The bar graph shows the root state posterior probabilities for each location (colored bars). Gray bars show probabilities obtained with 10 randomizations of the tip locations.

Fig. 4. Plausible historical migration pathways of turnip mosaic virus inferred using partial helper-component proteinase (HC-Pro*), protein 3 (P3*) and nuclear inclusion b (NIb*) regions. Details of the regions are given in the Methods. Migration routes only for Australia and New Zealand are shown, and only when supported by a Bayes factor (BF) greater than 10. Only the migration pathways for basal-B2 (a), world-B2 (b) and world-wB3 (c) isolates are shown. For each subgroup migration, the BF and estimated age in years before present (YBP) are shown (d). The 95% highest posterior density interval (HPD) interval for each age estimate is given in parentheses. We analysed basal-B group isolates instead of basal-B2 subgroup isolates because too few isolates belong to basal-B2 subgroup, and only the BFs from the basal-B2 subgroup are listed. BF values can be interpreted as follows: $10 \leq BF < 30$, strong support; $30 \leq BF < 100$, very strong support; and $BF \geq 100$, decisive support.

Table 1. Details of the data sets used for estimation of nucleotide substitution rate and time to the most recent common ancestor for turnip mosaic virus

Parameter	Protein encoding region*		
	Helper-component proteinase	Protein 3	Nuclear inclusion b
Best-fit substitution model	GTR + I + Γ_4	GTR + I + Γ_4	GTR + I + Γ_4
Best-fit molecular clock model	Relaxed uncorrelated exponential	Relaxed uncorrelated exponential	Relaxed uncorrelated exponential
Best-fit population growth model	Constant size	Constant size	Constant Size
Sequence length (nt)	927	897	891
No. of sequences	180	186	182
Sampling date range	1968-2012	1970-2012	1968-2012
Chain length (in millions)	100	100	100
TMRCA† (years)			
All isolates	610 (233-1156)‡	806 (274-1630)	679 (205-1502)
Australia			
basal-B2 subgroup	80 (119-52) [n=12]§	44 (65-25) [n=13]	36 (51-19) [n=2]
world-B2 subgroup	14 (20-9) [n=2)	9 (29-5) [n=2]	NA‖ [n=0]
world-B3 subgroup	27 (36-16) [n=2]	NA [n=1]	46 (67-28) [n=14]
New Zealand			
basal-B2 subgroup	14 (29-4) [n=3]	16 (28-4) [n=3]	22 (31-12) [n=3]
world-B2 subgroup	56 (83-30) [n=13]	16 (23-9) [n=10]	17 (24-11) [n=9]
world-B3 subgroup	NA [n=1]	NA [n=1]	68 (100-36) [n=5]
Substitution rate (nt/site/year)	1.47×10^{-3} (1.08×10^{-3}-1.89×10^{-3})	1.35×10^{-3} (9.50×10^{-4}-1.77×10^{-3})	1.30×10^{-3} (9.07×10^{-4}-1.77×10^{-3})
dN/dS¶	0.025	0.120	0.030
No. of variable sites#	511	536	493

Partial helper-component (HC-Pro), Protein 3 (P3*) and nuclear inclusion b (NIb*) regions (see Methods).
†Time to the most recent common ancestor (years).
‡95% highest posterior density interval in parentheses.
§The number of isolates in square brackets.
‖Not available.
¶Non-synonymous (dN) and synonymous (dS) substitution (dN/dS) ratios were calculated for three protein-encoding regions using the Pamilo-Bianchi-Li method.
#The number of variable sites.

Table 2. The estimate of the time of recombination events of turnip mosaic virus in Australia and New Zealand

Recombination site*	Parent (5' × 3')	Recombinant type†	Recombination age (YBP)‡	Stem age (YBP)	Crown age (YBP)
Australia					
nt 818	world-B3 × basal-B2	Inter	50-10	50-18	23-10
nt 1080	world-B2 × basal-B2	Inter	54-13	54-21	50-13
nt 1341	world-B3 × basal-B2	Inter	48-21	48-25	40-21
nt 1851	basal-B2 × basal-B2	Intra	35-9	35-14	22-9
nt 2530	basal-B2 × basal-B2	Intra	48-23	48-25	37-23
nt 2742	world-B3 × basal-B2	Inter	38-14	38-14	N/A§
nt 3475	basal-B2 × world-B3	Inter	51-22	51-26	49-22
New Zealand					
nt 5602	world-B3 × world-B3	Intra	49-20	49-37	46-20
nt 5665	world-B2 × world-B3	Intra	22-11	22-11	N/A
nt 6019	basal-B2 × basal-B2	Intra	75-19	75-28	70-19
nt 8071	world-B2 × world-B3	Intra	20-10	20-14	19-10
Both countries					
nt 1174	world-B2 × world-B2	Intra	27-14	27-18	24-14
nt 5219	world-B2 × world-B2	Intra	24-13	24-17	21-13
nt 6132	world-B2 × world-B3	Intra	28-16	28-20	25-16

*The ages of major recombination sites in Australia and New Zealand were estimated with reference to the results of Bayesian phylogenetic analyses shown in Figs. S6 (a) and S6 (b), respectively. The common recombination sites in both countries were estimated from the tree including all isolates (data not shown). Nucleotide positions show locations of individual genes numbered as in the original UK 1 genome (Jenner *et al*., 2000).
†Inter; interlineage recombination site, Intra; intralineage recombination site.
‡The oldest and youngest ages are shown. The oldest and the youngest ages were estimated from the stem and crown ages, respectively. Estimates are given in years before present (YBP).
§Not available