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. Gibbs8 and Kazusato Ohshima1,2

Correspondence
Kazusato Ohshima
ohshimak@cc.saga-u.ac.jp

Received 7 October 2014
Accepted 6 November 2014

Turnip mosaic virus (TuMV) is a potyvirus that is transmitted by aphids and infects a wide range of plant species. We investigated the evolution of this pathogen by collecting 32 isolates of TuMV, mostly from Brassicaceae plants, in Australia and New Zealand. We performed a variety of sequence-based phylogenetic and population genetic analyses of the complete genomic sequences and of three non-recombinogenic regions of those sequences. The substitution rates, divergence times and phylogeographical patterns of the virus populations were estimated. Six inter- and seven 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. The 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 that 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. To our knowledge, our study presents the first population genetic analysis of TuMV in Australia and New Zealand. We have shown that the time of migration of TuMV correlates well with the establishment of agriculture and migration of Europeans to these countries.

†Present address: 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.

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this study are AB989628–AB989659.

Seven supplementary figures and three tables are available with the online Supplementary Material.
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., 2008a, b; Gibbs & Ohshima, 2010). This is especially true 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 (Lefèvre 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) (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, mostly from the family Brassicaceae (Walsh & Jenner, 2002). TuMV, like other potyviruses, is transmitted by aphids in a 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 9833 nt. This is translated into one large polyprotein, which hydrolyses 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 region encoding protein 3 (P3) (Chung et al., 2008).

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 south-east Europe, Asia Minor and mid-Eurasia (Nguyen et al., 2013a, b; Tomimura et al., 2004). Crops of the Brassicaceae that are most commonly cultivated in Europe are Brassica spp., whereas both Brassica and Raphanus crops are important in Asia and Asian countries (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 systemically, causing a mosaic pattern on the systemically infected leaves, but do not infect Raphanus plants. [B] host-type isolates infect most Brassica spp. systemically, causing a mosaic pattern on systemically infected leaves, and infect Raphanus plants latently and occasionally. [BR] host-type isolates infect both Brassica and Raphanus plants systemically, causing a mosaic pattern on the systemically infected leaves. The basal-B cluster of (B) or B host-type isolates is most variable, is paraphyletic to the other lineages and has been isolated from both non- and Brassicaceae plants. The world-B cluster is the most variable and widespread cluster; most European isolates do not infect Raphanus spp., 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 was characterized by symptoms, host range, and sap and aphid transmission. More recent reports from Australia (Gibbs et al., 2008b; 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 34 of the isolates. We used 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 made 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 were 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, available in the online Supplementary Material). 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 were from Brassicaceae. The New Zealand samples included one from a crocus and 15 from Brassica spp. 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 Brassica rapa cv. Hakata-suwari plants, but occasionally infected Brassica 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 many of the Australian and New Zealand isolates were [B] host-infecting type. Interestingly, the 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 included the two published sequences from Australia and New Zealand. The genomes of 29 Australian and New Zealand isolates were 9798 nt (excluding 5′-end 35 nt primer sequences), whereas three New Zealand isolates (NZ403, NZ403B and NZ415) were 2 nt shorter in the 3′ non-coding region (NCR; 207 nt). 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) were 1086, 1374, 1065, 177, 156, 1932, 159, 576, 729, 1551 and 864 nt, respectively. All of the motifs reported for different potyvirus-encoded proteins were found.

Genetic recombination in Australia and New Zealand

The genomic sequences of the 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 as 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 as 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 of TuMV genomes. 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 chequered boxes denote 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 Fig. S2). The recombination sites newly identified in the present study (non-bold font) or those identified in earlier studies (bold font) are listed separately.

**Fig. 2.** Recombination maps of TuMV genomes of the 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 chequered boxes denote 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 Fig. S2). The recombination sites newly identified in the present study (non-bold font) or those identified in earlier studies (bold font) are listed separately.
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 European 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 the 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) included 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 Nla-Pro and NB regions (see Fig. 2). Trees were estimated using 225, 214 and 226 non- and intralineage recombinant sequences, respectively. The relationships of isolates were inferred by maximum likelihood (ML) implemented in PhyML (Fig. S2). These partitioned relationships of isolates were inferred by maximum likelihood (ML) implemented in PhyML (Fig. S2). These partitioned regions and realigned using gaps. When a sequence is a recombinant with two ‘parents’, it is split into two regions and realigned using gaps. In this way, one haplotype diversity value was 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 NB* 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 NB* regions. The best-supported demographic models were of constant population 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 dataset. The impact of non-random sampling on the test is unknown.

The mean estimated substitution rates were $1.47 \times 10^{-3}$ substitutions per site per year for HC-Pro*, $1.35 \times 10^{-3}$ substitutions per site per year for P3* and $1.30 \times 10^{-3}$ substitutions per site per year for NB* regions (Table 1). These estimates are potentially inflated by the inclusion of transient polymorphisms that would normally be removed by purifying selection over longer time frames (Duchêne et al., 2014; Gibbs et al., 2010; Wertheim & Kosakovsky Pond, 2011). 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 NB* regions (Table 1, Figs 3 and S4). The relationships between Australian and New Zealand isolates and European isolates were also confirmed using ML trees (Fig. S5).

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...
New Zealand isolates were only from these three subgroups world-B2 and world-B3 subgroups because Australian and for each phylogenetic subgroup, focusing on the basal-B2, tries and from Europe to Australia and New Zealand. indicated that TuMV migrated between European coun-
tries and from Germany to Australia was supported 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 recombination events related to the basal-B parents occurred earlier.

Plausible routes of TuMV migration into Australia and New Zealand

We investigated the likely routes of TuMV migration into Australia and New Zealand using a Bayesian phylogeo-
graphical 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 indicated that TuMV migrated between European coun-
tries and from Europe to Australia and 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 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 the 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 the NIb* region (BF=68) and took place 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 BP (Fig. 4b, c, d). There was also significant support (BF ≥100) for migration between the neighbouring countries within Europe and East Asia (Fig. S7).

These results were confirmed by ML trees of the HC-Pro*, P3* and NIb* regions (Fig. S5). These showed which isolates were closest to the Australian and New Zealand

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Helper-component proteinase</th>
<th>Protein 3</th>
<th>Nuclear inclusion b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best-fit substitution model</td>
<td>GTR + I + Γ₄</td>
<td>GTR + I + Γ₄</td>
<td>GTR + I + Γ₄</td>
</tr>
<tr>
<td>Best-fit molecular-clock model</td>
<td>Relaxed uncorrelated exponential</td>
<td>Relaxed uncorrelated exponential</td>
<td>Relaxed uncorrelated exponential</td>
</tr>
<tr>
<td>Best-fit population growth model</td>
<td>Constant size</td>
<td>Constant size</td>
<td>Constant size</td>
</tr>
<tr>
<td>Sequence length (nt)</td>
<td>927</td>
<td>897</td>
<td>891</td>
</tr>
<tr>
<td>No. of sequences</td>
<td>180</td>
<td>186</td>
<td>182</td>
</tr>
<tr>
<td>Chain length (in millions)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>TMRCA (years)†</td>
<td>All isolates 610 (233–1156)†</td>
<td>806 (274–1630)</td>
<td>679 (205–1502)</td>
</tr>
<tr>
<td>Australia</td>
<td>80 (119–52) [n=12]§</td>
<td>44 (65–25) [n=13]</td>
<td>36 (51–19) [n=2]</td>
</tr>
<tr>
<td>World-B2 subgroup</td>
<td>14 (20–9) [n=2]</td>
<td>9 (29–5) [n=2]</td>
<td>NA [n=0]</td>
</tr>
<tr>
<td>World-B3 subgroup</td>
<td>27 (36–16) [n=2]</td>
<td>NA [n=1]</td>
<td>46 (67–28) [n=14]</td>
</tr>
<tr>
<td>New Zealand</td>
<td>14 (29–4) [n=3]</td>
<td>16 (28–4) [n=3]</td>
<td>22 (31–12) [n=3]</td>
</tr>
<tr>
<td>Basal-B2 subgroup</td>
<td>56 (83–30) [n=13]</td>
<td>16 (23–9) [n=10]</td>
<td>17 (24–11) [n=9]</td>
</tr>
<tr>
<td>World-B3 subgroup</td>
<td>NA [n=1]</td>
<td>NA [n=1]</td>
<td>68 (100–36) [n=5]</td>
</tr>
<tr>
<td>Substitution rate (nt per site per year)</td>
<td>1.47 × 10⁻³ (1.08 × 10⁻³⁻)</td>
<td>1.35 × 10⁻³ (9.50 × 10⁻⁴⁻)</td>
<td>1.30 × 10⁻³ (9.07 × 10⁻⁴⁻)</td>
</tr>
<tr>
<td>dN/dS</td>
<td>0.025</td>
<td>0.120</td>
<td>0.030</td>
</tr>
<tr>
<td>No. of variable sites</td>
<td>511</td>
<td>536</td>
<td>493</td>
</tr>
</tbody>
</table>

NA, Not available.

*Partial helper-component (HC-Pro*), Protein 3 (P3*) and nuclear inclusion b (NIb*) regions (see Methods).
†TMRCA, Time to the most recent common ancestor (years).
§95% Highest posterior density interval in parentheses.
∥The number of isolates in shown in square brackets.
∥∥Non-synonymous (dN) and synonymous (dS) substitution (dN/dS) ratios were calculated for three protein-encoding regions using the Pamilo–Bianchi–Li method.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Helper-component proteinase</th>
<th>Protein 3</th>
<th>Nuclear inclusion b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best-fit substitution model</td>
<td>GTR + I + Γ₄</td>
<td>GTR + I + Γ₄</td>
<td>GTR + I + Γ₄</td>
</tr>
<tr>
<td>Best-fit molecular-clock model</td>
<td>Relaxed uncorrelated exponential</td>
<td>Relaxed uncorrelated exponential</td>
<td>Relaxed uncorrelated exponential</td>
</tr>
<tr>
<td>Best-fit population growth model</td>
<td>Constant size</td>
<td>Constant size</td>
<td>Constant size</td>
</tr>
<tr>
<td>Sequence length (nt)</td>
<td>927</td>
<td>897</td>
<td>891</td>
</tr>
<tr>
<td>No. of sequences</td>
<td>180</td>
<td>186</td>
<td>182</td>
</tr>
<tr>
<td>Chain length (in millions)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>TMRCA (years)†</td>
<td>All isolates 610 (233–1156)†</td>
<td>806 (274–1630)</td>
<td>679 (205–1502)</td>
</tr>
<tr>
<td>Australia</td>
<td>80 (119–52) [n=12]§</td>
<td>44 (65–25) [n=13]</td>
<td>36 (51–19) [n=2]</td>
</tr>
<tr>
<td>World-B2 subgroup</td>
<td>14 (20–9) [n=2]</td>
<td>9 (29–5) [n=2]</td>
<td>NA [n=0]</td>
</tr>
<tr>
<td>World-B3 subgroup</td>
<td>27 (36–16) [n=2]</td>
<td>NA [n=1]</td>
<td>46 (67–28) [n=14]</td>
</tr>
<tr>
<td>New Zealand</td>
<td>14 (29–4) [n=3]</td>
<td>16 (28–4) [n=3]</td>
<td>22 (31–12) [n=3]</td>
</tr>
<tr>
<td>Basal-B2 subgroup</td>
<td>56 (83–30) [n=13]</td>
<td>16 (23–9) [n=10]</td>
<td>17 (24–11) [n=9]</td>
</tr>
<tr>
<td>World-B3 subgroup</td>
<td>NA [n=1]</td>
<td>NA [n=1]</td>
<td>68 (100–36) [n=5]</td>
</tr>
<tr>
<td>Substitution rate (nt per site per year)</td>
<td>1.47 × 10⁻³ (1.08 × 10⁻³⁻)</td>
<td>1.35 × 10⁻³ (9.50 × 10⁻⁴⁻)</td>
<td>1.30 × 10⁻³ (9.07 × 10⁻⁴⁻)</td>
</tr>
<tr>
<td>dN/dS</td>
<td>0.025</td>
<td>0.120</td>
<td>0.030</td>
</tr>
<tr>
<td>No. of variable sites</td>
<td>511</td>
<td>536</td>
<td>493</td>
</tr>
</tbody>
</table>
**Fig. 3.** Bayesian maximum-clade-credibility chronogram inferred from the protein-encoding region of TuMV. The tree was estimated 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 (HPD) intervals of estimates of node ages. The bar graph shows the root state posterior probabilities for each location (coloured bars). Grey bars show the probabilities obtained with 10 randomizations of the tip locations.
clusters, and were therefore likely to be from the populations that provided the migrants. For example, the ML analysis showed that all the basal-B2 gene populations from Australia and New Zealand probably came from Germany (Figs 4 and 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) and 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.*, 2008b) 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 probable 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 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.

<table>
<thead>
<tr>
<th>Recombination site*</th>
<th>Parent (5’×3’)</th>
<th>Recombinant type†</th>
<th>Recombination age (YBP)‡</th>
<th>Stem age (YBP)</th>
<th>Crown age (YBP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nt 818</td>
<td>World-B3 × basal-B2</td>
<td>Inter</td>
<td>50–10</td>
<td>50–18</td>
<td>23–10</td>
</tr>
<tr>
<td>nt 1080</td>
<td>World-B2 × basal-B2</td>
<td>Inter</td>
<td>54–13</td>
<td>54–21</td>
<td>50–13</td>
</tr>
<tr>
<td>nt 1341</td>
<td>World-B3 × basal-B2</td>
<td>Inter</td>
<td>48–21</td>
<td>48–25</td>
<td>40–21</td>
</tr>
<tr>
<td>nt 1851</td>
<td>Basal-B2 × basal-B2</td>
<td>Intra</td>
<td>35–9</td>
<td>35–14</td>
<td>22–9</td>
</tr>
<tr>
<td>nt 2742</td>
<td>World-B3 × basal-B2</td>
<td>Inter</td>
<td>38–14</td>
<td>38–14</td>
<td>NA</td>
</tr>
<tr>
<td>nt 3475</td>
<td>Basal-B2 × world-B3</td>
<td>Inter</td>
<td>51–22</td>
<td>51–26</td>
<td>49–22</td>
</tr>
<tr>
<td><strong>New Zealand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nt 5602</td>
<td>World-B3 × world-B3</td>
<td>Intra</td>
<td>49–20</td>
<td>49–37</td>
<td>46–20</td>
</tr>
<tr>
<td>nt 5665</td>
<td>World-B2 × world-B3</td>
<td>Intra</td>
<td>22–11</td>
<td>22–11</td>
<td>NA</td>
</tr>
<tr>
<td>nt 8071</td>
<td>World-B2 × world-B3</td>
<td>Intra</td>
<td>20–10</td>
<td>20–14</td>
<td>19–10</td>
</tr>
<tr>
<td><strong>Both countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nt 6132</td>
<td>World-B2 × world-B3</td>
<td>Intra</td>
<td>28–16</td>
<td>28–20</td>
<td>25–16</td>
</tr>
</tbody>
</table>

NA, Not available.

*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 (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).
About 35 years ago. Thus, there is evidence that a minimum of three TuMV isolates evaded quarantine and entered Australia, and that 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 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 (16000–18000 km), rather than from South East Asia (5000–9000 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., 2013a, b; 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 Australia and New Zealand were in horticultural materials. In contrast, the isolates closest to the world-B2 and world-B3 migrants were from Brassica spp.

Our analysis of TuMV provides a significant snapshot of the evolution and emergence of a highly pathogenic virus in association with human migration 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 Brassicaceae in all of Oceania.

All collected samples were tested by direct double-antibody sandwich 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 earlier studies, TuMV isolates were usually cloned by single-lesion isolations. Consequently, we found very few mismatches in the sequences between the overlapping reverse transcription-PCR (RT-PCR) products (Nguyen et al., 2013a). Moreover, cloning is required when attempting to analyse recombinant events and the genetic structure of populations. In the present study, all of the isolates were inoculated onto 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 Nicotiana 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 on to young plants of Brassicaceae plants (Nguyen et al., 2013b). Inoculated plants were kept for at least 4 weeks in a glasshouse at 25 °C.

METHODS

Virus isolates. The Brassica crop-producing areas of Australia and New Zealand were surveyed during the growing seasons of 1994–2011.

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 (Fuji & Nakamae, 1999; 2000), one of scallion mosaic virus (Chen et al., 2002), one of narcissus yellow stripe virus (Chen et al., 2006) and two of narcissus late season yellow virus (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 TRANSLIGN (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 of 8922 nt. The polyprotein sequences were then joined with aligned 5′ and 3′ NCR sequences of each isolate. This produced sequences of 9087 nt, excluding the 35 nt 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
Fig. 4. Plausible historical migration pathways of TuMV inferred using partial helper-component proteinase (HC-Pro\(^\ast\)), protein 3 (P3\(^\ast\)) and nuclear inclusion b (NIb\(^\ast\)) regions. Details of the regions are given in Methods. (a–c) Migration routes for Australia and New Zealand only are shown, and only when supported by a Bayes factor (BF) \(>10\). Only the migration pathways for basal-B2 (a), world-B2 (b) and world-B3 (c) isolates are shown. (d) For each subgroup migration, the BF and estimated age in years before present (YBP) are shown. The 95% highest posterior density 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 the basal-B2
subgroup, and only the BF values from the basal-B2 subgroup are listed. BF values can be interpreted as follows: 10 < BF < 30, strong support; 30 < B < 100, very strong support; and BF > 100, decisive support.

(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 (Weisser, 1998) and SISCAN v.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 > 10.0 × 10^-6, were rechecked using the original PHYLPRO v.1 and SISCAN v.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 of 9710 nt excluding the 35 nt. 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 NeighborNet method in SPLITSTREE (Huson & Bryant, 2006) and using ML in PhyML v.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+G). 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 v.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 v.6 (Tamura et al., 2013).

Estimation of substitution rates and divergence times. Bayesian phylogenetic analyses were performed in BEAST v.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 UK1 genome), P3 (nt 2591–3463) and Nb (nt 7208–8068). Recombinant sequences and some nucleotides from recombination ends were discarded from the three regions (see Fig. 2). We called these regions HC-Pro*, P3* and Nb*, respectively. The sampling times of the sequences were used to calibrate the molecular clock. Bayes factors were used to select the best-fitting molecular-clock model. The spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912. Chamberlain, E. E. (1936). Turnip-mosaic. A virus disease of crucifers. New Zealand J Agr Res 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

REFERENCES


