Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road

Shusuke Kawakubo[a], Fangluan Gao[b], Shifang Li[c,d], Zhongyang Tan[a,e], Ying-Kun Huang[f], Charith Raj Adkar-Purushothama[g], Chennappa Gurikar[h], Phoowanarth Maneechoat[i], Pissawan Chiemsombat[j], Seint San Aye[k], Naruto Furuya[l], Oleksiy Shevchenko[m], Josef Špak[n], Dijana Škorić[o], Simon Y. W. Ho[p], and Kazusato Ohshima[a,q,1]

[a]Laboratory of Plant Virology, Department of Biological Sciences, Faculty of Agriculture, Saga University, 840-8502 Saga, Japan; [b]Institute of Plant Virology, Fujian Agriculture and Forestry University, 350002 Fuzhou, China; [c]State Key Laboratory of Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, 100193 Beijing, China; [d]Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, 571101 Hainan, China; [e]College of Biology, Hunan University, 410082 Changsha, China; [f]Yunnan Key Laboratory of Sugarcane Genetic Improvement, Sugarcane Research Institute, Yunnan Academy of Agricultural Sciences, 661699 Kaiyuan, China; [g]RNA Group/Groupe ARN, Département de Biochimie, Faculté de médecine des sciences de la santé, Pavillon de Recherche Appliquée au Cancer, Université de Sherbrooke, Sherbrooke, QC J1E 23 4K8, Canada; [h]Department of Food Processing and Technology, Faculty of Life and Allied Health Sciences, MS Ramaiah University, 560 054 Bangalore, India; [i]Department of Agriculture, Ministry of Agriculture and Cooperatives, Bangkok 10900, Thailand; [j]Department of Plant Pathology, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, 73140 Nakhon Pathom, Thailand; [k]Department of Plant Pathology, Yezin Agricultural University Yezin, 15013, Myanmar; [l]Laboratory of Plant Pathology, Faculty of Agriculture, Kyushu University, Fukuoka 819-0395, Japan; [m]Virology Department, Educational and Scientific Center Institute of Biology and Medicine, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine; [n]Institute of Plant Molecular Biology, Biology Centre CAS, 37005 České Budějovice, Czech Republic; [o]Department of Biology, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia; [p]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia; and [q]The United Graduate School of Agricultural Sciences, Kagoshima University, 890-0065 Kagoshima, Japan

Plant pathogens have agricultural impacts on a global scale and resolving the timing and route of their spread can aid crop protection and inform control strategies. However, the evolutionary and phylogeographic history of plant pathogens in Eurasia remains largely unknown because of the difficulties in sampling across such a large landmass. Here, we show that turnip mosaic potyvirus (TuMV), a significant pathogen of brassica crops, spread from west to east across Eurasia from about the 17th century CE. We used a Bayesian phylogenetic approach to analyze 579 whole genome sequences and up to 713 partial sequences of TuMV, including 122 previously unknown genome sequences from isolates that we collected over the past five decades. Our phylogeographic and molecular clock analyses showed that TuMV isolates of the Asian-*Brassica/Raphanus* (BR) and basal-BR groups and world-*Brassica*3 (B3) subgroup spread from the center of emergence to the rest of Eurasia in relation to the host plants grown in each country. The migration pathways of TuMV have retraced some of the major historical trade arteries in Eurasia, a network that formed the Silk Road, and the regional variation of the virus is partly characterized by different type patterns of recombinants. Our study presents a complex and detailed picture of the timescale and major transmission routes of an important plant pathogen.

turnip mosaic potyvirus | brassica vegetables | Eurasia | Silk Road | molecular evolution

Eurasia is the largest landmass on Earth and has a rich demographic, cultural, and economic history. It was one of the birthplaces of agriculture and hosts the richest variety of crops, including vegetable crops (1–3). Eurasia is also the likely area of origin of many plant pathogens (4–6), but there have been few detailed studies of exactly when and where these pathogens emerged. Possibly owing to a lack of genetic sequence data, there also remains a poor understanding of the migration pathways of plant viruses and other plant pathogens in Eurasia.

Turnip mosaic potyvirus (TuMV) is probably the most widespread and important plant virus and occurs throughout subtropical and temperate regions (7). It mainly damages dicotyledonous domestic brassica crops, which are important food vegetables and were probably developed from wild *Brassica* plants by plant breeders during the expansion of agriculture. First described

from brassica crops in 1921 in the United States (8, 9), TuMV has been ranked behind only cucumber mosaic cucumovirus as the most important virus infecting field-grown vegetables (10, 11).

Brassica vegetables, the main host plants of TuMV in modern agriculture (7, 10), mostly originated in the Mediterranean and Western Eurasia (12, 13). These economically important plants are commonly known as cabbages or mustard plants and include turnip, cauliflower, and broccoli. A possible wild ancestor of cabbage was originally found in Western and Southern Europe (14). *Brassica* vegetables then spread to East Asia, including Japan and Korea, while the first record of the cultivation of Chinese cabbage dates from the 15th century CE. The edible

Significance

Our article presents the most comprehensive reconstruction of the evolutionary and phylogeographic history of a major plant pathogen of brassica vegetables in Eurasia. Sampling across such a large landmass poses considerable challenges, and our study attempts to describe the spatial and temporal patterns of migration for a plant pathogen on a large scale. Our phylogeographic and molecular clock analyses show that the migration pathways of turnip mosaic potyvirus retrace some of the historical trade arteries of the Silk Road. This study demonstrates how a comprehensive genetic analysis can provide a large-scale view of the epidemiology and human-mediated spread of a plant pathogen across centuries of evolutionary history.

EVOLUTION

www.manaraa.com

radish (*Raphanus sativus*) possibly originated from ancestral wild radish (*Raphanus raphanistrum*) in the Mediterranean region and was domesticated in Asia prior to Roman times (15).

Within the genus *Potyvirus* (family *Potyviridae*), TuMV is closely related to viruses from monocotyledonous narcissus, scallion, wild onion, and yam to form the TuMV phylogenetic group (16–19). Potyviruses infect a wide range of flowering plants (16, 19, 20) and are spread by aphids in a nonpersistent manner and sometimes in seeds and infected living plant materials. They have a genome that is a single-stranded, positive-sense RNA of ~10,000 nucleotides (nt). The genome has one major open reading frame (ORF), which is translated into one large polyprotein, and a small overlapping ORF, a "pretty interesting *Potyviridae* ORF" (21). The polyprotein is autocatalytically hydrolyzed into at least 10 mature proteins (19, 22).

In terms of its evolution and epidemiology, the potyvirus TuMV is among the best studied of the plant-infecting RNA viruses. We have shown that it probably originated from a virus of wild orchids in Europe (7, 23). While adapting to wild and domestic Brassicaceae plants, TuMV spread from its center of emergence in Southern Europe, Asia Minor, and the Middle East to other parts of the world, including East Asia (24–27), Oceania (28), and the Americas (7, 29). However, the evolution and epidemiology of this virus in Eurasia remain poorly studied, leaving considerable uncertainty about its past migration pathways and its evolutionary dynamics and timescale.

In this study, we collected TuMV isolates from locations throughout Eurasia, including countries neighboring the center of emergence. We examined the biological characteristics of viral isolates that had been sampled from these countries across half a century of growing seasons. By analyzing 579 whole genome sequences, we estimated the evolutionary rate and timescale of TuMV and inferred its phylodynamic and phylogeographic history. Our study presents one of the largest and most detailed evolutionary and epidemiological analyses of a plant pathogen, providing a comprehensive picture of the temporal and spatial spread of a major plant pathogen across Eurasia.

Results and Discussion

Genomic Dataset. To resolve the migration pathways of TuMV across Eurasia, we assembled a dataset that included samples from throughout Eastern Europe and Asia (Fig. 1). We surveyed the incidence of TuMV infection during the growing seasons of 1974 to 2014 from agricultural crops and wild plants in these regions. We collected more than 1,000 plant samples showing mosaic and vein clearing, with some samples kept in a deep freezer or as freeze-dried materials until use. All samples were examined for the presence of TuMV infection. From the 205 samples that were positive for TuMV infection, we selected 122 that represented a range of sampling locations and host plants. TuMV isolates were biologically cloned and then examined for host-infecting types (pathogenicity), adding to our previous collections from Western Europe, Oceania, the Middle East, and Asia (7, 27) (*SI Appendix*, Table S1). Most of the Eurasian isolates infected *Brassica juncea* cv. Hakarashina and *Brassica rapa* cv. Hakatasuwari plants. However, a few infected cabbages collected in Asia, whereas many of the Asian isolates infected radishes. The pathogenicities of many TuMV isolates were associated with their country or district of sampling. We sequenced whole genomes from 122 Eurasian isolates in this study. Among the 579 whole genome sequences analyzed here, 507 genomic sequences were determined in our previous studies (23, 27, 28, 30, 31) or newly determined in the present study. The numbers of sequences used in each analysis are summarized in *SI Appendix*, Table S2.

Phylogenetic Relationships and Genomic Recombinants. The phylogenetic tree inferred from nonrecombinant sequences showed six

groups of TuMV (Fig. 2): Orchis, basal-*Brassica* (B), Iranian, Asian-*Brassica*/*Raphanus* (BR), basal-BR, and world-B. Basal-B, Iranian, and world-B are divided further into basal-B1 and B2, Iranian 1 and 2, and world-B1, B2, and B3 subgroups, respectively. This classification has been reported previously (7, 23, 27) and is likely to be comprehensive, given the worldwide sampling in our collections of TuMV isolates.

We inferred a maximum-likelihood phylogenetic tree using the polyprotein (major ORF) sequences of 579 isolates with information on geographic origin, host-infecting type, and original host (*SI Appendix*, Fig. S1). Phylogenetic generalized least squares (32) revealed an association between geographical distribution and host plants ($P < 0.05$), suggesting that some of the spatial structure is a consequence of the distribution of wild Brassicaceae, wild *Brassica*, and domestic *Raphanus* plants (categorized in *SI Appendix*, Fig. S1). These results might be influenced to some degree by nonrandom or uneven sampling in our dataset, although our sample is reasonably comprehensive in capturing variation in geographic range and host plants.

Of the 579 whole genomic sequences analyzed in this study, 418 have already been identified as recombinants in our previous studies (23, 27, 28, 30, 31). We identified recombination sites throughout the genomes and found 19 recombination type patterns not previously reported, with 49 recombinants among the 122 Eurasian genomic sequences in this study (*SI Appendix*, Fig. S2). Our analysis of 39 Eurasian whole genome sequences not sequenced in our laboratory, obtained from GenBank, showed that these, mostly from South Korea and China, were nonrecombinants or recombinants of basal-BR parents (*SI Appendix*, Fig. S2). Therefore, we inferred the phylogenetic tree using the 153 nonrecombinant sequences after discarding recombinants identified in this study and in previous studies (Fig. 2) (23, 27, 28, 30, 31). Our recombination analyses showed that world-B2 nonrecombinants were newly found in Ukraine (Fig. 2 and *SI Appendix*, Fig. S2) in this study, and in the Czech Republic, Denmark, and Turkey, as found in an earlier study (27). On the other hand, we did not find additional new world-B1 nonrecombinants in this study, besides the isolates reported previously from Poland and Russia (Fig. 2) (27). The world-B3 nonrecombinants are widely found in Europe, East Asia, and Southeast Asia (Fig. 2 and *SI Appendix*, Fig. S2) (27). The Asian-BR nonrecombinants are found mostly in China and Turkey, whereas the basal-BR nonrecombinants are found in Italy, Iran, China, and South Korea (Fig. 2 and *SI Appendix*, Fig. S2) (27).

The interlineage recombinants (i.e., those with parents from different major group lineages) with parents from the Asian-BR, basal-BR, or world-B3 groups were mainly distributed in South, Southeast, and East Asia (*SI Appendix*, Fig. S2) (27, 31). In contrast, interlineage recombinants from world-B1 parents were only found in Ukraine and world-B2 parents are distributed in Eastern Europe and Asia. The previously unidentified pattern of intralineage recombinants (i.e., those with parents from same major group lineage) from world-B3 parents was found mostly in Myanmar (*SI Appendix*, Fig. S2). Recombination sites of world-B and Asian-BR/basal-BR group parents in the protein 1 coding region, which has been identified as a recombination hotspot in the TuMV genome (31), were also widely seen in the recombinants of Eurasian isolates. None of the nonrecombinant isolates from South, Southeast, and East Asian countries fell into the Orchis, basal-B, or Iranian groups or the world-B1 and world-B2 subgroups. A more detailed analysis of the global recombination type patterns of TuMV genomes will be reported in a future publication. Finally, we selected recombination-cold protein-coding regions to allow a greater number of recombination-free sequences to be used for molecular clock and phylogeographic analyses.
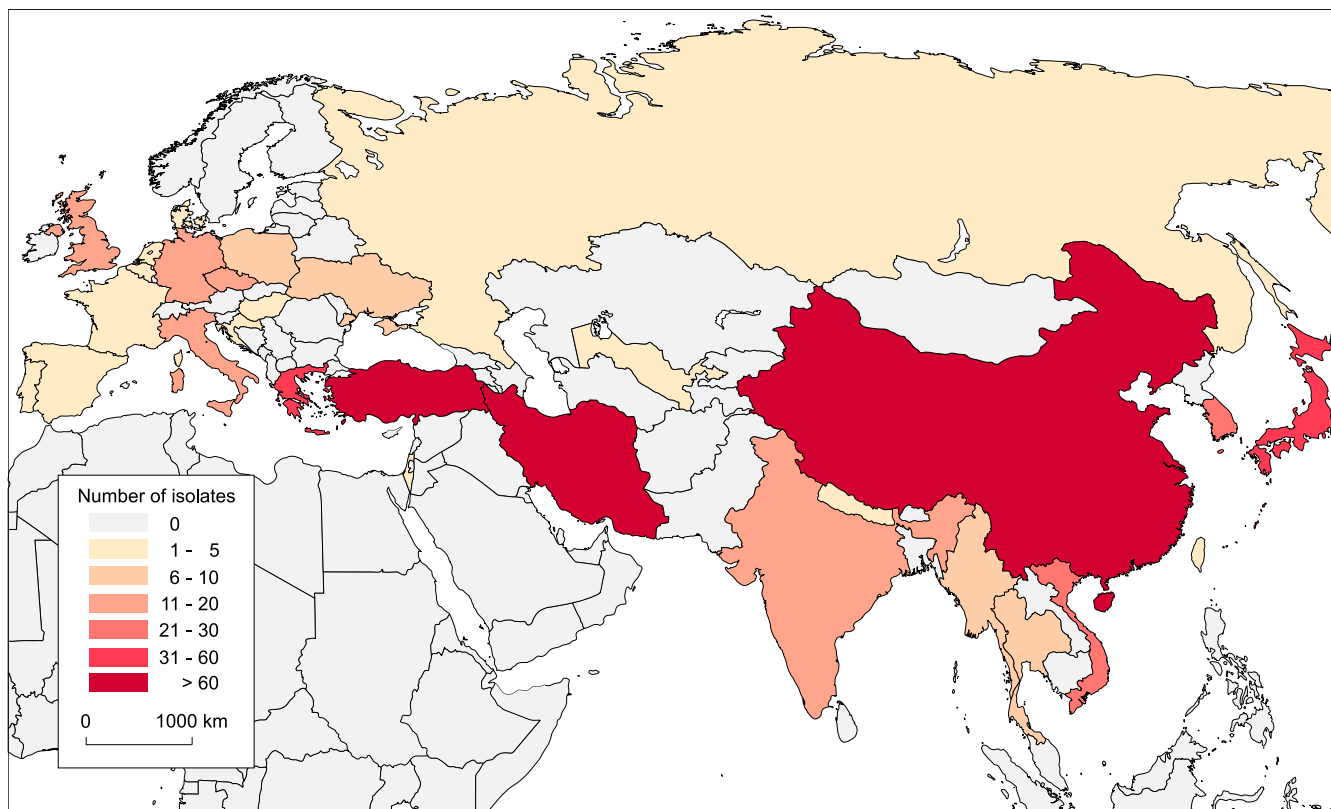
Fig. 1. Map of turnip mosaic virus sequence collections across Eurasia. A total of 520 isolate sequences from Eurasia were used from 579 world isolates, including 122 sequences determined in this study: China (mainland) (*n* = 63), Croatia (*n* = 2), Czech Republic (*n* = 13), India (*n* = 13), Myanmar (*n* = 9), Nepal (*n* = 3), Thailand (*n* = 9), Ukraine (*n* = 9), and Uzbekistan (*n* = 1). All of the isolates are listed in *SI Appendix*, Table S1. The map was obtained from https://365psd.com.
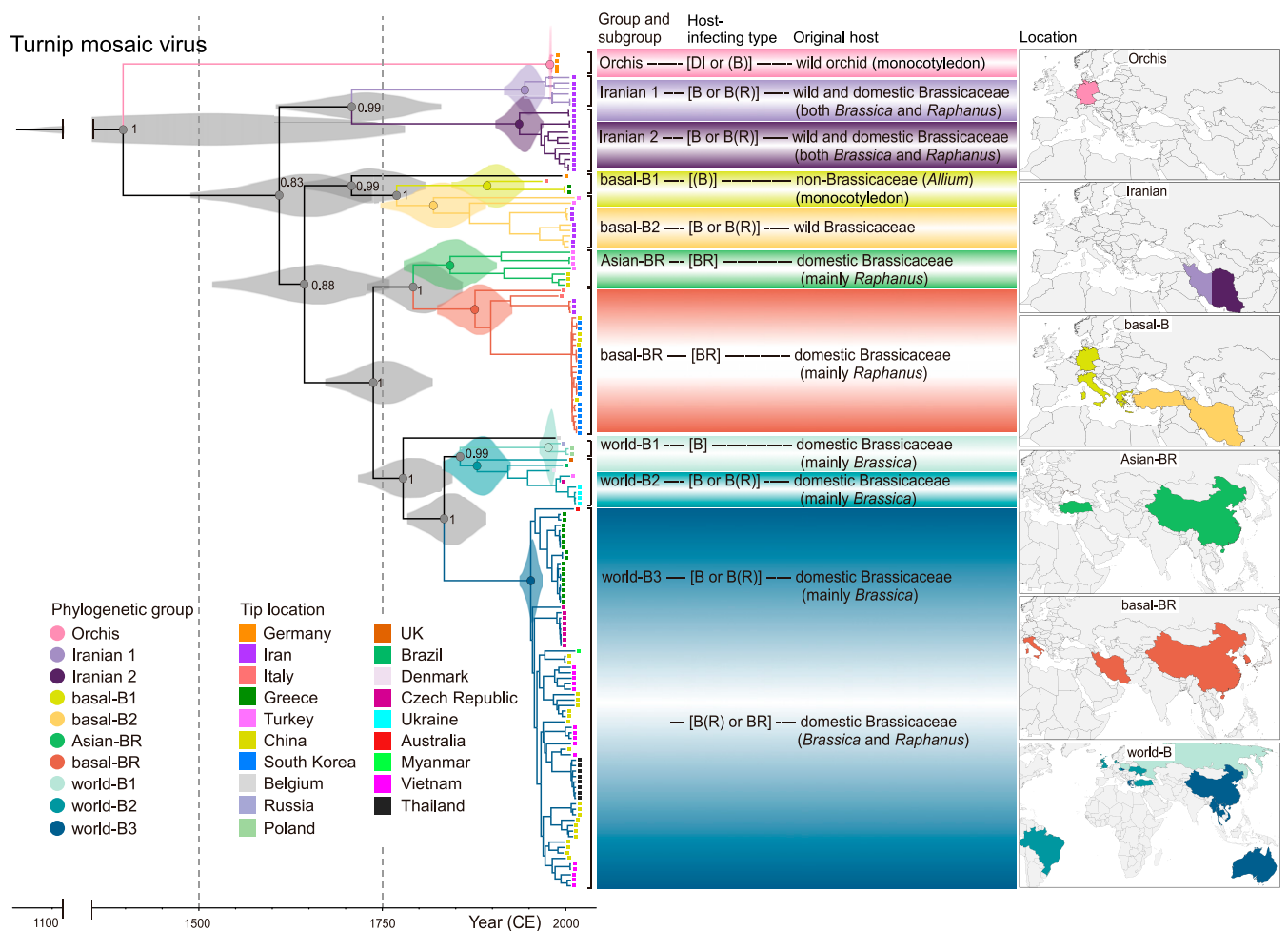
Origin and Emergence of TuMV.

Emergence and pathogenicity. Our results reveal a complex and dynamic evolutionary history of TuMV in Eurasia, characterized by numerous migration events and multiple shifts in pathogenicity (Fig. 2). The outgroup viruses of TuMV are Japanese yam mosaic virus (JYMV), scallion mosaic virus (ScaMV), wild onion symptomless virus, and narcissus viruses (18, 19), and those are isolated from monocotyledonous plants such as wild orchid. Although the closest extant relative of TuMV occurs in Europe, the center of emergence of this virus appears to have been in the Mediterranean region and the Middle East (27). The migration of TuMV in relation to its host plants and their pathogenicity from the center of emergence to the rest of Eurasia is largely unknown. The Orchis group consists of isolates from Germany and is the closest relative of the modern *Brassica*-infecting isolate lineages of TuMV (23). Orchis isolates generally do not infect both dicotyledonous *Brassica* and *Raphanus* plants or very rarely infect brassica plants latently (23). Furthermore, monocotyledonous *Allium* plants seem to be one of the intermediate hosts of Orchis and *Brassica*-infecting host plants. Therefore, TuMV probably first acquired the pathogenicity of *Allium* plants and then of wild and domestic *Brassica* plants, and then of *Raphanus* plants, rather than the reverse (Fig. 3). However, the collection of isolates that infect the intermediate hosts of Orchis and *Brassica*-infecting isolates, including *Allium*-infecting isolates, will help to resolve the history of pathogenicity in this virus.

The basal-B group is resolved as the sister lineage to the remaining *Brassica*-infecting groups but with low support (Fig. 2). This group includes the basal-B1 subgroup, which consists of isolates that originated from wild Brassicaceae

(*Alliaria* plants) and *Allium* plants from Germany, Italy, and Greece, and the basal-B2 subgroup, which consists of isolates that mostly originated from wild Brassicaceae in Turkey and Iran. Considering that the isolates of the basal-B1 subgroup originated from monocotyledonous *Allium* plants and have the greatest nucleotide diversity (mean 0.122, SD 0.036) among the phylogenetic groups, the basal-B1 subgroup is probably one of the oldest *Brassica*-infecting lineages. The Iranian isolates, which form a strongly supported monophyletic clade, were sometimes resolved as the sister group to *Brassica*-infecting groups in our analyses of partial genomic sequence (*SI Appendix*, Fig. S3). The lower nucleotide diversities of both Iranian 1 (mean 0.054, SD 0.021) and Iranian 2 (mean 0.045, SD 0.018) subgroups compared with those of basal-B1 and B2 (mean 0.098, SD 0.029) subgroups suggest that isolates of Iranian subgroups emerged more recently than those of basal-B subgroups. Although we had intended to collect Iranian isolates from various host plants and areas (*SI Appendix*, Table S1) (27), we have not collected many TuMV isolates from non-Brassicaceae plants that are known to be hosts in this country nor those from both Brassicaceae and non-Brassicaceae plants in neighboring countries experiencing armed conflicts. Therefore, the positions of Iranian groups can be evaluated more comprehensively with further sampling of TuMV in some of the countries surrounding the center of emergence. Such collections would also provide insights into how orchid-infecting lineages, which only infect brassica plants with difficulty, managed to jump the species barrier to become *Brassica*-infecting lineages (33, 34).

Evolutionary timescale. We focused here on the timescale of the spread of TuMV from the center of emergence to Eastern Eurasia for the three major groups as described above. This

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road

PNAS | 3 of 10
https://doi.org/10.1073/pnas.2021221118

www.manaraa.com

Subgroup | Group

Orchis

basal-B1 | basal-B
basal-B2

Iranian 1 | Iranian
Iranian 2

Asian-BR

basal-BR

world-B1
world-B2 | world-B
world-B3

Legend:
- Germany
- Italy
- Greece
- Turkey
- Iran
- China
- South Korea
- Belgium
- Russia
- Poland
- UK
- Brazil
- Denmark
- Czech Republic
- Ukraine
- Myanmar
- Vietnam
- Australia
- Thailand

0.1

Fig. 2. Maximum-likelihood tree of turnip mosaic virus isolates, inferred from 153 nonrecombinant sequences of the polyprotein (major ORF). Numbers at internal nodes indicate bootstrap percentages, based on 1,000 pseudoreplicates. The scale bar indicates 0.1 substitutions per site. The genomic sequences of isolates of NLSYV, NYSV, JYMV, and ScaMV were used as outgroup taxa. Details of the isolates are given in *SI Appendix*, Table S1.

builds on previous estimates of the timing of migrations in Europe (23), Oceania (28), and the Middle East (27), which showed a close correspondence of TuMV migrations with the establishment of agriculture and human movements (28).

Our analyses confirmed the presence of temporal signal in all of the datasets used in this study, based on a date-randomization test using clustered permutation (*SI Appendix*, Fig. S4) (35). We also found evidence of rate variation among lineages, with Bayes

Fig. 3. Overview of turnip mosaic virus evolution. Time-scaled maximum-clade-credibility tree inferred from partitioned sequences of the polyprotein (major ORF) of 153 nonrecombinant isolates. Phylogenetic groups and subgroups are labeled. Violin plots represent estimated posterior probability distributions for the ages of highlighted clades. Branch colors correspond to parental clade. Tips are colored by their collection location, as indicated by the color legend. The genomic sequence of the isolates of NLSYV, NYSV, JYMV, and ScaMV were used as outgroup taxa.

factors favoring an uncorrelated lognormal relaxed clock (36) over a strict clock (*SI Appendix*, Table S3). We firstly used 153 nonrecombinant polyprotein sequences to infer the evolutionary timescale (*SI Appendix*, Fig. S5 and Table S4). Analyses of partitioned, unpartitioned, and silent sites (27) yielded date estimates for the most recent common ancestor (MRCA) of 1395 CE (95% credible interval 1033 to 1717 CE), 1427 CE (95% credible interval 1089 to 1723 CE), and 1288 CE (95% credible interval 875 to 1685 CE), respectively. We also analyzed additional datasets in which the polyprotein sequences were gradually shortened or partially partitioned; the estimates from these datasets had overlapping 95% credible intervals, with no notable disparities.

We performed separate analyses of the partial helper-component proteinase (HC-Pro*; nt 1562 to 2494, numbers correspond to positions in the original UK1 [United Kingdom 1] genome) (37), protein 3 (P3*; nt 2591 to 3463), nuclear inclusion b (NIb*; nt 7208 to 8068), and complete coat protein (CP; nt 8759 to 9622) coding regions. These short, recombination-free sequences (HC-Pro*; 477, P3*; 487, NIb*; 498, and CP; 713, respectively) allowed us to use larger numbers of sequences compared with the longer sequence of the polyprotein (*SI Appendix*, Fig. S5 and Table S4). The mean date for the MRCA inferred from each protein-coding region ranged from 1126 CE

to 1634 CE, whereas those for only the Eurasian isolates ranged from 1152 CE to 1628 CE. All of these mean date estimates fall within the 95% credible intervals of the date estimates from the polyprotein sequences. As a further investigation of the robustness of our date estimates, we inferred the evolutionary rates and timescale from partial protein-coding sequences with or without discrete location states attached to the tips of the tree (*SI Appendix*, Fig. S5 and Table S4). We found that our estimates for the age of the MRCA differed when location states were integrated in the analysis. Such a discrepancy was also observed in a recent study of potato potyvirus Y (38).

Migration from the Center of Emergence to Eastern Eurasia.
The root of the phylogeny. To resolve the migration pathways of TuMV in Eurasia, we analyzed the sequence data using Bayesian phylogeographic analysis. The ancestral location at the root of the phylogeny was inferred using all isolates that were collected in Eurasia from HC-Pro*, P3*, NIb* (27, 28), and CP coding regions but not from polyprotein (see *Materials and Methods*). Because of their large genetic distances, which caused problems for rooting the tree, isolates of Orchis and Iranian groups were excluded from the datasets (*SI Appendix*, Fig. S6).

In all of our analyses, Italy was inferred to be the most probable location at the root of TuMV and at the base of each

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road

PNAS | 5 of 10
https://doi.org/10.1073/pnas.2021221118

www.manaraa.com

phylogenetic group, except for the world-B group (*SI Appendix*, Fig. S6). The inferred location of the common ancestor of the world-B group varied with the section of the genome being analyzed: Italy for HC-Pro*; Germany for P3* and CP; and the United Kingdom for NIb*. The location of the common ancestor of the world-B3 subgroup was inferred as Germany and introduced from Italy (1941 CE, 95% credible interval 1917 to 1961 CE) in our analysis of the HC-Pro* protein-coding region. In contrast, the location of the common ancestor was inferred as the United Kingdom and introduced from Germany (1935 CE, 95% credible interval 1916 to 1952 CE) in our analysis of the P3* protein-coding region. When the NIb* protein-coding region was analyzed, the location of the common ancestors of the world-B group and world-B3 subgroup were inferred as the United Kingdom, whereas they were both inferred as Germany when the CP region was used. Analyses of all four protein-coding regions yielded evidence of TuMV migrations from Mediterranean countries to Middle East and Asian countries (Fig. 4*B* and *SI Appendix*, Fig. S6). In particular, two of the introductions of TuMV into Turkey would have originated from Italy and Greece, whereas the other three introductions of TuMV into Japan, South Korea, and Northeast China would have occurred from Italy. The NIb* data supported an additional migration from Iran to India.

Migration pathways of each phylogenetic group. Our Bayesian phylogeographic analysis shows that the migration pathways of TuMV retraced some of the major historical trade arteries in Eurasia. This network of trade routes, known as the Silk Road, was important in the economic and cultural development of Eurasia. The Silk Road emerged in the Chinese Han Dynasty in the 2nd

century BCE and continued to be important until the 18th century CE (39–44).

Some nonrecombinants were found in the Asian-BR group from Turkey, indicating that the isolates in the Asian-BR group that infect both *Brassica* and *Raphanus* plants might have originated in Turkey or Italy (Fig. 2 and *SI Appendix*, Fig. S6). Asia Minor and the Mediterranean region, in the time of ancient Egypt, are believed to have been one of the sites of origin of wild radish (*R. raphanistrum*) (45). Indeed, we found many wild radish plants in the fields along the shore of the Aegean Sea during our collecting trips. A previous study found strong evidence of a past migration of TuMV from Turkey to Iran (27) and then to South Asia, where radish is one of the major crops and an important component of the cuisine. Our analyses yielded evidence of migrations from Nepal or Japan to Northeast China and six migrations from Northeast China to Central China, Southeast China, India, Iran, Japan, or Vietnam in the Asian-BR group. This suggests that Northeast China has served as a hub for the spread of the Asian-BR lineage to other parts of China and other Asian countries. Although we have no collections of TuMV from Pakistan and Afghanistan, our genetic analyses support a past migration from Iran to India (Fig. 4 and *SI Appendix*, Table S5). The Asian-BR group might have partially spread to the east in plant material carried through South and Southeast Eurasia, along a major network of maritime trade routes.

Isolates from the basal-BR group infect both *Brassica* and *Raphanus* plants. These isolates are characterized by their high virulence to *Raphanus* plants, with infected plants showing severe symptoms (24, 46). The basal-BR group was an epidemic in China, Japan, and South Korea after 2000 CE (25, 47). The
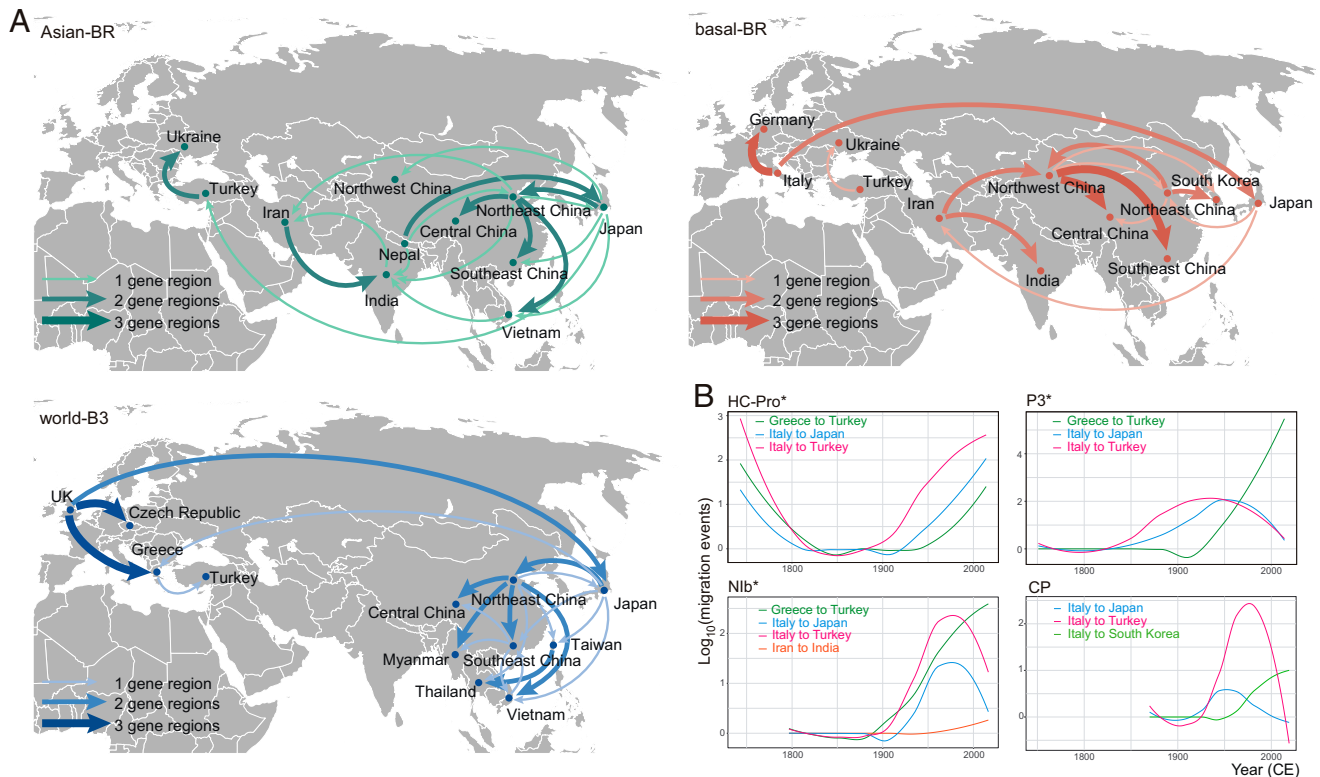
Fig. 4. Phylogeographic reconstruction of the spread of the turnip mosaic virus across Eurasia. (*A*) Supported spatial diffusion pathways of TuMV phylogenetic groups Asian-BR and basal-BR and subgroup world-B3. The total numbers of location state transitions were inferred from sequences of HC-Pro*, P3*, NIb*, and CP coding regions. The thickest arrows indicate diffusion pathways supported by three protein-coding regions; the arrows of intermediate thickness, supported by two protein-coding regions; and the thinnest arrows, supported by one protein-coding region. None of the diffusion pathways is supported by all four protein-coding regions. The map was obtained from https://365psd.com. (*B*) Migration events into Asia through time, inferred from each of the four protein-coding regions.

group possibly originated in Italy, given the phylogenetic distribution of Italian lineages within the group from the 18th to 20th century CE. For instance, the basal-BR group is most likely to have been transported to Japan in plant materials imported from Italy rather than by natural means such as flying aphids, as with how TuMV migrated from the United Kingdom and Germany to Australia and New Zealand (28). The migration of basal-BR isolates from Iran to the Xinjiang province in Northwest China (95% credible interval 1975 to 2009 CE) and then to Central China (95% credible interval 2002 to 2013 CE) followed the northern route of the ancient Silk Road. Our Bayesian analyses mostly identified recent migration pathways dated to the 20th century CE, but further sampling from other countries will fill the gaps in the inferred migration pathways.

The geographic spread of the world-B3 subgroup involved independent circulations on the Asian and European sides of Eurasia. These two groups differ in their pathogenicity, although both infected *Brassica* crops. Most European isolates did not infect *Raphanus* plants, whereas Asian isolates easily infect this plant species (Fig. 3 and *SI Appendix*, Table S1); the contrast in pathogenicity is probably the result of differences in the crops grown in Europe and Asia. Migration pathways are also inferred between the United Kingdom and Japan, the Czech Republic, and Greece. We found evidence of many migration pathways from Northeast China and Southeast China to Southeast Asian countries such as Myanmar, Thailand, and Vietnam. Northeast China was a center for the spread of world-B3 isolates in Asia, whereas the United Kingdom played a key role in the European spread. The world-B3 lineage probably acquired the pathogenicity on *Raphanus* during its journey from Western to Eastern Eurasia (34). Isolates from the world-B3 subgroup separated into an Asian cluster and a European cluster, as a result of an introduction from the United Kingdom to Japan in the 1900s. In all three phylogenetic (sub)groups, Northeast China has acted as an important area for the migration of TuMV.

The population dynamics of the Asian-BR, basal-BR, and world-B3 (sub)groups were each reconstructed from HC-Pro*, P3*, NIb*, and CP coding regions by the Bayesian skyline plot (*SI Appendix*, Fig. S7 and Table S3). We found a sudden increase in the population size of the basal-BR group around the 2000s, which was probably related to a rapid expansion after its introduction into China (24, 30, 48). The Asian-BR group and world-B3 subgroup spread widely in the early part of their divergence and experienced gradual increases in population size in the mid-1980s, and then both experienced population declines in the 2000s. These declines coincide with the rapid population expansion of the basal-BR group in East Asia, reflecting a change in the composition of viral lineages in this region.

Conclusions. Our detailed phylogeographic genomic study has revealed a complex picture of the timescale and major transmission routes of TuMV, showing that its spread followed the paths of some historical trade arteries in Eurasia. This adds to the growing evidence of the importance of the trade routes in facilitating the spread of pathogens across this vast landmass, including the circulation of the plague pathogen *Yersinia pestis* (49). Many crops are believed to have spread to East Asia via the Silk Road (50, 51). We found that TuMV was able to infect *Brassica* plants in the 17th century CE and spread to Eastern Eurasia over the past few centuries. Although our Bayesian analysis revealed a recent timescale of migration events (*SI Appendix*, Table S5), our date estimates are somewhat sensitive to the choice of phylogeographic model (*SI Appendix*, Fig. S5); it is possible that some of the important migrations of TuMV occurred during the later operation of the Silk Road. The results of our analyses suggest that the spread of TuMV was similarly mediated by human activity, given that this plant pathogen is probably spread primarily by aphids and would only travel across

long distances with difficulty. Our study shows the potential for genomic analyses of widely sampled virus isolates to provide detailed insights into the emergence and spread of important plant pathogens. Furthermore, this study reveals how the evolution of pathogens of agricultural crops can be shaped by historical trade routes. Finally, although less useful for RNA viruses, archival and archeological material could also offer a more precise view of the exact timing and mode of emergence of plant pathogens (52, 53).

Materials and Methods

Virus Isolates. During the growing seasons of 1974 to 2014, we surveyed the *Brassica* crop-producing areas in Eastern and Western Eurasia. All of the collected plant leaves were tested by double-antibody sandwich enzyme-linked immunosorbent assay (54) using the antiserum to TuMV virions (7). Details of the TuMV isolates, their place of origin, original host plant, year of collection, host-infecting type, and accession numbers are given in *SI Appendix*, Table S1 and Fig. S1.

Most of the isolates were sap-inoculated to *Chenopodium quinoa* plants using 0.01 M potassium phosphate buffer (PPB) (pH 7.0) and serially cloned through single lesions at least three times using chlorotic local lesions that appeared ~10 d after inoculation. The biological cloning step is important because TuMV isolates were often coinfected with cucumber mosaic cucumovirus and/or cauliflower mosaic caulimovirus (11, 55), and some plants contained a mixture of different TuMV isolates (7). However, some viruses from India and Myanmar were difficult to recover despite our several trials of inoculations onto *Nicotiana benthamiana*, turnip plants, and radish plants, so we omitted the biological cloning step using *C. quinoa* for these isolates. We carefully sequenced their genomes using three to four products amplified by RT-PCR. The overlapping regions of the products were 200 to 350 nt, and no mismatches were found between the regions. Hence, we avoided the possibility of artificial recombination events caused by different TuMV RNA molecules that might be detected in the whole genome sequences of biologically uncloned isolates.

Host Tests and Pathogenicity. Biologically purified TuMV isolates were propagated in *N. benthamiana* and *B. rapa* cv. Hakatasuwari (turnip). Plants infected systemically with each of the TuMV isolates were homogenized in 0.01 M PPB (pH 7.0), and the isolates were mechanically inoculated to young plants, as described previously (7). Inoculated plants were kept at 25 °C for at least 4 wk in a glasshouse.

Sequencing and Alignment. We determined the full genomic sequences of 122 TuMV isolates collected in Eastern Europe and Asia. The viral RNAs were extracted from TuMV-infected *N. benthamiana* or turnip leaves using Isogen (Nippon Gene). The RNAs were reverse transcribed by PrimeScript Moloney murine leukemia virus reverse transcriptase (Takara Bio) and amplified using high-fidelity Platinum *Pfx* DNA polymerase (Invitrogen). The products obtained by RT-PCR were separated by electrophoresis in agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen K. K.).

Sequences from each isolate were determined using three or four 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 200 to 350 nt. Each RT-PCR product was sequenced by primer walking in both directions using a BigDye Terminator version 3.1 Cycle Sequencing Ready Reaction kit (Life Technologies) and an Applied Biosystems 310 and 3130 Genetic Analyzer. Sequence data were assembled using BioEdit version 5.0.9 (56).

We assembled a dataset of 579 whole genome sequences (*SI Appendix*, Table S1), comprising the 122 whole sequences determined in this study and 457 published sequences from GenBank (retrieved in June 2017) (*SI Appendix*, Table S2). The genomic sequences of the isolates of narcissus late season yellows virus (NLSYV; accession numbers JQ326210, JX156421, and NC_023628), narcissus yellow stripe virus (NYSV; JQ395042, JQ911732, and NC_011541), JYMV (AB016500 and KJ701427), wild onion symptomless virus (NC_030391), and ScaMV (NC_003399) were used as outgroup taxa because they are members of the TuMV phylogenetic group.

The nucleotide sequences of the polyprotein-coding sequences were aligned by their amino acid translations using CLUSTAL_X2 (57) with TransAlign (58). The aligned nt were then reassembled to form whole genome sequences by adding the aligned 5′ and 3′ noncoding regions of RNA. This produced sequences of 9,681 nt that excluded the 35 nt that were used as primers for RT-PCR amplification.

EVOLUTION

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road

PNAS | 7 of 10
https://doi.org/10.1073/pnas.2021221118

www.manaraa.com

For convenience, we divided Eurasian isolates into five geographic regions: Eastern Europe (Croatia, the Czech Republic, Hungary, Poland, Russia, and Ukraine), Western Europe (Belgium, Denmark, France, Germany, Greece, Italy, Portugal, Spain, the Netherlands, and the United Kingdom), the Middle East (Iran, Israel, Turkey, and Uzbekistan), South (India and Nepal) and Southeast (Myanmar, Thailand, and Vietnam) Asia, and East Asia (mainland China, Japan, South Korea, and Taiwan). We further divided mainland China into four regions to trace TuMV migrations among the districts of China, taking into account the location of provinces from which we drew samples of TuMV: (*SI Appendix*, Fig. S8) Central China comprising Gansu, Shaanxi, and Sichuan provinces; Northeast China comprising Anhui, Beijing, Hebei, Jiangsu, Jilin, Liaoning, Shandong, and Zhejiang provinces; Northwest China comprising Xinjiang province; and Southeast China comprising Guangxi, Hunan, Jiangxi, and Yunnan provinces.

Recombination Analyses. Our earlier studies reported that approximately three-quarters of isolates from TuMV populations were recombinants (23, 31). To reconstruct the evolutionary history of the virus, we focus on non-recombinants and recombination-free genomic regions. Putative recombination breakpoints in all sequences were identified using Recombination Detecting Program (RDP) (59), GENECONV (60), BOOTSCAN (61), MAXCHI (62), CHIMAERA (63), and SISCAN (64) methods, implemented in the RDP4 package (65), and also the original SISCAN version 2 (64) program. These analyses were done using default settings for the different detection programs and a Bonferroni-corrected *P* value cutoff of 0.01. Each of the identified sites was examined individually, and phylogenetic approaches were used to verify the parent/donor assignments.

We tested for recombination in our dataset of 579 genome sequences. Having examined all sites with an associated *P* value of $<10^{-6}$ (i.e., the most likely recombination sites), we retained the intralineage recombinants and removed the interlineage recombinants. The identified recombination sites were treated as missing data in subsequent analyses. For convenience, we called these the "parental isolates" of the recombinants. Finally, TuMV sequences were also aligned without outgroup taxa, producing genome sequences of 9,693 nt that were concatenated sequences of aligned 5′ noncoding region, polyprotein, and 3′ noncoding region. We checked these for evidence of recombination using the programs described above. For further analyses of the evolutionary timescale and phylogeography of TuMV, we obtained sequences of the CP coding region from GenBank and tested for recombination using the same settings as described above. After we removed the isolates that had clear recombination sites, we discarded any isolates with unknown date or location of collection, leaving 713 world and 626 Eurasian isolates in the CP datasets (*SI Appendix*, Table S2).

Given that there were few nonrecombinants except from the center of emergence, we rely on using partial genomic sequences to resolve the migration pathways of the virus across Eurasia. However, we still have no collections of TuMV isolates from Central Asia, except for a single isolate from Uzbekistan; our attempts to contact plant pathologists in the region have been unsuccessful. Finally, we discarded the sequences that had recombination sites and used the sequences from four recombination-free protein-coding regions (HC-Pro*, P3*, NIb*, and CP) for our analyses of TuMV migration and timescale. The lengths of HC-Pro*, P3*, NIb*, and CP coding regions range from 849 to 873 nt.

Phylogenetic Analysis. We inferred the maximum-likelihood phylogeny of polyprotein, HC-Pro*, P3*, NIb*, and CP coding sequences using PhyML version 3 (66). The generalized time-reversible model with gamma-distributed among-site rate variation and a proportion of invariable sites was selected for all of the datasets using jModelTest version 2.1.7 (67). Branch support was estimated by bootstrapping with 1,000 pseudoreplicates. Nucleotide diversity was calculated for each phylogenetic (sub) group using DnaSP version 6.12 (68). We tested for correlations between spatial structure and different hosts of each sample using phylogenetic generalized least squares in *ape* (69) and *nlme* (70). We plotted the root-to-tip distances against the sampling times, using TempEst version 1.5.1 (71), to identify any problematic sequences resulting from mislabeling or contamination. These sequences were removed from the alignments.

Molecular Clock Analysis. We inferred the evolutionary timescale and substitution rate using a molecular clock calibrated by the sampling times of the sequences. To evaluate the temporal structure in the dataset, we conducted a date-randomization test with 10 clustered permutations (35) in the *TipDatingBeast* package (72). We assumed the presence of temporal structure when the 95% credible interval of the rate estimate from the original

dataset did not overlap with the 95% credible interval of any of the rate estimates from the date-randomized replicates. We also evaluated the degree of mutational saturation using the index of substitution saturation, I_{ss}, in DAMBE version 7.0.35 (73). Based on the analyses of the aligned HC-Pro*, P3*, NIb*, and CP coding sequences, the estimates of I_{ss} were four to five times lower than the critical value $I_{ss.c}$ for all datasets ($P < 0.05$). Therefore, there was little saturation across the sequences in the datasets of each of the four protein-coding regions. Analyses were first based on complete sequences of the polyprotein (nt 131 to 9,622, corresponding to the positions in the genome of the original TuMV-UK1 isolate).

We analyzed the datasets using Bayesian phylogenetic dating in BEAST version 1.8.4 (74), with the posterior distribution estimated by Markov chain Monte Carlo (MCMC) sampling. We ran at least two independent MCMC analyses for 10^8 iterations, sampling every 10^4 iterations. Tracer version 1.7.1 (75) was used to check for convergence between the two chains. We considered that sufficient sampling was indicated by effective sample sizes exceeding 200 for each parameter. The maximum clade credibility tree was computed in TreeAnnotator, a part of the BEAST package.

We used Bayes factors to compare two models of among-lineage rate variation (strict clock and uncorrelated lognormal relaxed clock) and three coalescent-based tree priors (constant size, exponential growth, and Bayesian skyline plot). Marginal likelihoods were estimated by path sampling and stepping-stone sampling (76). Bayes factors revealed support for the Bayesian skyline tree prior and the uncorrelated lognormal relaxed clock, which we applied to all of the datasets, except for the world-B3 subgroup of P3* region (*SI Appendix*, Table S3).

Phylogeographic Analysis. We used a Bayesian phylogeographical analysis (77) to infer the likely pathways of TuMV migration from the area of the center of emergence to Asian countries. This analysis was based on the recombination-free sequences of the four protein-coding regions: HC-Pro* (*n* = 472 to 477), P3* (*n* = 480 to 487), NIb* (*n* = 446 to 498), and CP (*n* = 626 to 713) (Fig. 4 and *SI Appendix*, Table S5). We also attempted an analysis using the 153 dated recombination-free polyprotein sequences but encountered serious problems with convergence and sufficient sampling. Better performance was observed in our analyses using HC-Pro*, P3*, NIb*, and CP coding region datasets, which also contained recombination-free sequences. We did not use the sequences that contained recombination sites between the four protein-coding regions or the sequences in the P1 coding region. Therefore, we investigated the pathways of migration from the area of the center of emergence to Asian countries using the HC-Pro*, P3*, NIb*, and CP coding regions (*SI Appendix*, Table S2).

We analyzed the data using BEAST (74), with each TuMV isolate assigned a discrete location state based on its country of sampling. Diffusions among the location states were modeled using an asymmetric model, with Bayesian stochastic search variable selection (77). We used the same settings as described for the Bayesian phylogenetic dating analyses above. The resulting diffusion rates were used to calculate Bayes factors using SpreaD3 (78). The migration pathways were interpreted as supported when the Bayes factor ≥ 3 and the node posterior probability ≥ 0.5.

To minimize the effect of sampling bias, we performed any necessary subsampling to keep the proportion of isolates from each location below 30%. To confirm that this filtering did not influence our results, we repeated the subsampling procedure at least three times. We also excluded any coded locations with only a single isolate. To determine whether the proportions of coded location states had a strong influence on the reconstruction of the location state at the root node, we compared the estimate with a randomized dataset of 10 replicates as a null distribution. We used a Python script to analyze the inferred load and direction of migration through time (available at https://github.com/admiralenola/globall4scripts). For this analysis, migration events were set to occur on the node based on the maximum clade credibility trees from our Bayesian phylogeographic analyses of sequence data from each of the four protein-coding regions.

Data Availability. All study data are included in the article and/or *SI Appendix*.

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread
along the former trade routes of the Silk Road

1. A. Snir *et al.*, The origin of cultivation and proto-weeds, long before Neolithic farming. *PLoS One* **10**, e0131422 (2015).
2. O. Bar-Yosef, "Multiple origins of agriculture in Eurasia and Africa" in *On Human Nature: Biology, Psychology, Ethics, Politics, and Religion*, M. Tibayrenc, F. J. Ayala, Eds. (Academic Press, New York, 2017), pp. 297–331.
3. G. R. Dixon, "Origins and diversity of *Brassica* and its relatives" in *Vegetable Brassicas and Related Crucifers*, G. R. Dixon, Ed. (CABI, Wallingford, 2006), chap. 1, pp. 1–33.
4. G. Agrios, *Plant Pathology* (Academic Press, ed. 5, 2005).
5. D. Saleh, J. Milazzo, H. Adreit, E. Fournier, D. Tharreau, South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. *New Phytol.* **201**, 1440–1456 (2014).
6. E. H. Stukenbrock, B. A. McDonald, The origins of plant pathogens in agroecosystems. *Annu. Rev. Phytopathol.* **46**, 75–100 (2008).
7. K. Ohshima *et al.*, Molecular evolution of Turnip mosaic virus: Evidence of host adaptation, genetic recombination and geographical spread. *J. Gen. Virol.* **83**, 1511–1521 (2002).
8. M. W. Gardner, J. B. Kendrick, Turnip mosaic. *J. Agric. Res.* **22**, 123–124 (1921).
9. E. S. Schultz, A transmissible mosaic disease of Chinese cabbage, mustard, and turnip. *J. Agric. Res.* **22**, 173–178 (1921).
10. J. A. Walsh, C. E. Jenner, Turnip mosaic virus and the quest for durable resistance. *Mol. Plant Pathol.* **3**, 289–300 (2002).
11. K. Ohshima *et al.*, Temporal analysis of reassortment and molecular evolution of Cucumber mosaic virus: Extra clues from its segmented genome. *Virology* **487**, 188–197 (2016).
12. I. C. Hedge, "A systematic and geographical survey of old world Cruciferae" in *The Biology and Chemistry of the Cruciferae*, J. G. Vaughan, A. J. Macleod, B. M. G. Jones, Eds. (Academic Press, London, 1976), pp. 1–46.
13. K. S. Talebi, T. Sajedi, M. Pourhashemi, *Forests of Iran: A Treasure from the Past, a Hope for the Future* (Springer Science & Business Media, 2013).
14. S. Pakash, X.-M. Wu, S. R. Bhat, "History, evolution, and domestication of *Brassica* crops" in *Plant Breeding Review*, J. Janick, Ed. (John Wiley & Sons, Inc., Hoboken, NJ, 2011), vol. 35, chap. 2, pp. 19–84.
15. Y. Kaneko, Y. Matsuzawa, "35–Radish: *Raphanus sativus* L" in *Genetic Improvement of Vegetable Crops*, G. Kalloo, B. O. Bergh, Eds. (Pergamon, Oxford), ed. 1, 1993), pp. 487–510.
16. A. Gibbs, K. Ohshima, Potyviruses and the digital revolution. *Annu. Rev. Phytopathol.* **48**, 205–223 (2010).
17. K. Ohshima, S. Korkmaz, S. Mitoma, R. Nomiyama, Y. Honda, First genome sequence of wild onion symptomless virus, a novel member of potyvirus in the turnip mosaic virus phylogenetic group. *Genome Announc.* **4**, e00851–e16 (2016).
18. K. Ohshima, S. Mitoma, A. J. Gibbs, The genetic diversity of narcissus viruses related to turnip mosaic virus blur arbitrary boundaries used to discriminate potyvirus species. *PLoS One* **13**, e0190511 (2018).
19. A. J. Gibbs, M. Hajizadeh, K. Ohshima, R. A. C. Jones, The potyviruses: An evolutionary synthesis is emerging. *Viruses* **12**, 132 (2020).
20. B. Moury, C. Desbiez, Host range evolution of potyviruses: A global phylogenetic analysis. *Viruses* **12**, 111 (2020).
21. B. Y. Chung, W. A. Miller, J. F. Atkins, A. E. Firth, An overlapping essential gene in the *Potyviridae*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5897–5902 (2008).
22. E. J. Lefkowitz *et al.*, Virus taxonomy: The database of the International committee on taxonomy of viruses (ICTV). *Nucleic Acids Res.* **46**, D708–D717 (2018).
23. H. D. Nguyen *et al.*, Turnip mosaic potyvirus probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. *PLoS One* **8**, e55336 (2013).
24. Y. Tomitaka, K. Ohshima, A phylogeographical study of the Turnip mosaic virus population in East Asia reveals an 'emergent' lineage in Japan. *Mol. Ecol.* **15**, 4437–4457 (2006).
25. Y. Tomitaka, T. Yamashita, K. Ohshima, The genetic structure of populations of Turnip mosaic virus in Kyushu and central Honshu, Japan. *J. Gen. Plant Pathol.* **73**, 197–208 (2007).
26. S. Korkmaz, Y. Tomitaka, S. Onder, K. Ohshima, Occurrence and molecular characterization of Turkish isolates of Turnip mosaic virus. *Plant Pathol.* **57**, 1155–1162 (2008).
27. R. Yasaka *et al.*, The timescale of emergence and spread of turnip mosaic potyvirus. *Sci. Rep.* **7**, 4240 (2017).
28. R. Yasaka *et al.*, Phylodynamic evidence of the migration of turnip mosaic potyvirus from Europe to Australia and New Zealand. *J. Gen. Virol.* **96**, 701–713 (2015).
29. K. Tomimura *et al.*, Comparisons of the genetic structure of populations of Turnip mosaic virus in West and East Eurasia. *Virology* **330**, 408–423 (2004).
30. K. Tomimura, A. J. Gibbs, C. E. Jenner, J. A. Walsh, K. Ohshima, 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 (2003).
31. K. Ohshima *et al.*, Patterns of recombination in turnip mosaic virus genomic sequences indicate hotspots of recombination. *J. Gen. Virol.* **88**, 298–315 (2007).
32. M. Bulmer, Use of the method of generalized least squares in reconstructing phylogenies from sequence data. *Mol. Biol. Evol.* **8**, 868–883 (1991).
33. Z. Tan *et al.*, Mutations in Turnip mosaic virus genomes that have adapted to *Raphanus sativus*. *J. Gen. Virol.* **86**, 501–510 (2005).
34. K. Ohshima, S. Akaishi, H. Kajiyama, R. Koga, A. J. Gibbs, Evolutionary trajectory of turnip mosaic virus populations adapting to a new host. *J. Gen. Virol.* **91**, 788–801 (2010).
35. S. Duchêne, D. Duchêne, E. C. Holmes, S. Y. W. Ho, The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906 (2015).
36. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
37. C. E. Jenner *et al.*, 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 (2000).
38. F. Gao, S. Kawakubo, S. Y. W. Ho, K. Ohshima, The evolutionary history and global spatio-temporal dynamics of potato virus Y. *Virus Evol.* **6**, veaa056 (2020).
39. P. Frankopan, *The Silk Roads: A New History of the World* (Bloomsbury Publishing, London, 2015).
40. V. Hansen, *The Silk Road: A New History* (Oxford University Press, Oxford, 2012).
41. F. H. Chen *et al.*, Special issue devoted to the civilization and environment along the ancient Silk Road preface. *Acta Geol. Sin.* **94**, I–VI (2020).
42. X. R. Liu, *The Silk Road in World History* (Oxford University Press, Oxford, 2010).
43. M. Fu, Revisiting ancient Silk roads: Origin and evolution [In Chinese version]. *Pacific Journal* **25**, 59–74 (2017).
44. E. Vadime, *The Silk Roads: Highways of Culture and Commerce* (Berghahn Books, 2000).
45. V. E. Rubatzky, M. Yamaguchi, *World Vegetables: Principles, Production and Nutritive Values* (Chapman & Hall, 1997).
46. F. Zhu *et al.*, Molecular characterization of the complete genome of three basal-BR isolates of Turnip mosaic virus infecting *Raphanus sativus* in China. *Int. J. Mol. Sci.* **17**, 888 (2016).
47. J. Gong *et al.*, Sequence variations among seventeen new radish isolates of Turnip mosaic virus showing differential pathogenicity and infectivity in *Nicotiana benthamiana*, *Brassica rapa*, and *Raphanus sativus*. *Phytopathology* **109**, 904–912 (2019).
48. X. Li *et al.*, The genetic structure of Turnip mosaic virus population reveals the rapid expansion of a new emergent lineage in China. *Virol. J.* **14**, 165 (2017).
49. M. A. Spyrou, K. I. Bos, A. Herbig, J. Krause, Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
50. H. Sanderson, "Roots and tubers" in *The Cultural History of Plants*, G. Prance, M. Nesbitt, Eds. (Routledge, New York, 2005), pp. 61–76.
51. J. Ye, The development and evolution of the Chinese cabbage in Ming and Qing Dynasties. *Agricultural History of China* **1**, 53–60 (1991).
52. S. Y. W. Ho, S. Duchêne, Dating the emergence of human pathogens. *Science* **368**, 1310–1311 (2020).
53. K. Yoshida *et al.*, Mining herbaria for plant pathogen genomes: Back to the future. *PLoS Pathog.* **10**, e1004028 (2014).
54. M. F. Clark, A. N. Adams, Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol.* **34**, 475–483 (1977).
55. R. Yasaka *et al.*, The temporal evolution and global spread of Cauliflower mosaic virus, a plant pararetrovirus. *PLoS One* **9**, e85641 (2014).
56. T. A. Hall, Bio Edit; a user-friendly biological sequence aliment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
57. M. A. Larkin *et al.*, Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
58. G. F. Weiller, TransAlign (Version 1.0, Research School of Biological Sciences, Canberra, Australia, 1999).
59. D. Martin, E. Rybicki, RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562–563 (2000).
60. S. A. Sawyer, GENECONV: A Computer Package for the Statistical Detection of Gene Conversion. Distributed by the author (Department of Mathematics, Washington University, St Louis, MO, 1999). Available at https://www.math.wustl.edu/~sawyer.
61. M. O. Salminen, J. K. Carr, D. S. Burke, F. E. McCutchan, Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11**, 1423–1425 (1995).
62. J. M. Smith, Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**, 126–129 (1992).
63. D. Posada, K. A. Crandall, Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13757–13762 (2001).
64. M. J. Gibbs, J. S. Armstrong, A. J. Gibbs, Sister-scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573–582 (2000).
65. D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

EVOLUTION

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road

PNAS | 9 of 10
https://doi.org/10.1073/pnas.2021221118

www.manaraa.com

66. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
67. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
68. J. Rozas *et al*., DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
69. E. Paradis, K. Schliep, Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
70. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, nlme: Linear and nonlinear mixed effects models. R package Version 3.1-137. https://cran.r-project.org/web/packages/nlme. Accessed 30 November 2020.
71. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
72. A. Rieux, C. E. Khatchikian, tipdatingbeast: An r package to assist the implementation of phylogenetic tip-dating tests using beast. *Mol. Ecol. Resour.* **17**, 608–613 (2017).
73. X. Xia, DAMBE6: New tools for microbial genomics, phylogenetics and molecular evolution. *J. Hered.* **108**, 431–437 (2017).
74. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
75. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
76. G. Baele *et al*., Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).
77. P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
78. F. Bielejec *et al*., SpreaD3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol. Biol. Evol.* **33**, 2167–2169 (2016).

Kawakubo et al.
Genomic analysis of the brassica pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of Eurasia