Whole-genome sequencing, trajectory tracking, and field investigation reveal origin and long-distance migration routes of wheat stripe rust in China

Yuxiang Li
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Jichen Dai
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Taixue Zhang
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Baotong Wang
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Conghao Wang
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Jiguang Zhang
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Qiang Yao
Academy of Agriculture and Forestry Science Qinghai University

Mingju Li
Agricultural Environment and Resource Institute, Yunnan Academy of Agricultural Sciences

Chengyun Li
Yunnan Agricultural University

Yueling Peng
Tibet Agricultural and Animal Husbandry College

Shiqin Cao
Gansu Academy of Agricultural Sciences

Gangming Zhan
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University

Fei Tao
Gansu Agricultural University
Haifeng Gao
Xinjiang Academy of Agricultural Sciences
Weili Huang
Xi'an Huang's Bio-technology Company Ltd.
Xiaojun Feng
Shaanxi Plant Protection Extension Station
Yingwen Bai
Baoji Plant Protection Extension Station
Zhuoma Qucuo
Tibet Agricultural and Animal Husbandry College
Hongsheng Shang
State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University
Chong Huang
National Agricultural Technology Extension and Service Center
Wancai Liu
National Agricultural Technology Extension and Service Center
Xiangming Xu
The National Institute of Agricultural Botany
Jiasui Zhan
Swedish University of Agricultural Sciences https://orcid.org/0000-0001-9250-0157
Xianming Chen
US Department of Agriculture, Agricultural Research Service https://orcid.org/0000-0002-9013-5974
Zhensheng Kang
Northwest A&F University
Xiaoping Hu (xphu@nwsuaf.edu.cn)
Northwest A&F University https://orcid.org/0000-0002-8155-7040

Article

Keywords:

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1427508/v1

Whole-genome sequencing, trajectory tracking, and field investigation reveal origin and long-distance migration routes of wheat stripe rust in China

Yuxiang Li[1#], Jichen Dai[1#], Taixue Zhang[1#], Baotong Wang[1#], Conghao Wang[1], Jiguang Zhang[1], Qiang Yao[2], Mingju Li[3], Chengyun Li[4], Yueling Peng[5], Shiqin Cao[6], Gangming Zhan[1], Fei Tao[7], Haifeng Gao[8], Weili Huang[9], Xiaojun Feng[10], Yingwen Bai[11], Zhuoma Qucuo[5], Hongsheng Shang[1], Chong Huang[12], Wancai Liu[12], Jiasui Zhan[13], Xiangming Xu[14], Xianming Chen[15], Zhensheng Kang[1*], and Xiaoping Hu[1*]

[1] State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University, Taicheng Road 3, Yangling, Shaanxi 712100, China.

[2] Key Laboratory of Agricultural Integrated Pest Management, Qinghai Province, Academy of Agriculture and Forestry Science, Qinghai University, Xining, Qinghai 810016, China.

[3] Yunnan Key Laboratory of Green Prevention and Control of Agricultural Transboundary Pests, Agricultural Environment and Resource Institute, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan 650205, China.

[4] State Key Laboratory for Conservation and Utilization of Bio-resources in Yunnan, Yunnan Agricultural University, Kunming, Yunnan 650201, China.

[5] Department of Plant Pathology, Tibet Agricultural and Animal Husbandry College, Linzhi, Tibet 860000, China.

[6] Wheat Research Institute, Gansu Academy of Agricultural Sciences, Lanzhou, Gansu 730070, China.

[7] Biocontrol Engineering Laboratory of Crop Diseases and Pests of Gansu Province, College of Plant Protection, Gansu Agricultural University, Lanzhou, Gansu 730070, China.

[8] Institute of Plant Protection, Xinjiang Academy of Agricultural Sciences, Urumqi, Xinjiang 830091, China.

[9] Xi'an Huang's Bio-technology Company Ltd., Xi'an, Shaanxi 710018, China.

[10] Shaanxi Plant Protection Extension Station, Xi'an, Shaanxi 710003, China.

[11] Baoji Plant Protection Extension Station, Baoji, Shaanxi 721001, China.

[12] National Agricultural Technology Extension and Service Center, Ministry of Agriculture, Beijing 100125, China.

[13] Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala 75007, Sweden.

[14] Pest & Pathogen Ecology, NIAB EMR, East Malling, West Malling, Kent ME19 6BJ, UK.

[15] Agricultural Research Service, United States Department of Agriculture and Department of Plant Pathology, Washington State University, Pullman, WA 99164-6430, USA.

# These authors contributed equally: Yuxiang Li, Jichen Dai, Taixue Zhang, and Baotong Wang

* These authors jointly supervised this work: Zhensheng Kang and Xiaoping Hu

e-mail: kangzs@nwsuaf.edu.cn; xphu@nwsuaf.edu.cn

Abstract

Understanding the origins and migration routes of phytopathogen inoculum is essential in predicting disease development and formulating control strategies. *Puccinia striiformis* f. sp. *tritici* (*Pst*), the causal agent of wheat stripe rust, is an airborne fungal pathogen threatening wheat production by long-distance migration. Due to large variation in geographic features, climatic conditions, and wheat production systems, inter-regional *Pst* dispersal routes in China remain largely unknown. In the present research, we sequenced 154 *Pst* isolates sampled from all the major wheat-growing regions in China to study the *Pst* population structure. Western Qinling Mountains, Himalayan region, and Guizhou Plateau were found to be centers of *Pst* origin in China. Combined with trajectory tracking and field disease surveys, long-distance *Pst* migration routes from individual origins were proposed. The present findings will improve current understanding of *Pst* origin and migration in China and emphasize the need for managing stripe rust at the national scale.

Introduction

Plant disease epidemics have changed the course of history in the civilization of human beings. With changes in climate, cropping system, and human activities, the increasing risk of plant disease outbreaks continues to threat the food security[1,2]. Long-distance airborne dispersal is an important route for numerous pathogens to invade and establish in new territories, especially for fungal and fungal-like pathogens, causing large-scale epidemics in a short time. Spores of these airborne pathogens can be transported passively for up to thousands of kilometers by wind to spread diseases[3-6]. Investigating the pathogen inoculum origins and dispersal routes is essential in designing effecting disease monitoring scheme, predicting disease development at large scale, and developing control measures at an appropriate scale, including deploying resistance cultivars, applying fungicides, and implementing quarantine measures[7,8].

Wheat stripe (yellow) rust, caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a devastating airborne disease threatening wheat production in more than 60 countries and is responsible for significant crop losses each year[9-11]. Major stripe rust epidemics have been reported in China[12], the USA[9], the UK[13], New Zealand[14], Australia[4], Pakistan[15], India[16], and many other countries[10,17], which resulted in grain losses up to tens of million tonnes annually. Long-distance dispersal of urediniospores is an important approach for *Pst* to spread across regions, countries or continents to cause epidemics[6,18,19]. Through analysis of *Pst* isolates from different continents, Himalayan and Mediterranean were putatively identified as centers of *Pst* origin and possible dispersal routes to other regions were proposed[20,21]. However, in most stripe rust epidemic regions, the *Pst* origin and dispersal route are still unclear.

Because of complicated topographies, diverse climatic conditions, and differing cropping systems in wheat-producing regions nationwide in China, wheat stripe rust has a unique epidemic characteristics in the region compared with other countries[22]. Interregional epidemics of stripe rust occur almost every year in China, causing an average of 1 million tonnes yield loss annually. Severe epidemics occurred in 1950, 1964, 1990, 2002, 2017, and 2020 resulted in accumulative yield loss nearly 16.0 million tonnes[23,24]. The Chinese National Rust Consortium has been researching epidemic development and management since its establishment in 1950. Wheat-planting areas within 97 to 135°E (longitude) and 22 to 53°N (latitude) in the inland are considered as the main *Pst* epidemic system, which can be further separated into 15

epidemic regions[22,25], whereas *Pst* in Xinjiang and Tibet is considered as independent epidemic systems from the main one in the inland because of the geographic separation by mountains and deserts[22]. Among these regions, Longnan, representing mountainous regions in southern Gansu and northwestern Sichuan in the western Qinling Mountain terrain, is considered as the main *Pst* oversummering region, which provides *Pst* inoculum for spreading to overwintering zones, and subsequently to the main wheat production region next spring. In general, long-distance dispersal of *Pst* occurs from west (oversummering) to east (overwintering) in autumn, and from south to north in spring[22]. Inferring *Pst* dispersal routes between different epidemic regions relies mainly on field surveillances of the onset and subsequent *Pst* development, but without clear dispersal routes at a fine scale. In recent decades, extensive studies have been carried out to understand *Pst* dispersal patterns using molecular markers. Most of these studies have focused on investigating and determining genetic relationship of *Pst* populations between Longnan and other regions[26-29]. Following the similar approach, some features of *Pst* epidemics in Xinjiang and Tibet have been studied[30-32]. Additionally, long-distance *Pst* dispersal has been detected in Sichuan, Yunnan, and Guizhou in southwestern China based on trajectory analysis[33]. However, all these studies have centered on the *Pst* dispersal at the regional scale rather at the national scale.

According to the earlier field surveys and population genetic studies, an outline of *Pst* dispersal routes in China has been proposed. Longnan is widely believed as the source of *Pst* inoculum in China, affecting epidemic development in Sichuan Basin and Huang-Huai-Hai areas. Urediniospores produced in Longnan after oversummering are dispersed to the Sichuan Basin to infect wheat plants in the autumn, and then *Pst* overwinters and produces more urediniospores there, resulting in spring epidemics in the same region. In addition, *Pst* inoculum from Longnan can also disperse to Shaanxi, Henan, and Hubei provinces to overwinter, providing the inoculum for the large area in Huang-Huai-Hai wheat-growing region, leading to the epidemic in the following spring. Therefore, stripe rust epidemic level in Longnan is directly related to that in Sichuan Basin and Huang-Huai-Hai wheat-growing region in most years[25-27]. However, some exceptions have been found in the wheat growing season in 2000-2001 and 2016-2017[34,35]. In 2001, stripe rust occurred intensively in the Sichuan Basin, Yunnan, and Guizhou, but only lightly in Longnan[34]. In 2017, severe stripe rust occurred in the Huang-Huai-Hai region, but *Pst* occurred at a low level in Longnan in the autumn 2016, the least since 2000[35]. As these observations differ from the main proposed *Pst* dispersal routes, further studies are necessary to understand *Pst* inoculum source and

dispersal routes at a fine scale in China.

Rapid development of genomic and transcriptomic sequencing technologies, bioinformatics, and aerodynamics modelling has enabled an integrated study on identifying *Pst* origin and dispersal routes. By employing a wealth of variants from high-throughput sequences, genetic diversities, and dispersal routes were studied in the *Pst* populations of the UK and Australia[36-38]. In the present study, we investigated *Pst* origins and dispersal routes in China using population genetic analysis, trajectory cluster calculation, and field survey. Through the analysis, we identified the Western Qinling Mountain, Himalayan region, and Guizhou Plateau as three centers of *Pst* in China, from which *Pst* is spread to the main wheat-growing regions nationwide. We specifically investigated the dispersal routes derived from these centers and assessed contributions of each origin to the nationwide *Pst* epidemics.

Results

Sampling, sequencing, and variation detection

Samples were collected from wheat stripe rust epidemic areas in eleven provinces/autonomous regions across China (Supplementary Table 1, Fig. 1a). After multiplication of urediniospores in greenhouse, *Pst* DNA from individual samples was extracted and sequenced. After filtering, a total of 893.54 Gb clean reads were obtained from 154 isolates (Supplementary Table 2). Quality and completeness of selected *Pst* genomes (see methods) were evaluated to select the appropriate reference genome (Supplementary Table 3, Supplementary Table 4). Considering the number of scaffolds, N50 scaffold length, and percentage of complete BUSCOs, DK09_11 was selected as the reference genome. A total of more than 1.83 million variants were generated from alignments with DK09_11. The mapping rates ranged from 58.36% to 95.65% for individual isolates with an average of 88.95% (Supplementary Table 5). After further quality filtering, ~840,000 high-quality SNPs were obtained. These SNPs were primarily located in intergenic, downstream, and upstream regions, and their impact on gene functions were mainly caused by missense, start and stop codon gained and lost, and splice donor/acceptor variant (Supplementary Table 6).

Phylogeny and population structure of *Pst*

After further selection based on linkage disequilibrium analysis to reduce ascertainment bias, a total of 39,453 SNPs were used to construct phylogenetic trees.

Based on the maximum likelihood (ML), two lineages (Cluster 1 and Cluster 2) were identified (Fig. 1b). Cluster 1 containing isolates only from Gansu, Tibet, and Guizhou was close to the root of phylogeny, whereas Cluster 2 comprised isolates from all the regions. This was supported by Principal component analysis (PCA); the first two PCs, explaining 27.4% and 12.5% of the total variance, indicated that Cluster 1 and Cluster 2 were separate groups (Fig. 1c). Two-cluster population structure was also supported by admixture analysis (Fig. 1d).

To understand the relationships among isolates within each cluster, the two clades from the phylogeny were analyzed separately. Cluster 1 was further partitioned into three sub-groups, containing isolates from Gansu (Cluster 1.1), Tibet (Cluster 1.2), and Guizhou (Cluster 1.3) as indicated by bootstrap values (Fig. 2a), which was supported by both PCA (Fig. 2b) and admixture analysis[39] (Fig. 2c). Likewise, Cluster 2 was further partitioned into four sub-groups using an unrooted phylogenetic tree (Fig. 3a), also supported by PCA (Fig. 3b) and admixture analysis (Fig. 3c). Samples in Cluster 2.2 were mainly from Sichuan and Yunnan in the southwest of China. Clusters 2.3 and 2.4 comprised isolates mostly from the winter-spring wheat area, including Qinghai, Xinjiang, and Tibet, located in the western China (Fig. 1a, Fig. 3d). As Cluster 2.1 comprised isolates from multiple locations, it was further divided into three sub-populations, including isolates from Yunnan and Guizhou (Cluster 2.1.1), and isolates mainly from Hubei, Henan, Shaanxi, Ningxia, Qinghai, and Sichuan (Cluster 2.1.2 and Cluster 2.1.3) (Supplementary Fig. 1).

Centers of *Pst* inoculum origin in China

Isolates from Gansu, Tibet, and Guizhou in Cluster 1 were closely related to the root of the phylogenetic tree, suggesting they were the likely origins of *Pst* in China (Fig. 1b). To explore the genetic diversity in all clusters, nucleotide diversity indices (π and θ_ω) were calculated. Cluster 1 had significantly ($P < 0.001$) higher diversity indices (π = 0.0054, θ_ω = 0.0025) than Cluster 2 (π = 0.0027, θ_ω = 0.0022). The highest genetic diversity was found in Gansu, followed by Tibet and Guizhou in Cluster 1, and the lowest was in the spring and winter wheat-growing area in Cluster 2.3 and Cluster 2.4 (Fig. 1b, Fig. 4a). The observed heterozygosity (*Ho*) was significantly higher (*P* adjusted < 0.05) than the expected heterozygosity (*He*) in Cluster 1.1, Cluster 1.2, Cluster 1.3, and Cluster 2.1 (Fig. 4b). Neutrality test with Tajima's *D* further confirmed the high heterozygosity of *Pst* in Clusters 1.1, Cluster 1.2, and Cluster 1.3 (Supplementary Fig. 2). Based on the genetic diversity, heterozygosity, phylogeny, and

geographic location, Gansu, Tibet, and Guizhou representing western Qinling Mountains, Himalayan region, and Guizhou Plateau, respectively, could be considered as primary origins of *Pst* in China, from which *Pst* is dispersed to other wheat-growing areas.

Genetic exchanges of *Pst*

To identify migration routes of *Pst* from inoculum sources, pairwise genetic differentiations (F_{ST}) was determined between possible centers of origin in Cluster 1 and sub-populations in Cluster 2. *Pst* isolates from Sichuan-Yunnan in Cluster 2.2 were identified to be the hotspot area where genetic exchanges with Gansu in Cluster 1.1 (F_{ST} = 0.233), Tibet in Cluster 1.2 (F_{ST} = 0.278), and Guizhou in Cluster 1.3 (F_{ST} = 0.282) were inferred to have taken place (Supplementary Table 7, Supplementary Fig. 3). Paired genetic differentiations identified the closest relationship between Cluster 2.2 and Cluster 2.1 (F_{ST} = 0.065). As Cluster 2.1 comprised isolates from multiple locations, detailed analysis was carried out for each subgroup in Cluster 2.1 (Supplementary Table 7). When compared with Sichuan-Yunnan in Cluster 2.2, lowest F_{ST} value was for Cluster 2.1.3, followed by Cluster 2.1.2 and Cluster 2.1.1 (Supplementary Table 8, Supplementary Fig. 4).

Genetic introgression analysis was performed to estimate the genetic relationship among different clusters. Thirty-five four-taxon topologies were acquired from phylogenetic analysis with *D* statistics. Significant genetic introgression between P2 and P3 (see methods) were identified in the ten topologies (*P* adjusted < 0.001). Among all the significant topologies, genetic introgressions were mainly identified between Cluster 1.1, Cluster 1.2, and Cluster 1.3 (P3) with Cluster 2.2 (P2). Extensive genetic introgression was indicated between Cluster 2.1 and Cluster 2.2 (Table 1, Fig. 5a, Fig. 5b). Consistent with genetic differentiation, *D* statistics suggested *Pst* in the western Qinling Mountains, Himalayan region, and Guizhou Plateau were mainly dispersed to the Sichuan-Yunnan region, and then exchanged frequently with multiple locations in Cluster 2.1 (Fig. 5a, Fig. 5b). In agreement with the phylogenetic trees constructed with RAxML, a parallel phylogeny was developed with TreeMix, suggesting a single migration event (m = 1) between Cluster 1.1 and Cluster 2.2 (Fig. 5c, Supplementary Fig. 5).

Trajectory analysis of *Pst* migration routes

To determine the spread of *Pst* from the centers of origin, the air trajectory of *Pst* in Longnan, Himalayan region, and Guizhou Plateau were analyzed. *Pst* in Longnan were mainly dispersed to the western and eastern Gansu, Sichuan Basin, and southern Ningxia at 100-3000 m above ground level (AGL, Supplementary Fig. 6). At 1000-3000 m AGL, *Pst* from Tianshui in Longnan were mostly dispersed to Pingliang in eastern Gansu with the average trajectory frequency (ATF) of 23%, Linxia in western Gansu (ATF = 17%), Guyuan in southern Ningxia (ATF = 17%), and Mianyang in the Sichuan Basin (ATF = 12%) (Supplementary Fig. 6). Based on the distribution of wheat-planting areas, Sichuan Basin and eastern Gansu were regarded as the major potential *Pst* epidemic regions threatened by the inoculum from Longnan in the autumn (Fig. 6a). Further trajectory analysis in autumn elaborated *Pst* inoculums in eastern Gansu were responsible for the autumn seedling infection in western Shaanxi (ATF = 40%), southwestern Shaanxi (ATF = 29%), central Shaanxi (ATF = 17%), Sichuan Basin (ATF = 15%), Henan and Hubei (ATF = 5%) at 1000-3000 m AGL (Supplementary Fig. 6).

Inoculums from Linzhi of Tibet located in the Himalayan region were mainly responsible for disease incidence in Tibet and Sichuan Basin (ATF ≥ 16%). In addition to the Sichuan Basin, another possible dispersal route was to southern Qinghai and Longnan, but with low frequencies (ATF = 3% - 5%) (Fig. 6b, Supplementary Fig. 7). Trajectory tracking analysis suggested an important dispersal route from western Xinjiang to eastern Qinghai (ATF = 12%) in April and May at 3000-5000 m AGL (Supplementary Fig. 8), implying that *Pst* in Xinjiang and Tibet could both affect *Pst* development in Qinghai. Bijie and Liupanshui are the major stripe rust epidemic regions in Guizhou Plateau. Average trajectory distribution of the two regions mainly covered areas in the Sichuan Basin in autumn and spring with the ATF of 23%. In addition, *Pst* dispersal from Bijie and Liupanshui to Hubei was also possible with 5% ATF in autumn and 12% in spring (Fig. 6c, Supplementary Fig. 9). As the intersection of *Pst* trajectories from three centers of origin, *Pst* inoculums in the Sichuan Basin were mainly circulated inside the basin (ATF > 50%) and dispersed to southwestern Shaanxi (ATF = 12%) and western Shaanxi (ATF = 7%) in spring at 100-3000 m AGL (Fig. 6d, Supplementary Fig. 10).

To confirm and trace sources of *Pst* from the overwintering regions in the Sichuan Basin, Henan, and Hubei, backward trajectory analyses were performed in each representative location. Backward trajectory aggregation suggested that *Pst* inoculum

in the Sichuan Basin resulted most likely from Longnan and eastern Gansu, and those in Henan and Hubei from eastern Gansu and central Shaanxi (Supplementary Fig. 11).

Field investigations

As the *Pst* epidemics caused yield losses of 1.40 and 1.65 million tonnes in 2017[23] and 2020, the onsets of *Pst* occurrences during the 2016-2017 and 2019-2020 cropping seasons were investigated. In 2016-2017, the start dates for stripe rust on wheat seedlings were in Baoji and Aba in eastern Shaanxi and Longnan in the early November, followed by the *Pst* overwintering regions in the Sichuan Basin, Henan, and Hubei. Stripe rust was observed in Yunnan and Gansu in late December. As *Pst* can both oversummer and overwinter in these locations, it is not possible to distinguish whether the disease resulted from local inoculum or from other regions. The onset dates of stripe rust in Anhui, Jiangsu, Shanxi, Shandong, and Hebei were in the period of March to April the following spring (Supplementary Table 9); stripe rust in these regions arose from infection by external inoculum. In general, the onsets of disease occurrence in the 2019-2020 crop season were earlier than in the 2016-2017 season, but the temporal sequence of starting dates in different regions was essentially the same (Supplementary Table 10). Although field surveillances might insufficiently accurate, the basic dispersal routes could be inferred from the temporal sequence of disease occurrence in different regions.

Migration routes of *Pst* in China

Based on the population genetic structure, air trajectory, and temporal sequence of disease occurrence in different regions, we propose the main dispersal routes of *Pst* in China (Fig. 7). Longnan is deduced as an important origin and inoculum source of *Pst* in China, where urediniospores could both oversummer and overwinter. In autumn, a major dispersal route from Longnan is to the Sichuan Basin and eastern Gansu. Once in eastern Gansu, *Pst* can then proliferate and disperse mainly to central Shaanxi, Sichuan Basin, Henan and Hubei in the same period (autumn), where urediniospore multiplication may take place in winter. In the following spring, *Pst* from the overwintering regions could be further spread both northward and eastward to Shanxi, Hebei, Anhui, Jiangsu, and Shandong to cause epidemics. Tibet is another center where *Pst* can survive all year-round. As stripe rust occurs in Tibet mostly from April to June, *Pst* urediniospore inoculum from Tibet is likely to affect the disease mainly in northern Sichuan Basin, where wheat harvest is in early June. Urediniospores from Tibet could also contribute to stripe rust in Qinghai and Longnan, where winter and

spring wheat crops are harvested from July to October. Guizhou Plateau, located in southwestern China, can be considered as another source of inoculum where stripe rust occurs in all wheat-growing seasons. After oversummering, urediniospore inoculum in Guizhou Plateau may be dispersed mostly to the Sichuan Basin in autumn, where urediniospores could multiply in winter. Additionally, exchanges of urediniospores between central Guizhou and Hubei can take place in autumn and spring. As a recipient of urediniospores from the three centers, urediniospores produced in the Sichuan Basin are likely to be mainly responsible for the epidemics in the basin, even though urediniospores could be spread from the basin to southwestern regions (Fig. 7).

Discussion

As a distinct epidemiological region of stripe rust, the exact *Pst* origins and dispersal routes in China remain largely elusive. Recently, much intensive research has been conducted to investigate *Pst* dispersal patterns in China, but focusing on studying relationships of *Pst* populations among a few regions with a limited number of molecular markers[26-28,40-42]. Analyses based on a limited of markers may not provide sufficient resolution to elucidate *Pst* origins and dispersal routes because of the rapid evolution and large-scale dispersal of *Pst*[36,37]. In the present study, we studied *Pst* origins and possible dispersal routes among all major wheat production regions in China based on population genetic analysis of abundant SNP markers, air trajectories, and field monitoring. We have identified the western Qinling Mountains, Himalayan region, and Guizhou Plateau as three main centers of *Pst* origins and proposed major *Pst* dispersal routes in China.

Longnan in the western Qinling Mountains has long been regarded as the *Pst* inoculum source responsible for the stripe rust epidemics throughout China because of the high diversity of *Pst* pathotypes and genotypes in the region[24,27,40,41]. We confirmed that this region is a major *Pst* origin and inoculum source in the present study. We identified two main *Pst* dispersal routes from Longnan to the Sichuan Basin and eastern Gansu on the basis of population genetic differentiation analysis, air trajectories, and field surveys, consistent with previous studies[26,27]. Recently, we also proposed Tibet in the Himalayan region as another possible center of *Pst* origin based on the *Pst* race and genotypic compositions[31,32], which is supported by the present results. Furthermore, the present results also suggested that there might a limited dispersal between *Pst* in Tibet and other regions, especially Sichuan Basin, which agrees with previous results

on *Pst* pathotypes[31]. Previous hypothesis of Guizhou as a possible *Pst* inoculum source is solely based on air trajectory analysis without molecular evidence[33]. This hypothesis is supported by the present research based on both population genetic analysis and air trajectory modelling. As this region may provide inoculum for the Sichuan Basin and Hubei, key regions for inoculum multiplication during overwintering, it should not be neglected in the national stripe rust monitoring and management programme. *Pst* populations in the three main centers have high diversity and heterozygosity, which are mainly caused by mutation, sexual recombination, and somatic recombination[43], making them as the most important regions for new *Pst* races to evolve. Since many *Berberis* spp. susceptible to *Pst* grow in Longnan[44], Himalaya region[45] and Guizhou Plateau[46] as well as natural *Pst* infection occurring on alternate host plants[19,66,67], sexual recombination may have contributed to the rapid evolution and high diversity of *Pst* in these three regions.

Pst urediniospores can be dispersed for up to thousands of kilometers[6,18,19], resulting in recurrent exchanges of spores among different regions. Continual interchanges of *Pst* inoculum may result in some decoupling of phylogenetic clusters from geographic regions. In the present study, *Pst* isolates from Gansu, Tibet, and Guizhou were grouped in Cluster 1, as well as in Cluster 2. Samples from Gansu were identified in the same cluster with those from Qinghai, Tibet, Xinjiang, Ningxia, and Shaanxi, indicating exchanges of *Pst* inoculum in the vast northwestern region of China. These results are consistent with previous findings of close relationships in *Pst* populations between Gansu and Ningxia[29]; Gansu and Shaanxi[27]; and among Gansu, Qinghai, and Xinjiang[28]. In addition to Gansu, most isolates from Tibet were in the same cluster as those from Xinjiang and Qinghai. Both the air trajectory modelling and genetic structure results suggest that Qinghai could be a region that may receive urediniospores from both Xinjiang and Tibet in the spring. Isolates from Qinghai were separated into Cluster 2.3 and Cluster 2.4, which could be related to geographical separation by the Laji Mountain. Similarly, some *Pst* isolates in Guizhou and Yunnan were in the same clusters, which can be interpreted on the basis of geographic features and airflow trajectory. Population structure analysis indicated that Sichuan-Yunnan is a single epidemic region where *Pst* from various sources meets; this is supported by the highest genetic diversity of sub-populations in Cluster 2. Population differentiation analysis confirmed frequent exchanges between Sichuan-Yunnan and subclusters within Cluster 2.1 belonging to large areas in China. Recurring interchanges of *Pst* among Sichuan, Yunnan, and Guizhou is a possible reason explaining the dispersal route between Guizhou and Sichuan-Yunnan[47].

In China, the Huang-Huai-Hai plain is the major wheat-producing region, producing around 70% of the national wheat grain[48]. Overwintering inoculum in Henan and Hubei is important in affecting spring rust occurrence in the Huang-Huai-Hai wheat-growing region. Field surveys showed that stripe rust is always found first on the older wheat leaves in Henan and Hubei in the spring, suggesting that the epidemic is largely caused by the local overwintering *Pst* inoculum. Trajectory tracking and field surveys in the current study indicated that Longnan is the primary source of *Pst* inoculum and responsible for overwintering inoculum in Henan and Hubei spread via eastern Gansu and central Shaanxi. Previous field observations found that disease incidences of Longnan and central Shaanxi are related to that of Hubei and Henan in most years. However, during the severe stripe rust epidemic in 2017, rust in Henan and Hubei was more related to the epidemics in Yunnan and Guizhou instead of Longnan and Shaanxi. Therefore, as a secondary source, inoculum in Guizhou Plateau can significantly affect the epidemics in the Huang-Huai-Hai region depending on climatic conditions. Although the Himalayan region is another center of *Pst* inoculum source, inoculum from this region is unlikely to directly cause rust in the Huang-Huai-Hai wheat-producing region, but mainly contributes to rust development in Tibet and Sichuan Basin.

The present findings on *Pst* inoculum sources and possible key dispersal routes are valuable for implementing effective strategies for *Pst* monitoring and management. Reducing *Pst* oversummering inoculum in the inoculum source and conducting effective management of rust in those areas along the important dispersal routes should be an efficient strategy to prevent severe rust from occurring in the main wheat production regions. Growing resistant cultivars and using fungicides when needed can be used to reduce *Pst* inoculum in the oversummering regions. These approaches should also be considered in the other stripe rust epidemic regions in case rust control in the key overwintering regions is not as effective as desired.

Methods

Sample collection, purification, and multiplication

A total of 154 *Pst*-infected wheat leaves were sampled in 11 provinces throughout China, and only one leaf bearing uredinia was sampled from a single infection spot in a single field. Collected samples were stored in a desiccator at 4°C no longer than a

week before being used to inoculate wheat seedlings for inoculum multiplication. To obtain fresh urediniospores for inoculation, the infected leaf was placed on a water-soaked paper towel in the dew chamber for 16h in the dark at 10°C. Urediniospores from a single fresh uredium were used to inoculate the seedlings of Mingxian 169, a wheat cultivar highly susceptible to all identified *Pst* races so far in China. Incubation of inoculated seedlings and subsequent collection of urediniospores from infected seedlings followed procedures previously described[49,50]. Multiplication of urediniospores were repeated until sufficient quantity for DNA extraction.

DNA extraction and whole-genome sequencing

Genomic DNA was extracted directly from harvested urediniospores using the cetyltrimethyl ammonium bromide (CTAB) method with some modifications[51]. Quality and quantity of genomic DNA were examined with 0.8% agarose gel electrophoresis and NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA). Whole-genome sequencing was performed on the Illumina NovaSeq 6000 platform by Novogene Co., Ltd. (Beijing, China). The library of DNA for each *Pst* isolate was constructed with the 150 bp paired-end fragmented sequences using the NEB Next® Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA). All raw sequence reads were deposited in the National Center of Biotechnology Information (NCBI) under the directory of BioProject PRJNA809046. The read depth coverage of each isolate was assessed using SAMtools version 1.9[52]. Raw reads quality were evaluated with FastQC version 0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), reads with inferior quality were removed using Trimmomatic version 0.39[53].

Detection and annotation of variants

Genomic variants were identified following the previously reported framework[54]. To select a reference genome, *Pst* genomes (11-281[55], 93-210[56], 104E137A-[57], DK09_11[58], and PST-130[59]) sequenced by third-generation sequencing technology were evaluated through statistical assessment with benchmark universal single-copy orthologs (BUSCO) version 3.0.2 (https://gitlab.com/ezlab/busco). Error-corrected reads of 154 *Pst* isolates were aligned with the selected reference genome using BWA-mem version 0.7.17 with the default settings[60]. The obtained files in the SAM format were converted to the BAM format using SAMtools. The BAM files were cleaned, sorted, and validated with the Picard tools version 2.23.9 (https://broadinstitute.github.io/picard/). Variant calling was implemented with the

Genome Analysis Toolkit (GATK) version 4.1.9.0 HaplotypeCaller and ApplyBQSR (https://gatk.broadinstitute.org/) using the indexed filtered BAM files as input. Two-round callings were conducted to identify SNPs and indels throughout the genomes, as previously reported[61]. Hardfiltering was applied to tease apart substandard variants and select high-quality SNPs and indels using GATK VariantFiltration. Parameters were set as "QD<2.0 || FS>60.0 || MQ<40.0 || MQRankSum<-12.5 || ReadPosRankSum<-8.0 || SOR>3" according to the recommendation in GATK Best Practice Pipeline with modifications (https://github.com/gatk-workflows/). Filtered variants of each isolate were merged into a single genome Variant Call Format (gVCF) file using GATK CombineGVCFs. Joint genotyping was implemented on the combined file using GATK GenotypeGVCFs. The output file was filtered using VCFtools version 0.1.13 (https://vcftools.github.io), with the parameters "--maf 0.05 --mac 3 --minQ 200 --max-missing 0.3 --min-meanDP 38 --max-meanDP 71", to keep the high-quality variants with the coverage of read depth oscillated less than 30% from the average. Variants were further annotated by SnpEff version 4.3t[62].

Phylogenetic and population structure analyses

To ensure adjacent SNPs were not linked, the squared correlation (r^2) was calculated to estimate the linkage disequilibrium[2] between each pair of SNPs in a window of 50 SNPs with 1 SNP shifted forward across the genome using the PLINK version 1.9[63]. A SNP was filtered out when r^2 was greater than 0.2. Remaining unlinked SNPs were used to construct maximum likelihood (ML) phylogenetic trees with 1000 replications using RAxML version 8.2.12[64]. SNPs of *Puccinia triticina* (SRR11479796) detected using the same approach were used as an outgroup. A rapid Bootstrap analysis with the GTRGAMMAX model was performed. Phylogenetic trees were visualized and annotated using iTOL version 6 (https://itol.embl.de/). A bootstrap value of at least 60 indicated that the node is well supported. Population genetic structures was examined using ADMIXTURE version 1.3.0 with K values ranged from 2 to 10 for 500 replications[65]. Principal component analysis (PCA) was conducted based on the SNP data using PLINK version 1.9[63], and the first two PCs were plotted using the *ggplot2* package in the R program version 4.1.2.

Genetic diversity and heterozygosity identification

The genetic diversity of *Pst* isolates in each cluster identified by phylogeny and population genetic structure was evaluated by nucleotide diversity parameters (π and θ_ω). π was computed in 5,000 bp sliding windows using VCFtools version 0.1.13

(https://vcftools.github.io). θ_ω was calculated using the formula $\theta_\omega = \frac{K}{a_n}$, $a_n = \sum_{i=1}^{n-1} \frac{1}{i}$, where *K* is the number of segregating sites and *n* stands for the total number of haploids[66]. The significant level of genetic diversity indices π and θ_ω was tested between different groups with a non-parametric Kruskal-Walli's tests. To determine the genetic variation in the population, observed and expected heterozygosity were estimated by the formula: $Ho = \frac{N-O(hom)}{N}$ and $He = \frac{N-E(hom)}{N}$ using PLINK version 1.9, where *N* is the total number of alleles, and $O(hom)$ and $E(hom)$ are the observed and expected number of homologous alleles, respectively[67]. A permutation test was performed to examine the significant variation between observed and expected heterozygosity. Neutrality test with Tajima's *D* was further used to confirm the heterozygosity level among different clusters using VCFtools version 0.1.13 with the sliding windows of 5 000 bp. Significant level derived from Tajima's *D* among different clusters were detected with Kruskal-Walli's test.

Determination of population differentiation, introgression, and migration

Population differentiation between pairwise clusters was calculated by paired Weir and Cockerham's population fixation statistics (F_{ST}) using VCFtools version 0.1.13. Genetic introgression analyses were computed using the *D* statistics (also known as ABBA-BABA test) to identify evidence of genetic introgression between populations using Dsuite software[68]. The *D* and f_4-ratio can be presented as a given four-taxon topology ((P1, P2), P3, O), ABBA pattern means that P1 harbors an ancestral allele "A" derived from outgroup O, while P2 and P3 share a different allele "B" at a SNP site. Similarly, BABAs pattern shows P1 and P3 share an allele "B", while P2 retains the allele "A" derived from outgroup O. Under the null hypothesis, which indicates no genetic introgression occurs, equal frequencies should be observed in ABBA and BABA patterns, whereas a significant deviation indicates introgression occurred between P3 and P1 or P2. f_4-ratio shows the mixing proportions of admixture events[39]. Migration events among different populations on the maximum likelihood phylogeny were estimated using TreeMix with high-quality SNPs filtered by LD[69]. The number of migration events (m) between any two populations extended from 1 to 7 were tested with 20 replications, and the optimal number of events were investigated using the *OptM* package in the R program version 4.1.2[70].

Trajectory simulation and traceability analysis of *Pst* migration

According to earlier epidemiological studies of *Pst* in China, representative locations were selected to perform trajectory simulation of *Pst* using Hybrid Single-Particle Lagrangian Integrated Trajectory version 4 (HYSPLIT-4)[71]. Weekly meteorological data from 2010 to 2019 were obtained from the National Oceanic and Atmospheric Administration (NOAA) website (ftp://arlftp.arlhq.noaa.gov/pub/archives/gdas1). The data were converted to 1° latitude-longitude (360 by 181) grids using the NOAA's Air Resource Laboratory archiving program. Altitudes and simulation time periods were selected based on different geographical locations to mimic biological events of *Pst* dispersal (Supplementary Table 11, Supplementary Table 12). Using the meteorological data as the input for the Trajectory model in HYSPLIT-4, trajectory frequencies were calculated for *Pst* dispersal from each location at 100-1000 m, 1000-3000 m, and 3000-5000 m above ground level (AGL) depending on characteristics of geographic locations. As *Pst* spores can survive in the air for up to 120 h[33], each simulation was run for 120 h. Average trajectory frequency (ATF) spanning of the entire ten years for each location was calculated. The range of those high-frequent trajectories was visualized using the *ggplot2*, *tmap* and *sp Gallery* packages in R version 4.1.2. To trace the potential source of *Pst* dispersal in Sichuan Basin, Henan, and Hubei from October 1st to November 30th (Supplementary Table 11), softwares MeteoInfoMap and TrajStat were applied for backward trajectory simulation analysis (http://meteothink.org/index.html) using the meteorological data from NOAA. The angle distance clustering method was used to compute backward trajectory clusters in the altitude extended from 100 m to 3000 m AGL (http://meteothink.org/docs/trajstat/cluster_cal.html).

Field surveillance surveys on stripe rust occurrence

Occurrence of wheat stripe rust in the main epidemic regions in China was determined by the National Agro-Tech Extension and Service Center and Regional Plant Protection Extension Stations in these regions. Two most recent epidemics in 2017 and 2020, were selected to estimate the temporal sequence of disease occurrence among the selected regions. The disease onset in the representative locations was estimated based on the field monitoring. *Pst* dispersal routes were inferred based on the earliest dates of stripe rust observations and dates of sowing and harvesting.

Data availability

Raw sequenced reads of 154 samples have been deposited in National Center for Biotechnology Information BioProject database (PRJNA809046).

Code availability

Code analysed in this study have been described in Methods and are available on GitHub at https://github.com/yuxiang-li/code_Pst_population.

References

1. Fisher, M.C. et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186-194 (2012).
2. Ristaino, J.B. et al. The persistent threat of emerging plant disease pandemics to global food security. *Proc Natl Acad Sci USA* **118**, 23 (2021).
3. Bowden, J., Gregory, P. & Johnson, C. Possible wind transport of coffee leaf rust across the Atlantic Ocean. *Nature* **229**, 500-501 (1971).
4. Wellings, C.R. *Puccinia striiformis* in Australia: a review of the incursion, evolution, and adaptation of stripe rust in the period 1979–2006. *Aust J Agric Res* **58**, 567-575 (2007).
5. Dyer, A., Matusralt, M., Drenth, A., Cohen, A. & Splelman, L. Historical and recent migrations of *Phytophthora infestans*: chronology, pathways, and implications. *Plant Dis* **77**, 653-661 (1993).
6. Brown, J.K. & Hovmøller, M.S. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* **297**, 537-541 (2002).
7. Nagarajan, S. & Singh, D.V. Long-distance dispersion of rust pathogens. A*nnu Rev Phytopathol* **28**, 139-153 (1990).
8. Goss, E.M. Genome-enabled analysis of plant-pathogen migration. *Annu Rev Phytopathol* **53**, 121-135 (2015).
9. Chen, X.M. Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Can J Plant Pathol* **27**, 314-337 (2005).
10. Wellings, C.R. Global status of stripe rust: a review of historical and current threats. Euphytica **179**, 129-141 (2011).
11. Hovmøller, M.S., Sørensen, C.K., Walter, S. & Justesen, A.F. Diversity of *Puccinia striiformis* on cereals and grasses. *Annu Rev Phytopathol* **49**, 197-217 (2011).
12. Chen, W.Q. et al. Race dynamics, diversity, and virulence evolution in *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust in China from 2003 to 2007. *Plant Dis* **93**, 1093-1101 (2009).
13. Johnson, R. & Taylor, A.J. Isolates of *Puccinia striiformis* collected in England from the wheat varieties Maris Beacon and Joss Cambier. *Nature* **238**, 105-106 (1972).
14. Beresford, R.M. Stripe rust (*Puccinia striiformis*), a new disease of wheat in New Zealand. *Cereal Rusts Bulletin* **10**, 35-41 (1982).

15. Duveiller, E., Singh, R.P. & Nicol, J.M. The challenges of maintaining wheat productivity: pests, diseases, and potential epidemics. Euphytica **157**, 417-430 (2007).
16. Prashar, M., Bhardwaj, S.C., Jain, S.K. & Datta, D. Pathotypic evolution in *Puccinia striiformis* in India during 1995–2004. *Aust J Agric Res* **58**, 602-604 (2007).
17. Chen, X.M. Pathogens which threaten food security: *Puccinia striiformis*, the wheat stripe rust pathogen. Food Secur **12**, 239-251 (2020).
18. Hovmøller, M.S., Justesen, A.F. & Brown, J.K.M. Clonality and long-distance migration of *Puccinia striiformis* f. sp. *tritici* in north-west Europe. *Plant Pathol* **51**, 24-32 (2002).
19. Kim, K.S. & Beresford, R.M. Use of a spectrum model and satellite cloud data in the simulation of wheat stripe rust (*Puccinia striiformis*) dispersal across the Tasman Sea in 1980. *Agric For Meteorol* **148**, 1374-1382 (2008).
20. Ali, S. et al. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia striiformis* f. sp. *tritici*. PLoS Pathog **10**, e1003903 (2014).
21. Sharma-Poudyal, D. et al. Molecular characterization of international collections of the wheat stripe rust pathogen *Puccinia striiformis* f. sp. *tritici* reveals high diversity and intercontinental migration. Phytopathology **110**, 933-942 (2020).
22. Zeng, S.M. & Luo, Y. Long-distance spread and interregional epidemics of wheat stripe rust in China. Plant Dis **90**, 980-988 (2006).
23. Ma, Z.H. Researches and control of wheat stripe rust in China. *J of Plant Prot* **45**, 1-6 (2018).
24. Wan, A.M. et al. Wheat stripe rust epidemic and virulence of *Puccinia striiformis* f. sp. *tritici* in China in 2002. *Plant Dis* **88**, 896-904 (2004).
25. Li, Z.Q. & Zeng, S.M. Wheat rust in China. *China Agriculture Press, Beijing, China.* (2002).
26. Liang, J.M. et al. Population genetic structure and the migration of *Puccinia striiformis* f. sp. *tritici* between the Gansu and Sichuan Basin populations of China. *Phytopathology* **106**, 192-201 (2016).
27. Wang, C.H., Li, Y.X., Wang, B.T. & Hu, X.P. Genetic analysis reveals relationships among populations of *Puccinia striiformis* f. sp. *tritici* from the Longnan, Longdong and Central Shaanxi Regions of China. *Phytopathology* doi.org/10.1094/PHYTO-07-20-0312-R (2022).
28. Wan, Q., Liang, J.M., Luo, Y. & Ma, Z.H. Population genetic structure of *Puccinia striiformis* in northwestern China. *Plant Dis* **99**, 1764-1774 (2015).
29. Liang, J.M., Wan, Q., Luo, Y. & Ma, Z.H. Population genetic structures of *Puccinia*

635 *striiformis* in Ningxia and Gansu provinces of China. *Plant Dis* **97**, 501-509 (2013).
636 30. Zhan, G.M. et al. Virulence and molecular diversity of the *Puccinia striiformis* f. sp.
637 *tritici* population in Xinjiang in relation to other regions of western China. *Plant Dis* **100**,
638 99-107 (2016).
639 31. Hu, X.P. et al. Race composition of *Puccinia striiformis* f. sp. *tritici* in Tibet, China.
640 *Plant Dis* **96**, 1615-1620 (2012).
641 32. Hu, X.P. et al. Population genetic analysis of *Puccinia striiformis* f. sp. *tritici*
642 suggests two distinct populations in Tibet and the other regions of China. *Plant Dis* **101**,
643 288-296 (2017).
644 33. Wang, H.G., Yang, X.B. & Ma, Z.H. Long-distance spore transport of wheat stripe
645 rust pathogen from Sichuan, Yunnan, and Guizhou in southwestern China. *Plant Dis*
646 **94**, 873-880 (2010).
647 34. Wan, A.M. et al. Occurrence of wheat stripe rust and monitoring of physiological
648 races of *Puccinia striiformis* f. sp. *tritici* in China in 2000 ~ 2001. *J Plant Prot* **28**, 5-8
649 (2002).
650 35. Huang, C. et al. Epidemics analysis of wheat stripe rust in China in 2017. *J Plant*
651 *Prot* **44**, 162-166 (2018).
652 36. Hubbard, A. et al. Field pathogenomics reveals the emergence of a diverse wheat
653 yellow rust population. *Genome Biol* **16**, 23 (2015).
654 37. Bueno-Sancho, V. et al. Pathogenomic analysis of wheat yellow rust lineages
655 detects seasonal variation and host specificity. *Genome Biol Evol* **9**, 3282-3296 (2017).
656 38. Ding, Y. et al. Incursions of divergent genotypes, evolution of virulence and host
657 jumps shape a continental clonal population of the stripe rust pathogen *Puccinia*
658 *striiformis*. *Mol Ecol* **00**, 1-19 (2021).
659 39. Patterson, N. et al. Ancient Admixture in Human History. *Genetics* **192**, 1065-1093
660 (2012).
661 40. Liang, J.M., Liu, X.F., Tsui, C.K., Ma, Z.H. & Luo, Y. Genetic structure and
662 asymmetric migration of wheat stripe rust pathogen in western epidemic areas of China.
663 *Phytopathology* **111**, 1252-1260 (2021).
664 41. Huang, M.M. et al. Analyses of wheat yellow rust populations reveal sexual
665 recombination and seasonal migration pattern of *Puccinia striiformis* f. sp. *tritici* in
666 Gangu, Northwestern China. *Phytopathology* doi.org/10.1094/PHYTO-12-20-0558-R
667 (2021).
668 42. Wang, C.C. et al. Genetic diversity and population structure of *Puccinia striiformis*
669 f. sp. *tritici* reveal its migration from central to eastern China. *Crop Prot* **128**, 104974
670 (2020).

671 43. Charlesworth, D. Balancing selection and its effects on sequences in nearby
672 genome regions. *PLoS Genet* **2**, e64 (2006).
673 44. Zhao, J. et al. Identification of eighteen *Berberis* species as alternate hosts of
674 *Puccinia striiformis* f. sp. *tritici* and virulence variation in the pathogen isolates from
675 natural infection of barberry plants in China. *Phytopathology* **103**, 927-934 (2013).
676 45. Wang, Z.Y. et al. Virulence variations of *Puccinia striiformis* f. sp. *tritici* isolates
677 collected from *Berberis* spp. in China. *Plant Dis* **100**, 131-138 (2016).
678 46. Li, S. et al. Identification of eight *Berberis* species from the Yunnan-Guizhou
679 plateau as aecial hosts for *Puccinia striiformis* f. sp. *tritici*, the wheat stripe rust
680 pathogen. *J Integr Agric* **20**, 1563-1569 (2021).
681 47. Wang, H., Yang, X.B. & Ma, Z. Long-Distance Spore Transport of wheat stripe
682 Rust Pathogen from Sichuan, Yunnan, and Guizhou in Southwestern China. *Plant Dis*
683 **94**, 873-880 (2010).
684 48. Wang, E., Chen, C. & Yu, Q. Modeling the response of wheat and maize
685 productivity to climate variability and irrigation in the North China Plain. in *18th World*
686 *IMACS/MODSIM Congress* 2742-2748 (Citeseer, 2009).
687 49. Wan, A.M. & Chen, X.M. Virulence characterization of *Puccinia striiformis* f. sp.
688 *tritici* using a new set of *Yr* single-gene line differentials in the United States in 2010.
689 *Plant Dis* **98**, 1534-1542 (2014).
690 50. Li, Y.X., Wang, M.N., See, D.R. & Chen, X.M. Ethyl-methanesulfonate
691 mutagenesis generated diverse isolates of *Puccinia striiformis* f. sp. *tritici*, the wheat
692 stripe rust pathogen. *World J Microbiol Biotechnol* **35**, 28 (2019).
693 51. Chen, X.M., Line, R.F. & Leung, H. Relationship between virulence variation and
694 DNA polymorphism in *Puccinia striiformis*. *Phytopathology* 83, 1489-1497 (1993).
695 52. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**,
696 2078-2079 (2009).
697 53. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
698 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
699 54. DePristo, M.A. et al. A framework for variation discovery and genotyping using
700 next-generation DNA sequencing data. *Nat Genet* **43**, 491-498 (2011).
701 55. Li, Y.X., Xia, C.J., Wang, M.N., Yin, C.T. & Chen, X.M. Genome sequence resource
702 of a *Puccinia striiformis* isolate infecting wheatgrass. *Phytopathology* **109**, 1509-1512
703 (2019).
704 56. Xia, C.J. et al. Genomic insights into host adaptation between the wheat stripe rust
705 pathogen (*Puccinia striiformis* f. sp. *tritici*) and the barley stripe rust pathogen (*Puccinia*
706 *striiformis* f. sp. *hordei*). *BMC Genom* **19**, 664 (2018).
707 57. Schwessinger, B. et al. A near-complete haplotype-phased genome of the

dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high interhaplotype diversity. *MBio* **9**, e02275-17 (2018).
58. Schwessinger, B. et al. Distinct life histories impact dikaryotic genome evolution in the rust fungus *Puccinia striiformis* causing stripe rust in wheat. *Genome Biol Evol* **12**, 597-617 (2020).
59. Vasquez-Gross, H., Kaur, S., Epstein, L. & Dubcovsky, J. A haplotype-phased genome of wheat stripe rust pathogen *Puccinia striiformis* f. sp. *tritici*, race PST-130 from the Western USA. *Plos One* **15**, e0238611 (2020).
60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
61. Li, Y.X., Xia, C.J., Wang, M.N., Yin, C.T. & Chen, X.M. Whole-genome sequencing of *Puccinia striiformis* f. sp. *tritici* mutant isolates identifies avirulence gene candidates. *BMC Genom* **21**, 1-22 (2020).
62. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
63. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
65. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664 (2009).
66. Watterson, G.A. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**, 256-276 (1975).
67. Chen, N., Cosgrove, E.J., Bowman, R., Fitzpatrick, J.W. & Clark, A.G. Genomic consequences of population decline in the endangered Florida scrub-Jay. *Curr Biol* **26**, 2974-2979 (2016).
68. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite-Fast *D*-statistics and related admixture evidence from VCF files. *Mol Ecol Resour* **21**, 584-595 (2021).
69. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nat Preced* doi.org/10.1038/npre.2012.6956.1 (2012).
70. Fitak, R.R. *OptM*: estimating the optimal number of migration edges on population trees using Treemix. *Biol Methods Protoc* **6**, bpab017 (2021).
71. Draxler, R.R. & Hess, G.D. An overview of the HYSPLIT_4 modeling system of trajectories, dispersion, and deposition. *Aust Meteorol Mag* **47**, 295-308 (1998).

Acknowledgement

This research is supported by the National Key Research and Development Program of China (2021YFD1401000) and National Natural Science Foundation of China (31471731 and 31772102).

Author contribution

X.H. and Z.K. designed and managed the research project; Y.L. assisted experimental design; X.H. and Y.L. performed data analysis and wrote the manuscript; J.D. conducted population genetic analysis. T.Z. and B.W. performed air trajectory analysis; J.Z. and C.W. participated in inoculating and harvesting spores; B.W., Q.Y, M.L, C.L., Y.P., S.Q., H.G., W.H., X.F., Y.B., and Z.Q. participated in disease surveys and samplings; G.Z. and F.T. assisted spore production; B.W., C.H., and W.L. conducted disease surveys; H.S., J.Z., X.X., and X.C. offered suggestions for the study and revised the manuscript. All the authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

762 **Table1 The values of *D* statistics between different quartet taxon topologies**

P1	P2	P3	*D* statistic	*Z* score[a]	*P* value	*P* adjusted[b]	f_4-ratio[c]
Cluster 2.1	Cluster 2.2	Cluster 1.1	0.2340	19.19	0	0**	0.2013
Cluster 2.3	Cluster 2.2	Cluster 1.1	0.2234	11.84	0	0**	0.1957
Cluster 2.4	Cluster 2.2	Cluster 1.1	0.1834	8.283	6.000×10^{-17}	3.000×10^{-16}**	0.1672
Cluster 2.1	Cluster 2.2	Cluster 1.2	0.2219	16.71	0	0**	0.1631
Cluster 2.3	Cluster 2.2	Cluster 1.2	0.2058	9.856	0	0**	0.1553
Cluster 2.4	Cluster 2.2	Cluster 1.2	0.1630	6.372	9.230×10^{-11}	4.080×10^{-10}**	0.1279
Cluster 2.1	Cluster 2.2	Cluster 1.3	0.1795	13.20	0	0**	0.1372
Cluster 2.3	Cluster 2.2	Cluster 1.3	0.2008	10.29	0	0**	0.1520
Cluster 2.4	Cluster 2.2	Cluster 1.3	0.1272	4.994	2.690×10^{-7}	1.150×10^{-6}**	0.1032
Cluster 2.3	Cluster 2.2	Cluster 2.1	0.06574	3.709	0.0001041	0.0003311**	2.347

763

764 [a] The *Z* scores are the normalization of *D* statistic values.

765 [b] The *P* values calculated from multiple *D* statistics are adjusted using the Benjamini Hochberg method. "*" represents < 0.001; and

766 "**" < 0.0001.

767 [c] The f_4-ratios are the mixing proportions of admixture events.

Figure captions

Fig. 1. Geographic distribution and population genetic structure of 154 *Puccinia striiformis* f. sp. *tritici* (*Pst*) isolates.

(a) Geographical distribution of the 154 *Pst* isolates collected from eleven provinces in China. Different colors represent the different sampling sites. (b) The maximum likelihood phylogenetic tree analysis of the 154 *Pst* isolates using single-nucleotide polymorphisms (SNPs) pruned by linkage disequilibrium (LD), the star symbols on the branches indicate bootstrap values equal or greater than 60%. Cluster 1 and Cluster 2 were marked with blue and red, respectively. Samples collected from different provinces were represented with rounds in various colors. (c) Principal component analysis plots of SNPs from the 154 isolates filtered with LD. The blue dots represent Cluster 1, while red dots designate Cluster 2. (d) Ancestry coefficient analysis of the *Pst* accessions. The accessions were grouped into two clusters with K = 2.

Fig. 2. The population genetic structures of *Puccinia striiformis* f. sp. *tritici* (*Pst*) accessions in Cluster 1.

(a) The maximum likelihood phylogenetic tree of *Pst* isolates in Cluster 1. The clusters colored with blue, red, and green represent Cluster 1.1, Cluster 1.2, and Cluster 1.3, respectively. Stars on the branches indicated bootstrap values equal or greater than 60%. (b) Principal component analysis plot of the *Pst* isolates within Cluster 1. The blue, red, and green dots represent different clusters. (c) Ancestry coefficient analysis of the *Pst* isolates in Cluster 1. The isolates were grouped into three clusters at K = 3.

Fig. 3. The population genetic structures of *Puccinia striiformis* f. sp. *tritici* (*Pst*) accessions in Cluster 2.

(a) Non-root phylogeny of *Pst* isolates in Cluster 2. Population structures colored with blue, red, green, and yellow represents Cluster 2.1, Cluster 2.2, Cluster 2.3, and Cluster 2.4, respectively. Bootstrap values greater than 60% were marked with stars. (b) Principal component analysis plots of the *Pst* isolates in Cluster 2 using single-nucleotide polymorphisms. Different clusters are shown in different colors. (c) Ancestry coefficient analysis of the *Pst* isolates in Cluster 2. The accessions were classified into four clusters with K = 4. (d) Stacked histogram of four clusters in Cluster 2. Proportion of *Pst* accession collected from different provinces are marked with different colors.

Fig. 4. The genetic diversity and heterozygosity of *Puccinia striiformis* f. sp. *tritici* (*Pst*) sub-populations in China.

(a) The box plot of the nucleotide diversity of *Pst* subpopulations. Genetic diversity was tested with π and $\theta\omega$. Letters on the top of each box indicate significant levels of Kruskal-Walli's tests (P < 0.05). (b) The histogram of the observed and expected heterozygosity of each cluster. A permutation test was performed to examine the significant variation between observed and expected heterozygosity. P values were adjusted using the Benjamini Hochberg method. * represents P adjusted < 0.05 and ** means P adjusted < 0.01. Error bars represent the standard errors of the observed

and expected heterozygosity for samples in each cluster.

Fig. 5. The genetic differentiation, introgression, and migration in *Puccinia striiformis* f. sp. *tritici* (*Pst*) sub-populations.

(a) The highly significant taxon topologies detected from *D* statistics (*P* adjusted < 0.001). Cluster 1.1, 1.2, 1.3, and 2.1 in red box of each group represent P3. The red and blue lines show the different introgression patterns with P3. (b) The results of *D* statistics among *Pst* clusters. Rows represent nodes within the tree topology, and columns represent clusters regarded as P3. Each cell shows the f_b values between a tree node (rows) and each cluster (column). Grey cells show the comparisons cannot be tested with *D* statistics. Stars marked on the cell shows significant tests at *P* adjusted greater than 0.001. (c) The migration event detected with TreeMix analysis. Migration arrows are colored according to their weight. Horizontal branch length is proportional to the amount of genetic drift that has occurred on the branch. Scale bar shows ten times the average standard error of the entries in the sample covariance matrix.

Fig. 6. Trajectory analysis of *Puccinia striiformis* f. sp. *tritici* (*Pst*) migration routes in essential regions in China.

(a) The air trajectory within the major potential *Pst* epidemic regions. The green and orange arrows show the air trajectory in spring and autumn, respectively. The simulation period in Tianshui and Pingliang is October 1st to November 30th in autumn, in Mianyang is January 1st to April 10th in spring, and in Nanyang and Xiangyang is February 20th to April 10th. (b) The air trajectory from the Tibet regions in spring. The simulation period is April 1st to May 31st. (c) The aerial transportation from the Guizhou regions. The blue arrows show the air trajectory in spring and autumn. The simulation periods in Bijie and Liupanshui are October 1st to November 30th in autumn and February 20th to April 10th in spring. (d) The migration of air trajectory from the Sichuan regions in spring with the simulation period of January 1st to April 10th.

Fig. 7. The major migration routes of *Puccinia striiformis* f. sp. *tritici* (*Pst*) in China.

Migration routes inferred from population genetics, air trajectory and disease occurrence. The red arrow represents the migration route supported by all analyses. The dark blue arrow shows the migration routes supported by population genetics and air trajectory. The green arrows designate the migration routes supported by disease survey and air trajectory. The blue arrow indicates the migration routes were solely supported by air trajectory. The width of each arrow shows the possibility of the migration routes.

Supplementary Figure Legends

Supplementary Fig. 1. The population structure of *Puccinia striiformis* f. sp. *tritici* (*Pst*) isolates of Cluster 2.1. (a) The maximum likelihood tree analysis of *Pst* isolates in Cluster 2.1 used SNPs pruned by linkage disequilibrium. Tree structures were colored with blue, red, and green colors represents the Cluster 2.1.1, Cluster 2.1.2, and Cluster 2.1.3, respectively. The blue, red, and green strips represent Cluster 2.1.1, Cluster 2.1.2, and Cluster 2.1.3 severally. Grey star marks mean the tree nodes with bootstrap values more than 60%. (b) Stacked histogram of different clusters representing the sampling sites components.

Supplementary Fig. 2. The neutral tests of *Puccinia striiformis* f. sp. *tritici* (*Pst*) clusters. The width of violin plot in different height represents the density of Tajima's *D* values. The diamonds represent the median Tajima's *D* values of *PST* clusters. The letters on the top of the boxes means the significant levels after tested with method of Kruskal-Walli's test ($P < 0.05$).

Supplementary Fig. 3. The genetic differentiation (F_{ST}) between different *Puccinia striiformis* f. sp. *tritici* (*Pst*) clusters. The sizes of the squares represent the values of F_{ST} between different clusters. The color bar from blue to red means the values of F_{ST} from low to high.

Supplementary Fig. 4. The genetic differentiation (F_{ST}) between the subsets of Cluster 2.1 and other clusters in Cluster 2. The sizes of the squares represent the values of F_{ST} between different clusters. The color bar from blue to red means the values of F_{ST} from low to high.

Supplementary Fig. 5. The optimal migrate events number (*m*) selected from the second-order rate of change (Δm) across values of *m* by *Optm*.

Supplementary Fig. 6. The frequencies of air trajectory analysis in Tianshui and Pingliang of Gansu provinces. The simulation period is October 1st to November 30th. (a) The frequencies of air trajectory originating from Tianshui at the altitude from 100 m to 1000 m above ground level (AGL). (b) The frequencies of air trajectory derived from Tianshui at 1000 m to 3000 m AGL. (c) The frequencies of air trajectory from Pingliang at 100 m to 1000 m AGL. (d) The frequencies of air trajectory originating from Pingliang at 1000 m to 3000 m AGL. The colors represent the densities of air trajectory

Supplementary Fig. 7. The frequencies of air trajectory analysis in the Tibet regions during spring. The frequencies of air trajectory originating from Linzhi. The simulation period is April 1st to May 31st. The color depths represent the densities of air trajectory in April to May at the altitude from 3000-5000 m above ground level.

Supplementary Fig. 8. The frequencies of air trajectory analysis in the Xinjiang region.

The simulation period is April 1st to May 20th. (a) The frequencies of air trajectory originating from the Aksu region of western Xinjiang in spring at 3000-5000 above ground level (AGL). (b) The frequencies of air trajectory originating from Kazakh Autonomous Prefecture of Ili in western Xinjiang in spring at 3000-5000 m AGL. (c) Schematic diagrams of migration routes derived from western Xinjiang

Supplementary Fig. 9. The frequencies of air trajectory analysis in Bijie and Liupanshui regions of Guizhou province. The simulation periods are October 1st to November 30th in autumn and February 20th to April 10th in spring. (a) The frequencies of air trajectory originating from Bijie at 1000-3000 m above ground level (AGL) in autumn. (b) The frequencies of air trajectory derived from Bijie at 1000-3000 m AGL in spring. (c) The frequencies of air trajectory from Liupanshui at 1000-3000 m AGL in autumn. (d) The frequencies of air trajectory from Liupanshui at 1000-3000 m AGL in spring. The colors represent the densities of air trajectory

Supplementary Fig. 10. The frequencies of air trajectory analysis in regions during spring. The frequencies of air trajectory originating from Mianyang. The simulation period is January 1st to April 10th. The colors represent the densities of air trajectory at the altitude from 100-1000 m above ground level. The colors represent the densities of air trajectory

Supplementary Fig. 11. The backward trajectory analysis in important *Puccinia striiformis* f. sp. *tritici* (*Pst*) overwintering regions in China. The simulation period is October 1st to November 30th. (a) The backward trajectory analysis of the air trajectory gathered in Mianyang at the altitude from 100-3000 m above ground level (AGL) in autumn. (b) The backward trajectory analysis of the air trajectory gathered in Nanyang at the altitude from 100-3000 m AGL in autumn. (c) The backward trajectory analysis of the air trajectory gathered in Xiangyang at the altitude from 100-3000 m AGL in autumn. The green lines show the specific air trajectories. The black arrows show the potential air trajectory associated with potential *Pst* transportation

Figures

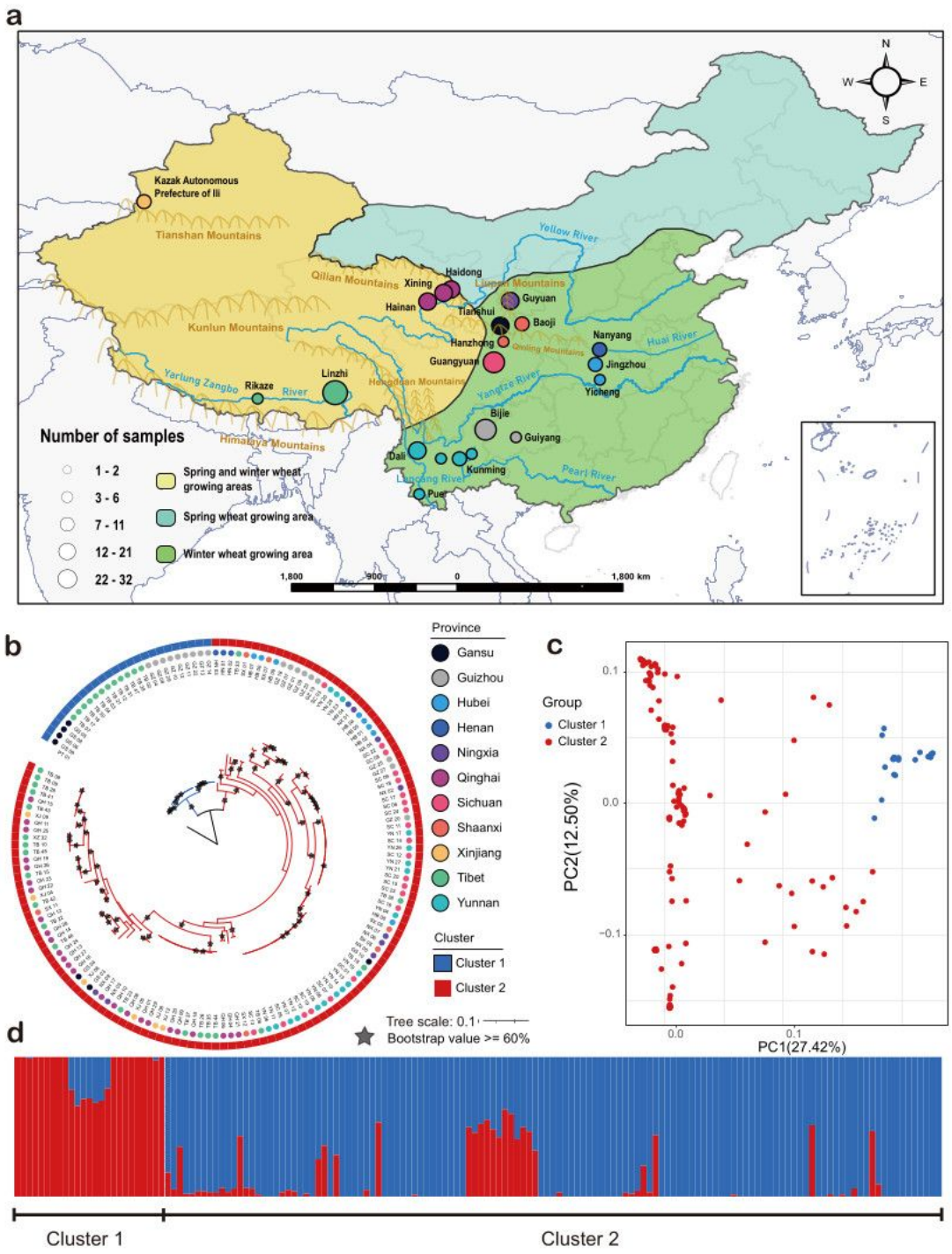


Figure 1

Geographic distribution and population genetic structure of 154 Puccinia striiformis f. sp. tritici (Pst) isolates.

(a) Geographical distribution of the 154 Pst isolates collected from eleven provinces in China. Different colors represent the different sampling sites. (b) The maximum likelihood phylogenetic tree analysis of the 154 Pst isolates using single-nucleotide polymorphisms (SNPs) pruned by linkage disequilibrium (LD), the star symbols on the branches indicate bootstrap values equal or greater than 60%. Cluster 1 and Cluster 2 were marked with blue and red, respectively. Samples collected from different provinces were represented with rounds in various colors. (c) Principal component analysis plots of SNPs from the 154 isolates filtered with LD. The blue dots represent Cluster 1, while red dots designate Cluster 2. (d) Ancestry coefficient analysis of the Pst accessions. The accessions were grouped into two clusters with K = 2.

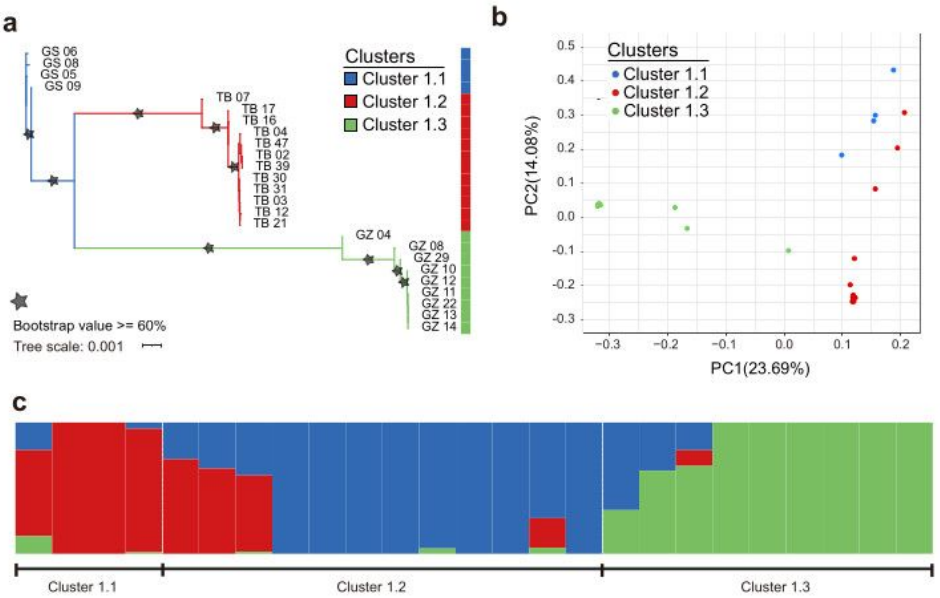


Figure 2

The population genetic structures of Puccinia striiformis f. sp. tritici (Pst) accessions in Cluster 1.

(a) The maximum likelihood phylogenetic tree of Pst isolates in Cluster 1. The clusters colored with blue, red, and green represent Cluster 1.1, Cluster 1.2, and Cluster 1.3, respectively. Stars on the branches indicated bootstrap values equal or greater than 60%. (b) Principal component analysis plot of the Pst

isolates within Cluster 1. The blue, red, and green dots represent different clusters. (c) Ancestry coefficient analysis of the Pst isolates in Cluster 1. The isolates were grouped into three clusters at K = 3.

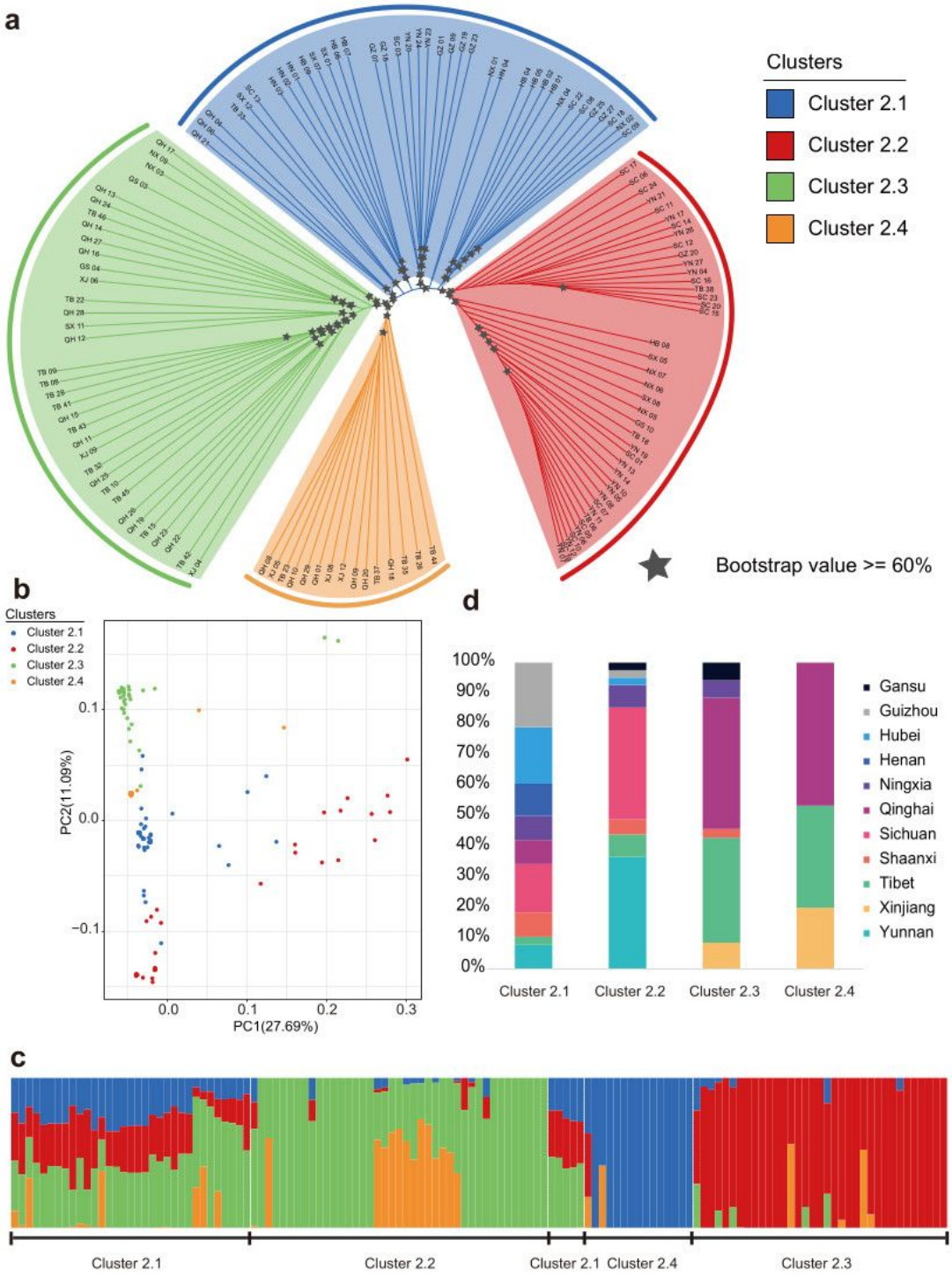
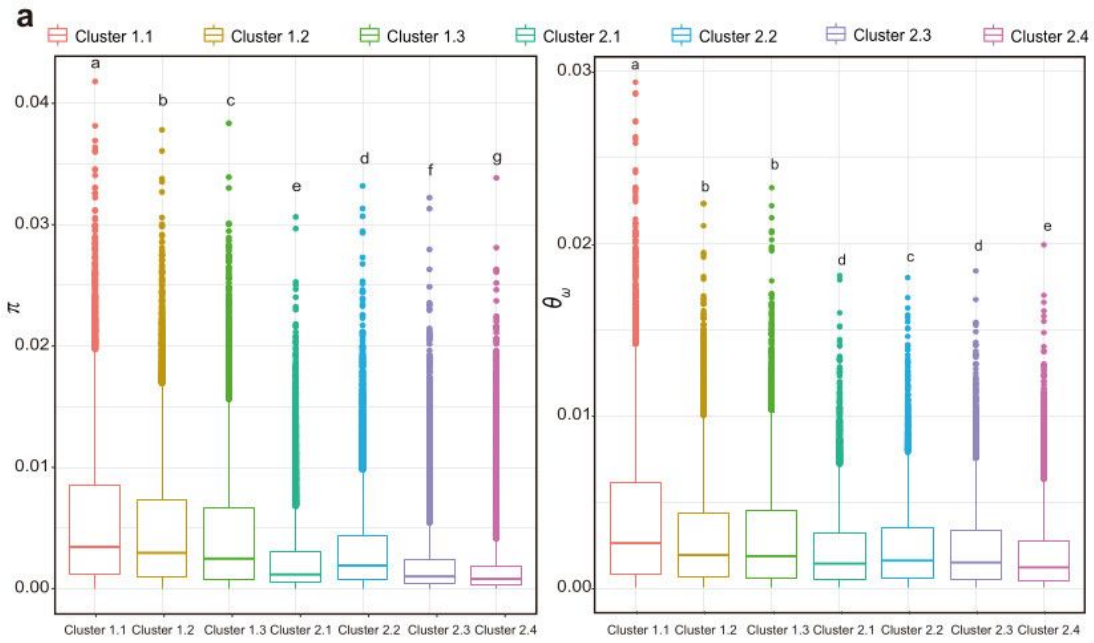


Figure 3

The population genetic structures of Puccinia striiformis f. sp. tritici (Pst) accessions in Cluster 2.

(a) Non-root phylogeny of Pst isolates in Cluster 2. Population structures colored with blue, red, green, and yellow represents Cluster 2.1, Cluster 2.2, Cluster 2.3, and Cluster 2.4, respectively. Bootstrap values greater than 60% were marked with stars. (b) Principal component analysis plots of the Pst isolates in Cluster 2 using single-nucleotide polymorphisms. Different clusters are shown in different colors. (c) Ancestry coefficient analysis of the Pst isolates in Cluster 2. The accessions were classified into four clusters with K = 4. (d) Stacked histogram of four clusters in Cluster 2. Proportion of Pst accession collected from different provinces are marked with different colors.



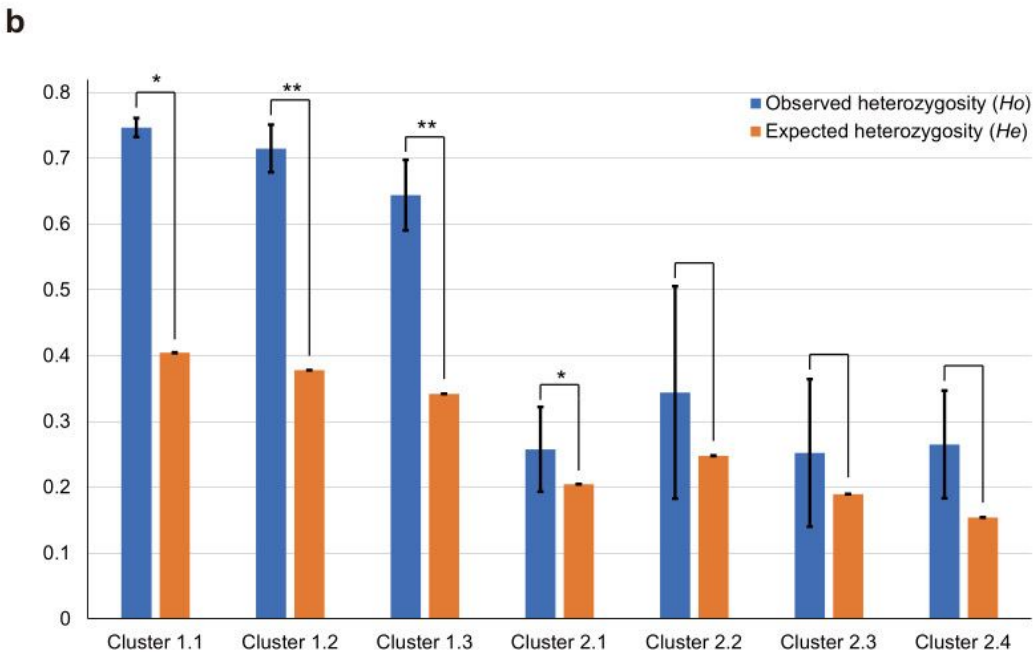


Figure 4

The genetic diversity and heterozygosity of Puccinia striiformis f. sp. tritici (Pst) sub-populations in China.

(a) The box plot of the nucleotide diversity of Pst subpopulations. Genetic diversity was tested with π and θω. Letters on the top of each box indicate significant levels of Kruskal-Walli's tests ($P < 0.05$). (b) The histogram of the observed and expected heterozygosity of each cluster. A permutation test was performed to examine the significant variation between observed and expected heterozygosity. P values were adjusted using the Benjamini Hochberg method. * represents P adjusted < 0.05 and ** means P adjusted < 0.01. Error bars represent the standard errors of the observed and expected heterozygosity for samples in each cluster.

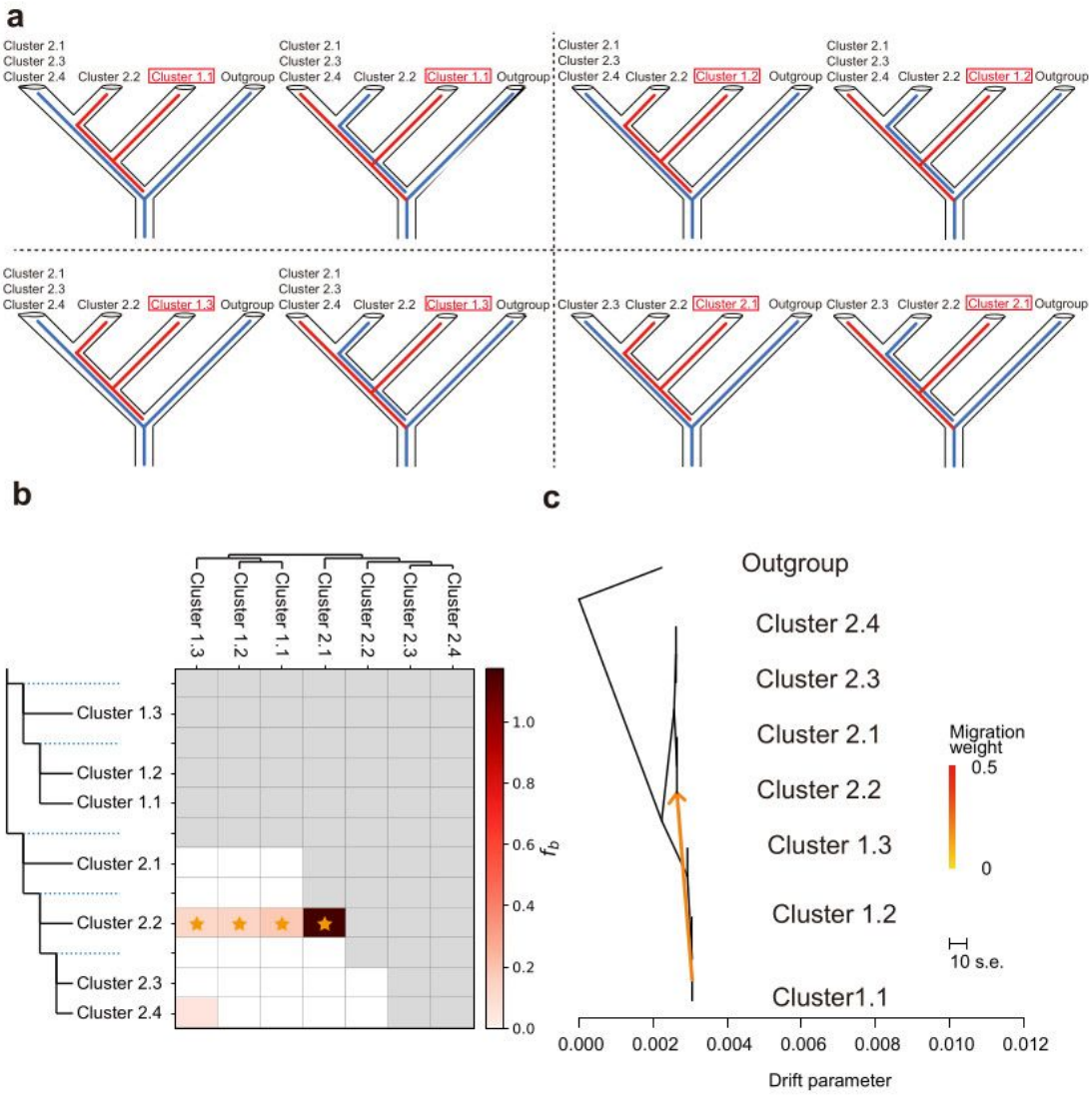


Figure 5

The genetic differentiation, introgression, and migration in Puccinia striiformis f. sp. tritici (Pst) sub-populations.

(a) The highly significant taxon topologies detected from D statistics (P adjusted < 0.001). Cluster 1.1, 1.2, 1.3, and 2.1 in red box of each group represent P3. The red and blue lines show the different

introgression patterns with P3. (b) The results of D statistics among Pst clusters. Rows represent nodes within the tree topology, and columns represent clusters regarded as P3. Each cell shows the fb values between a tree node (rows) and each cluster (column). Grey cells show the comparisons cannot be tested with D statistics. Stars marked on the cell shows significant tests at P adjusted greater than 0.001. (c) The migration event detected with TreeMix analysis. Migration arrows are colored according to their weight. Horizontal branch length is proportional to the amount of genetic drift that has occurred on the branch. Scale bar shows ten times the average standard error of the entries in the sample covariance matrix.

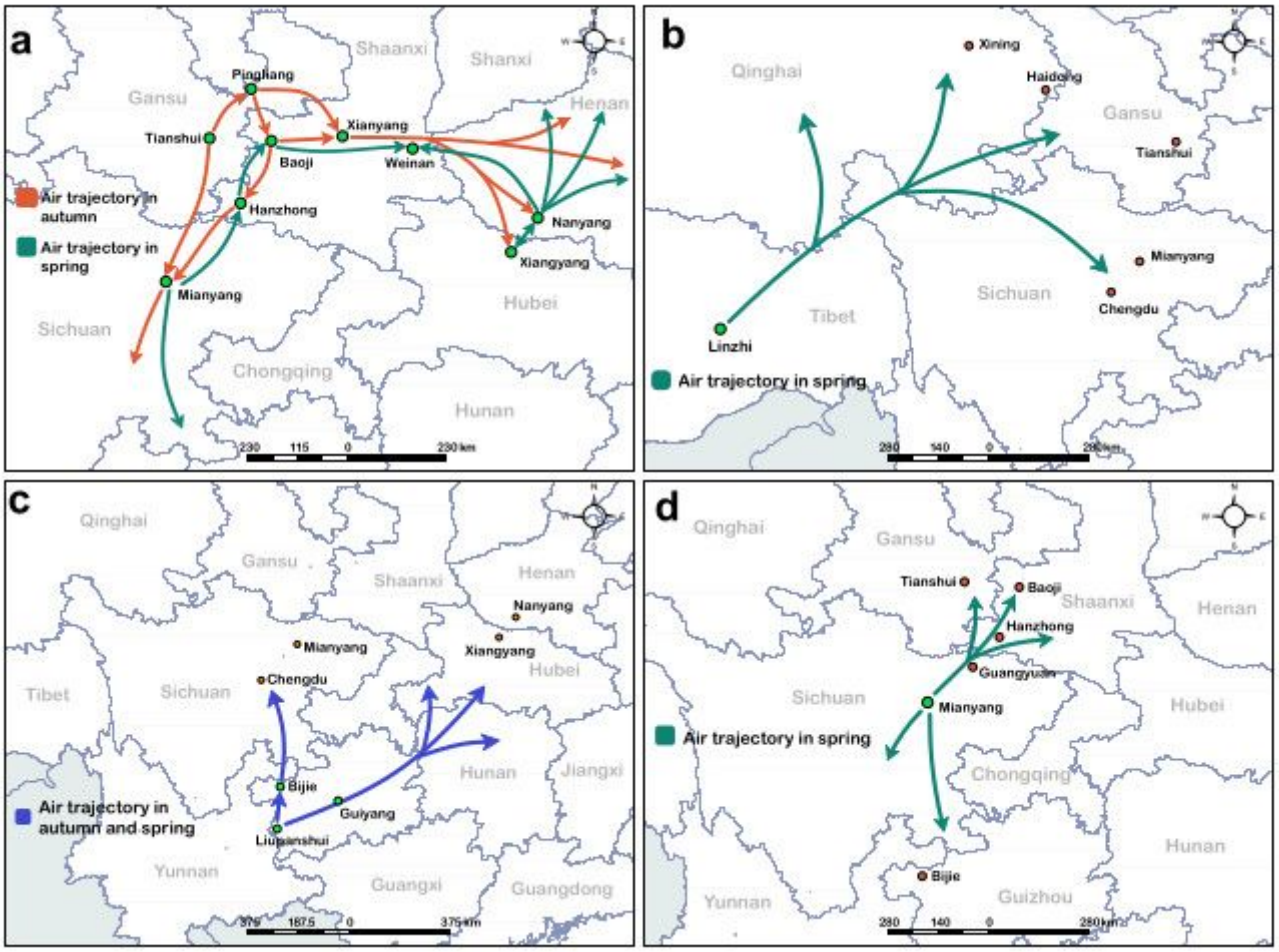


Figure 6

Trajectory analysis of Puccinia striiformis f. sp. tritici (Pst) migration routes in essential regions in China.

(a) The air trajectory within the major potential Pst epidemic regions. The green and orange arrows show the air trajectory in spring and autumn, respectively. The simulation period in Tianshui and Pingliang is October 1st to November 30th in autumn, in Mianyang is January 1st to April 10th in spring, and in Nanyang and Xiangyang is February 20th to April 10th. (b) The air trajectory from the Tibet regions in spring. The simulation period is April 1st to May 31st. (c) The aerial transportation from the Guizhou regions. The blue arrows show the air trajectory in spring and autumn. The simulation periods in Bijie and

Liupanshui are October 1st to November 30th in autumn and February 20th to April 10th in spring. (d) The migration of air trajectory from the Sichuan regions in spring with the simulation period of January 1st to April 10th.

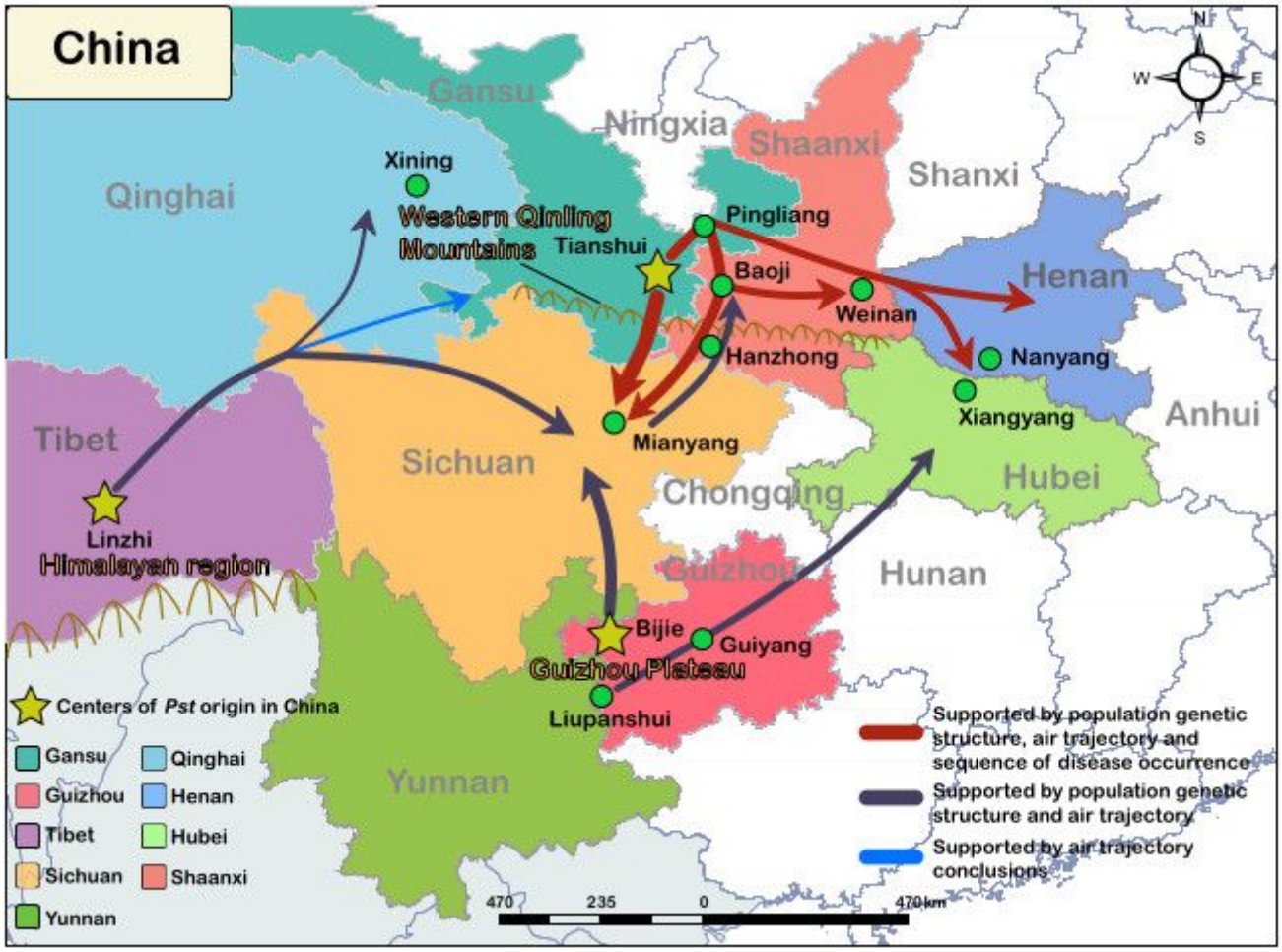


Figure 7

The major migration routes of Puccinia striiformis f. sp. tritici (Pst) in China.

Migration routes inferred from population genetics, air trajectory and disease occurrence. The red arrow represents the migration route supported by all analyses. The dark blue arrow shows the migration routes supported by population genetics and air trajectory. The green arrows designate the migration routes supported by disease survey and air trajectory. The blue arrow indicates the migration routes were solely supported by air trajectory. The width of each arrow shows the possibility of the migration routes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable125.xlsx

- SupplementaryFiguresandTables.pdf
- codesubmission.zip
- reportingsummar.pdf