Russian Wheat Aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction?

B. ZHANG[1, 2, 3], O. R. EDWARDS[3, 4], L. KANG[2*], S. J. FULLER[1, 3]

[1]Faculty of Science & Technology, Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia; [2]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; [3]Cooperative Research Centre for National Plant Biosecurity, LPO Box 5012, Bruce, ACT 2617, Australia; [4]CSIRO Ecosystem Sciences, Centre for Environment and Life Sciences, Underwood Avenue, Floreat, WA 6014, Australia.

Keywords: *Diuraphis noxia*, native range, geographic isolation, genetic structure, population expansion, wheat domestication

Running title: Population genetics of *Diuraphis noxia*

***Correspondence**:

Le Kang, email: lkang@ioz.ac.cn; fax: +86-10-64807099

Abstract

In this study we explore the population genetics of the Russian wheat aphid (*Diuraphis noxia*), one of the world's most invasive agricultural pests, in northwestern China. We have analyzed the data of ten microsatellite loci and mitochondrial sequences from 27 populations sampled over two years in China. The results confirm that the Russian wheat aphids (RWAs) are holocyclic in China with high genetic diversity indicating widespread sexual reproduction. Distinct differences in microsatellite genetic diversity and distribution revealed clear geographic isolation between RWA populations in northern and southern Xinjiang, China, with gene flow interrupted across extensive desert regions. Despite frequent grain transportation from north to south in this region, little evidence for RWA translocation as a result of human agricultural activities was found. Consequently, frequent gene flow among northern populations most likely resulted from natural dispersal, potentially facilitated by wind currents. We also found evidence for the long-term existence and expansion of RWAs in China, despite local opinion that it is an exotic species only present in China since 1975. Our estimated date of RWA expansion throughout China coincides with the debut of wheat domestication and cultivation practices in western Asia in the Holocene. We conclude that western China represents the limit of the far eastern native range of this species. This study is the most comprehensive molecular genetic investigation of the RWA in its native range undertaken to date, and provides valuable insights into the history of the association of this aphid with domesticated cereals and wild grasses.

Introduction

Biological invasions have occurred in many ecosystems and have evoked concern in evolutionary ecology and biological conservation (Pysek *et al.* 2008), as they are an important factor influencing global change (Bright 1999). Comparative studies to examine an invasive species in both its introduced and native range can improve understanding of how a non-indigenous species shapes its new environment (Scott 2007). Such studies not only provide information on the basic biological characteristics of an invader, but can also provide knowledge of the genetic background of the founding population of an invasive species (Ross *et al.* 2003, 2007; Ross & Shoemaker 2008), the dispersal pattern (Goodisman *et al.* 2001) and the invasion pathway of a species throughout its introduced range (Bonizzoni *et al.* 2004). Data of this kind improve our ability to predict the array of evolutionary responses and impacts that may result, as well as the future distribution of the invasive species.

In this study, we analyze the population genetics of the Russian wheat aphid (RWA), *Diuraphis noxia* Kurdjumov, one of the world's most invasive agricultural pests, in western China. RWAs infest native grasses and cereal crops, however they are most noted for their potential to severely damage grains such as wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare*) and their capacity for rapid population growth (Smith *et al.* 2004; Burd *et al.* 2006; Jyoti *et al.* 2006). The native distribution of RWAs is believed to center on the Iranian-Turkestanian mountain range and extends to southern Russia, the Middle East, and central Asia (Kovalev *et al.* 1991), with the earliest documented record of damage coming from Ukraine in the early 1900s. RWAs gradually spread to most European and North African countries during the early part of the 20th Century at which time it gained recognition as an emerging global pest. It was during the 1970's and 1980's that RWAs began to rapidly spread, causing severe crop damage in major grain producing areas in Europe, Africa and the Americas (Kovalev *et al.* 1991; Stary 1999; Smith *et al.* 2004).

RWAs were first observed in north-western China in 1975 at Tacheng in the Xinjiang Uyghur Autonomous Region (Zhang *et al.* 1999a). RWAs have not been detected in any other province in China. There is some dispute as to whether the RWA is an exotic or native species in China, with most Chinese entomologists regarding it as an invasive pest (Zhang *et al.* 1999a, b), possibly because it was around this time that invasive populations of RWAs were first reported in South Africa (1978), Mexico (1980), North America (1986) and South America (1988).

In recent years most research on RWAs has focused on documenting the biology and genetics of this species in its invasive range (Shufran *et al.* 2007; Shufran & Payton 2009; Liu *et al.* 2010) and much emphasis has been placed on documenting variant biotypes and discovering resistance genes in wheat and barley cultivars (Puterka *et al.* 1992; Basky 2002; Haley *et al.* 2004; Burd *et al.* 2006). Population genetic studies on RWAs from central Asia, including China, have not been undertaken. A significant body of research does exist however on the biology of this species in China. RWAs exhibit a holocyclic life cycle in China (Zhang *et al.* 1999a) with parthenogenesis the predominant mode of reproduction in late spring and summer, and sexual reproduction occurring in October. Cold-resistant eggs are laid in late October which over-winter on the basal leaves of the host plants (Zhang *et al.* 1999a). Invasive populations of RWA have been characterized as primarily anholocyclic (obligatory parthenogenetic), although the appearance of sexual females and eggs has been reported recently in North America and Argentina (Kiriac *et al.* 1990; Clua *et al.* 2004).

Host plants of RWA include cultivar crops, such as wheat, barley, and oats, and native grasses, wild oats and rye. Variable population growth rates and relative virulence on wheat and barley have been reported amongst invasive populations of RWA (Puterka *et al.* 1992; Basky 2002; Smith *et al.* 2004; Jimoh *et al.* 2011), however little is known about the level of

host adaptation in native populations of RWA. Host-based adaptation has been reported in other aphid species (Ferrari *et al.* 2006; Charaabi *et al.* 2008; Peccoud *et al.* 2009), and in the greenbug (*Schizaphis graminum*), another cereal aphid, mitochondrial data suggest that genotypes associated with cultivated cereals have a single origin (Shufran *et al.* 2000). Parthenogenetic reproduction is thought to facilitate sympatric host specialization in aphids (Sunnucks *et al.* 1997); parthenogenesis is also likely a key factor leading to the dominance of single genotypes ("superclones") across space and time (Abbot 2011).

It is not yet clear what biological, genetic and/or ecological factors are responsible for RWA invasiveness, and which factors are limiting its range expansion after establishment. RWAs quickly spread through most of the wheat growing districts in the western USA soon after its introduction in 1986, but did not expand its range significantly to the east (Smith *et al.* 2004). Large-scale dispersal is important in facilitating the expansion of aphid populations in both their native and invasive ranges (Dolatti *et al.* 2005; Michel *et al.* 2009; Shufran & Payton 2009). Aphid dispersal morphs (alatae) exhibit weak flying ability (Loxdale *et al.* 1993; Zhang *et al.* 2008), with most movement across long distances attributed to wind-aided dispersal (Venette & Ragsdale 2004). Monitoring insect movement using traditional ecological methods is problematic (Roderick 1996). Genetic methods are now used widely to examine the levels of migration among populations and provide answers to a range of questions relating to movement patterns and population demographic history.

Here we report results of the most comprehensive population genetic study yet undertaken on RWAs. We investigate the patterns of spatial and temporal genetic differentiation among sampled populations and infer possible dispersal mechanisms. We provide evidences for historical demographic population expansion throughout western China and predict the potential for future expansion of this species in other wheat growing districts with similar geographic features in China.

Materials and Methods

Aphid sampling

Russian wheat aphids (RWAs) were collected from wheat fields (*Triticum aestivum* L.) in northern and southern Xinjiang including desert, oasis and mountain foothill regions. In total, eighteen sites were sampled including fifteen in the north and three in south, from May to June of 2009 and 2010 (Fig. 1 & S1). Nine sites were sampled in both years to provide a temporal comparison. Up to fifty colonies were identified at each site and one parthenogenetic, wingless female aphid was collected from each plant. Consecutive samples at a location were collected a minimum of 50 meters apart, or in different fields, to minimize the chance of sampling aphids from the same colony. RWA specimens were preserved in 100% ethanol until DNA extraction.

DNA extraction and amplification

Total genomic DNA was extracted from single adult aphids using a salting-out method (Sunnucks & Hales 1996). All RWAs were screened for 12 microsatellite loci, including three cross-species loci developed from *Sitobion* aphids (Sa4Σ – Simon *et al.* 1999; Sm11 – Wilson *et al.* 1997; Sm23 – Wilson *et al.* 2004), and nine loci newly developed from RWAs. Microsatellite loci were amplified in a total volume of 10μl containing 10 nmol of fluorescent-labeled primers (Sangong Company, China), 0.5 U Taq, 1× PCR Buffer, 0.3 mM each dNTP, 2mM $MgCl_2$ (TaKaRa Taq™, Takara Biomedical) and 20ng of aphid DNA. PCR cycling conditions followed Shufran and Payton (2009), except that different annealing temperatures were used. Electrophoresis of the amplification products was conducted in a capillary sequencer ABI3730×1 (Applied Biosystems), with an internal size ladder (500 LIZ). Allele sizes were analyzed using GeneMapper (version 3.0, Applied Biosystems) and allele designation was confirmed following visual examination.

We also sequenced two mitochondrial DNA regions: partial cytochrome oxidase I (CO1) and a

continuous fragment centered on NADH dehydrogenase subunit 6 (ND6), including partial NADH subunit 4L, two tRNA genes, total ND6, and partial cytochrome B. The 436bp CO1 gene was amplified using the primers C1-J-1718 and C1-N-2191 (Simon *et al.* 1994), and the ND6 fragment (837bp) was amplified using the primers N4L-J9648 and CB-N10608 (Simon *et al.* 2006). The PCR protocol and cycling conditions followed Shufran & Payton (2009), except that ExTaq (TaKaRa Taq™, Takara Biomedical) was used. PCR products were purified using an ABgene Ultra PCR Clean-Up Kit (Thermo Scientific) and run on an ABI3130 sequencer.

Genetic Diversity

Genetic diversity estimates were calculated using FSTATv2.9.3. (Goudet 2001) and included: observed and expected heterozygosity (Ho & He), allele size range, number of alleles (Na), allelic richness (Ar), and the *f* estimator of *F*is and significance values (Weir & Cockerham, 1984). Allele frequencies, Hardy-Weinberg equilibrium (HWE), and linkage disequilibrium tests were calculated using Genepop v4.0 with 1000 iterations and 100 Markov Chain approximations (Raymond & Rousset 1995; Rousset 2008). Significance was assessed following Bonferroni correction (Rice 1989). Micro-Checker v2.2.3 was used to test for large allele dropout (Van Oosterhout *et al.* 2004). Null allele frequencies for each locus were estimated using Cervus v3.0 (Marshall *et al.* 1998). All individuals were also classified according to multilocus genotype (MLG) in GenClone v2.0 (Arnaud-Haond & Belkhir, 2007). Genetic diversity was analyzed based on gross genotypic diversity (GGD), which was calculated as G/N, with G equal to the number of MLGs, and N equal to the sample size (Llewellyn *et al.* 2003).

Mitochondrial DNA sequences were aligned and edited using BioEdit v7.0.0 (Hall 2004) and MEGA v4.1 (Tamura *et al.* 2007). The number and frequency of haplotypes were calculated using DnaSP v5 (Librado & Rozas 2009), and a phylogeographic network was inferred using TCS (Clement *et al.* 2000). We also calculated Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997)

implemented in Arlequin v3.5.1.2 (Excoffier *et al.* 2005) to infer deviations from neutrality and to detect demographic changes or selection (Fu & Li 1993).

Genetic Differentiation

Pairwise *F*st estimates were calculated from the microsatellite data using Arlequin and exact G tests of allelic differentiation were calculated using Genepop. The datasets were analyzed by year, and one site, Fuhai, was excluded because of low sample size. A Mantel test implemented in Genepop (using 10000 permutations) was used to examine whether there was a relationship between *F*st and geographic distance. The sampling coordinates were recorded in GPS, and the straight-line distance between each pairwise locality was calculated using Google Earth (Google inc., Mountain View CA).

Three clustering methods were used to identify population structure. Firstly, a Bayesian Markov Chain Monte-Carlo (MCMC) method implemented in Structure v2.1 (Pritchard *et al.* 2000) was used. An admixture model was assigned by assuming independent allelic frequencies with 100,000 iterations of MCMC after a 20,000 burn-in period, and ten independent runs for each *K* were evaluated. To select the most likely *K* value, we adopted two criteria: first, the *K* reached a plateau in the Ln(*K*) plot, and $\triangle K$ attained its maximal value (Evanno *et al.* 2005); and second, a parsimony method was used in which the lowest *K* is selected that captures most differentiation among populations (DiLeo *et al.* 2010). We then used Distruct v1.1 (Rosenberg 2004) to display the bar plot under the most likely *K* value. Secondly, factorial correspondence analysis (FCA) was carried out in Genetix v4.05 (Belkhir *et al.* 1996-2004) to examine the three-dimensional spatial distribution of genetic variation for each individual. Finally, an analysis of molecular variance (AMOVA) was conducted in Arlequin to confirm population clusters and to differentiate the variation component among populations and years.

We used the microsatellite data to examine evolutionary scenarios of expansion and gene flow among sites using DIYABC v 0.7 (Cornuet *et al.* 2008), MIGRATE v3.2.7 (Beerli 2008) and BayesAss 1.3 (Wilson & Rannala 2003). DIYABC estimates the posterior distributions of different evolutionary scenarios by generating simulated data and comparing selected simulated data that are closest to the observed data (Cornuet *et al.* 2008). Five scenarios of simultaneous expansion were examined using four geographically widespread sites - Qapqia (Yili Valley, north-west Xinjiang), Yumin (north-west Xinjiang), Mori (north-east Xinjiang) and Wuqia (south Xinjiang) - and an unsampled site as the origin of expansion. We assumed a stable effective population size (N_e), a transitory bottleneck (db=5) and a generalised stepwise model (GSM) of mutation. 250000 simulated datasets were produced for each scenario and the 15000 closest simulations to the observed data were compared using logistic regression.

MIGRATE detects gene flow over historical timescales - up to $4N_e$ generations in the past. It is implemented using a maximum likelihood model with two long chains, followed by ten short chains recorded at the sampling increment of 100 iterations, and with a burn-in of 10000 iterations. The program was run five times using different random seeds. BayesAss estimates recent migration rates with 95% confidence intervals. Five independent runs with different initial random seeds were undertaken using 20 million iterations and a 10 million burn-in chain to check the congruence.

Demographic Changes in Population Size

Changes in demographic history are known to affect the frequency of alleles, the distribution of mutations, and the coalescent times of gene copies. Two tests were used to determine whether the microsatellite data displayed any signature for past population expansion or contraction. Firstly, using the program Bottleneck v1.2.02 (Cornuet & Luikart 1996), observed and expected heterozygosity were compared to detect any heterozygote excess (Piry *et al.*

1999). We also used Bottleneck to test for mode-shift. Secondly, *k* and *g* tests were used to detect any signal of population expansion in the ancestral generations (Reich & Goldstein 1998; Reich *et al.* 1999; Bilgin 2007). Negative *k* values at each locus indicate population expansion. A low value of *g* (under 1) can be interpreted as evidence of population expansion.

The mitochondrial data were also examined for evidence of population expansion using a pairwise mismatch distribution implemented in Arlequin. The goodness-of-fit of the observed data to a simulated model of expansion was tested with the sum of squared deviations (SSD) and raggedness index. The age of expansion was estimated with the formula ô = 2ìt, where ì equals the aggregate mutation rate across all nucleotides per generation and t is the expansion time in generations. We also adopted Ramos-Onsins and Rosas's R2 test (Ramos-Onsins & Rozas 2002) in DnaSP to complement the power of the pairwise mismatch distribution. The R2 test was conducted using coalescent simulations with 1000 replicates and 95% confidence intervals.

Results

Genetic Diversity

Twelve microsatellite loci were screened for 1040 RWA colonies sampled in western China in 2009 and 2010. Two of the cross-species loci (Sm11 and Sm23) were discarded as a high number of scoring errors were detected. The remaining ten loci were polymorphic (Appendix S2) and could be confidently scored (i.e. no large allele dropout or scoring errors were detected using Micro-Checker). Only one locus (Dn1) was potentially affected by null alleles, having a null allele frequency greater than 0.1, however no significant departure from HWE was found for this locus. Significant deviation from HWE was identified in five of the 27 tests as a result of heterozygote deficit or excess. Although a small proportion of linkage disequilibrium tests indicated significant linkage, no consistent pattern between any particular pair of loci was evident, therefore the ten loci are providing independent assessments of genetic variation.

Within each site, the highest allelic number and richness was found in Haba, with 11.1 and 6.08 respectively (Table 1). In contrast, the lowest allelic number was found in Pishan with 2.7, and lowest allelic richness in Cele with 2.51 (Table 2). Sites located in northern Xinjiang, including the regions surrounding Tacheng, Altay and Urumqi, presented similar average gene diversities during both years. An ANOVA revealed that sites in the south had significantly reduced gene diversity ($F=3.68$, df=3,22, $p=0.027$) and allelic richness ($F=5.36$, df=3,21, $p=0.007$) compared with the north.

A total of 928 MLGs were identified from 1040 RWAs based on the data from ten microsatellite loci (Table 1 & 2). The number of MLGs shared within a site ranged from 0 to 8, with the highest sharing occurring in Cele. Four sites were entirely composed of unique MLGs. Interestingly only one MLG was shared among sites (between two individuals from Pishan and Cele). No MLGs were shared among years at any site.

Concatenated, 1272bp of mitochondrial DNA was obtained from 178 RWAs. Eighteen haplotypes were identified, with one common haplotype found at all sites (relative frequency: 88.8%), and seventeen rare haplotypes found at low frequencies (0.5-1.1%). Three haplotypes were shared among sites: Hap1 (universal), Hap3 (found at Yumin and Qapqia), and Hap7 (found at Haba and Toli). Hap10 was found in two individuals from Wuqia (Fig. 2). The remaining fourteen haplotypes were unique to one site. Mori in north-east Xinjiang had the highest nucleotide diversity as well as significant Tajima's D and significant Fu's Fs values (Appendix S3). Twenty variable sites were found and although eighteen of these occurred among protein coding regions, the majority of single base pair mutations were transitions (12/18) and synonymous mutations (13/18).

The TCS network revealed a star-like pattern centered on the widely distributed Hap1 (Fig. 2). From the central haplotype (Hap1), fifteen haplotypes diverged by one mutation, one haplotype (Hap13) diverged by two mutations and another haplotype (Hap9) diverged by three mutations.

Genetic Differentiation (nDNA)

Population differentiation was analyzed using pairwise *F*st values and exact tests of allelic differentiation. In 2009, pairwise *F*st values among northern sites were generally low, ranging from 0.0055 to 0.1129 (Table 3A). The majority of pairwise comparisons of *F*st among northern sites were significant indicating restricted gene flow between sites. In 2010, the majority of pairwise comparisons of *F*st were significant among northern sites again, however, more importantly a much higher level of differentiation was detected between northern and southern sites (Table 3B). The highest *F*st value was between Cele in the south and Berqin in the north (*F*st=0.3768), and the average *F*st values of southern sites (Wuqia, Pishan, and Cele) to the other eleven northern sites were 0.112, 0.16, and 0.266, respectively. Furthermore, the

pairwise *F*st value between the two southern sites, Wuqia and Cele, was also very large 0.27. These data indicated that gene flow is considerably restricted among southern sites and between northern and southern sites.

Mantel tests based on the 2009 data (only northern sites were sampled) did not reveal a significant correlation between *F*st and geographic distance (r=0.25, *p*=0.17). However in 2010, both northern and southern sites were sampled and a strong pattern of isolation by distance was detected (r=0.57, *p*<0.0001).

An AMOVA was conducted using 2010 data and separating sites into three groups (1. Wuqia, 2. Cele and Pishan, and 3. northern sites). The proportion of variance among groups (12.42%) was larger than that found among sites within groups (4.43%), and the fixation index (*F*ct=0.124) was significant, indicating extremely restricted gene flow among the three groups (Appendix S4). We also analyzed temporal differentiation among the nine sites that were sampled in both 2009 and 2010. Pairwise *F*st and exact tests revealed significant differentiation between years in all populations except Emin (Appendix S5). Genetic variation between years resulted in a fixation index (*F*sc=0.028) greater than that for among sites (*F*ct=0.007), suggesting that more structure exists within a site when sampled from one year to the next than among sites sampled within a single year.

Population Structure

Similar patterns of hierarchical structure were obtained using individual-based clustering in Structure and three-dimensional factorial correspondence analysis (FCA). Both methods revealed three clusters (k=3) among northern sites sampled in 2009 (Fig. 3A, Appendix S6A). However, no distinct groups could be discerned that corresponded to any of the 13 sites, indicating that all individuals sampled were of mixed ancestry. Further increasing k in Structure did not reveal any distinct subdivisions. An analysis of 2010 data using Structure revealed four

clusters corresponding to three regions with distinctive population groups: 1) Wuqia, 2) Cele and Pishan, and 3) all other northern sites (Fig. 3B). The FCA analysis also identified the three southern sites as distinct from the northern sites, with Pishan genetically intermediate between Cele and Wuqia (Fig. 4). The three axes explained over 50% of the variation among the sites. Structure (k=2) and FCA identified a varying degree of admixture amongst the northern populations in 2010 (Fig 3B, Appendix S6B).

Evolutionary scenario testing using DIYABC revealed higher posterior probabilities for simultaneous expansion from the three northern sites analysed (Qapqia: 0.370, 95% CI 0.283-0.456; Yumin: 0.365, 95% CI 0.279-0.451; Mori: 0.235, 95% CI 0.169-0.302) than from southern Xinjiang (Wuqia: 0.005, 95% CI 0.002-0.007) or an unsampled alternative (0.025, 95% CI 0.014-0.0037). Yumin and Qapqia abut the border with Kazakhstan and showed slightly higher posterior probabilities than Mori (north-east Xinjiang) as being the expansion origin. Similarly, MIGRATE estimates of long-term gene flow were significantly asymmetric based on non-overlapping 95% confidence intervals (Appendix S7), indicating that Yumin and Qapqia may be expansion origins. Additionally, the most divergent mitochondrial haplotype was found at Qapqia further suggesting that this site may represent the ancestral origin of RWAs in China. Given the low level of haplotype sharing detected (only three haplotypes shared out of 18), it is interesting to note that Yumin and Qapqia shared haplotype 3 (Fig. 2). However, when we used BayesAss to look for evidence of recent gene flow between north and south Xinjiang, no trace of migration was detected among Yumin, Qapqia, Mori and Wuqia (non-migration rate: 0.833, 95% CI 0.675-0.992).

RWA Population Demographic History

Population demographic history examined using Bottleneck and Kgtest displayed little evidence for past population fluctuation (Table 4). Significant heterozygote deficits were only detected at three sites. Therefore, the reduction in allele number within populations was

348 probably due to founder events rather than rapid decline in population size. Likewise, the
349 L-shaped mode of allele frequency distribution suggests a long-term stable population size.
350 Furthermore, the *k* test was not significant for most sites indicating that the allele length
351 distribution was not significantly different from a binomial distribution and that the population
352 size has been steady. The *g* tests were also not significant providing further evidence of stable
353 population size. However, when considering all 18 sites as one population, the *k* test indicated
354 that significant population expansion has occurred throughout western China. Although the *g*
355 test value was not significant, it was less than one, thus supporting the conclusion of past
356 population expansion in western China.
357
358 The mtDNA data also provided evidence of rapid demographic expansion, with the universal
359 haplotype at the center of a star-like cluster formed by the 17 rare haplotypes (Fig. 2).
360 Furthermore, the pairwise mismatch distribution was unimodal, with a strong peak evident at
361 zero, which steeply declined from zero to one base pair. The goodness-of-fit tests were not
362 significant (p(SSD)=0.52 and p(Harpending's RI)=0.68), and evidence for highly significant
363 population expansion was detected in the R2 statistic (R2=0.08347, p=0.002), Tajima's D
364 (D=-2.39352, p<0.01), and Fu's Fs (Fs=-28.395, p<0.0001). The estimated generation time
365 since expansion for Chinese populations was approximately 3,200 years, based on ô value of
366 0.146 and 1.77%/MY as mutation rate based on the rate given by Papadopoulou *et al.* (2010)
367 for beetle mtDNA.
368

Discussion

This study has investigated the population genetics, demographic history and evolutionary adaptation of the Russian wheat aphid in its rarely investigated, far eastern native range in China. We have also rejected the hypothesis that this invasive pest had been introduced into Western China in the last couple of decades. An understanding of the levels and patterns of genetic variation in native populations can provide valuable insights into the factors that have facilitated the recent global invasion by this damaging pest species.

Genetic Diversity of RWAs in China

The microsatellite data revealed high genetic diversity and large numbers of MLGs. No MLGs were shared between two consecutive sampling years at any single site and very few MLGs were shared within and among sites in the Xinjiang region, strongly supporting previous research that sexual reproduction is prevalent in China (Zhang *et al.* 1999a). High population densities of RWAs in China, together with little, recent migration among sites may have also contributed to the high genetic diversity found in this study.

Consistently, our findings revealed significantly higher genetic diversity of RWAs in northern sites compared with southern, suggesting limited gene flow among and possible founder events in southern sites. A gradual reduction in genetic diversity and gene flow was evident, declining from Wuqia, the most northerly of the southern sites, to Pishan and Cele (the most southerly located site). Of all the sites sampled, Cele was the least diverse having the lowest allelic richness and a number of MLGs shared among individuals within the population. From this, we surmise that the population in Cele was probably founded by very few RWAs – possibly colonising from Pishan. In contrast, the northern sites exhibited roughly equivalent levels of microsatellite variation. While the mtDNA data were generally less informative due to low levels of variation, one site in the north-east (Mori) displayed the highest diversity.

Genetic diversity within a site was correlated with geographic location and latitude; northern sites had higher diversity than southern sites. One possible explanation is that different patterns of introduction and establishment of RWAs occurred in the two regions. Given that ecological and environmental conditions in the north and south are quite different, RWAs would have experienced different selection pressures, potentially on different hosts and different ecological conditions influenced by climate and geography. In southern Xinjiang, microclimatic variation will have a strong effect on RWA populations as they occur in mountain regions above 2000 meters elevation (even above 3300 meters in Taxkorgan; Du 2000). In northern Xinjiang, RWAs occur at elevations ranging from 700-1000 meters, mostly on plains or flat areas. Broad (or macro) scale fluctuations in climate will have a greater influence in the north and elevation is less likely to be a barrier to insect dispersal or migration compared with the south. Furthermore, grain fields in the south are predominantly cultivated in small patches (ie. oases) that are discontinuously located along the edge of the Taklamakam Desert and the Tarim River basin. Conversely, cultivated fields and wild grasslands are continuously distributed along the northern slope of the Tianshan Mountain range, offering RWAs a selection of host plants on which they can live or use as stepping stones to migrate. Finally, in southern Xinjiang farmers plant only winter wheat and have one wheat-growing season per year, while in northern Xinjiang farmers plant both winter and spring wheat each year, with an overlapping growth season from April to June. As a result, RWAs can persist over longer time periods in the north and because of plentiful food resources their survival and reproductive success may be enhanced.

The high genetic diversity observed at microsatellite loci contrasted markedly with the low level of mtDNA genetic diversity that we observed in the Chinese RWA populations. Only eighteen haplotypes were identified from 178 RWA individuals, and seventeen of these were rare and found at very low frequency. This level of mtDNA diversity is still much higher than that found in

invasive RWA populations, which have no mtDNA variation (Shufran *et al.* 2007; Shufran & Payton 2009). In other aphid species, anholocyclic populations have mitochondrial haplotypes that are distinct from holocyclic populations, and often exhibit reduced mtDNA diversity (Martinez-Torres *et al.* 1997).

Gene Flow among RWA Populations in Xinjiang

All methods of population structure analysis used in this study provided unequivocal support for strong differentiation among Chinese RWA populations relative to geography. Little evidence of gene flow between northern and southern regions was found. The Tianshan Mountain range segregates Xinjiang into northern and southern regions and the dominant wind direction is from west (Siberia) to east (China). The wind from north to south across the mountain range is weak and unlikely to facilitate passive RWA dispersal and although not conclusive evidence, RWAs have not been found along the southern slope of the Tianshan Mountains. However, aphids have been found suspended in air currents and are thought to be capable of long distance (100's of kilometers) flight (Dixon 1998; Delmotte *et al.* 2002). In this study, the low level of gene flow between northern and southern Xinjiang suggests that RWAs probably have a low active flying capacity and this may be due to demographic or behavioural factors.

Experiments have shown that live adult RWAs can survive and produce a viable colony after three days without food and water (Vitou and Edwards unpublished data). Therefore, it cannot be discounted that live adult RWAs may be transported on seedlings or human artifacts over long distances. In fact, wheat seeds are transferred frequently between northern and southern Xinjiang as Yili and Tacheng have wheat breeding centres that provide on an annual basis, high-quality improved seeds to wheat growers located throughout Xinjiang ("Greater Mekong Subregion Agricultural Information Network"). Because of high shipping costs, forage grass species or wheat seedlings are not transferred between northern and southern regions.

Consequently, as we detected little evidence of short-term gene flow from north to south, RWAs are probably not frequently transported by human agricultural activities. As more wheat fields are planted, the possibility remains however, that over time, aphid populations may expand into new areas via natural pathways (flight or wind currents).

Historical Expansion of RWAs in China

The accepted opinion is that the original native eastern distribution of RWAs included northern Kazakhstan (Kovalev *et al.* 1991) and therefore, it is logical to suppose that RWAs could have been present along mountain ranges from central Asia (ie. Kazakhstan) to western China before they were first detected in the 1970s. Our study has provided strong evidence for a long-term association of RWAs with wheat and possibly other cereals in western China.

Our mtDNA data indicate a relatively recent population expansion of Chinese RWAs during the last three thousand years. Although this estimate only provides an approximation, it is concordant with historical climate change events in central Asia and the spread of cereal domestication and cultivation practices. During the last 11000 years, the warm wet climate of the Holocene (Richerson *et al.* 2001) provided a relatively stable, warm, and CO_2-rich environment facilitating rapid plant growth. During this time, plant domestication and associated cultivation spread rapidly. Wheat domestication was first recorded in the Fertile Crescent (including the modern day Turkey, Iran, Iraq, Syria, Lebanon, Jordan, Palestine and Israel) in 9500-7500BC (Bellwood, 2001; Diamond, 2002) and spread eastward to central Asia by 7000-6000BC, to north-western China by 4600-2000BC (Li *et al.* 2007; Thornton & Schurr 2004) and then to the Indian subcontinent by 3,500-3,000BC (An *et al.* 2005). The earliest published record of wheat in Xinjiang comes from 2000BC (Thornton & Schurr 2004), a point in time when the Silk Road first became an active conduit for trade and agriculture between western and eastern Asia. We hypothesize that the expansion of RWAs in western China suggested by our mtDNA results was facilitated by agricultural activities associated with the

476 human domestication of wheat.

477

478 Our microsatellite data also revealed a signal of population expansion when all sites were
479 combined. Most sites displayed a very slight growth trend, indicating long-term co-evolution of
480 the RWA with its host in natural habitats. Thus, our data are consistent with the theory that
481 long-term effective population size should be in general, closer to the actual size during the
482 remission period than that in the initial expansion and growth period (Motro & Thomson 1982).
483 In addition, high gene flow among populations of RWA in the north during the expansion and
484 growth period probably enhanced the homogenizing effect, as has been found during an
485 outbreak event of the migratory locust, *Locusta migratoria* (Chapuis *et al.* 2009).

486

487 Our results from the mtDNA and microsatellite data are difficult to reconcile. The high gene
488 flow we observed among northern Xinjiang RWA populations indicates that there should also
489 be gene flow with populations in neighbouring Kazakhstan, which all available evidence
490 suggests is within the native range of RWAs (Kovalev *et al.* 1991). If so, why would the mtDNA
491 point to a recent population expansion? It is possible that RWAs did not exist in Xinjiang before
492 the arrival of domesticated wheat. However, an alternative explanation is that the widespread
493 planting of domesticated wheat changed the population structure of RWAs across their entire
494 native range by selecting for wheat-adapted genotypes. Exclusively parthenogenetic
495 reproduction during the wheat growing season would facilitate the fixation of a single
496 wheat-adapted maternal lineage (a “superclone”), as has been observed in other aphid
497 species (Vorburger 2006; Abbot 2011; Harrison & Mondor 2011). Under this hypothesis, all
498 existing RWAs in Xinjiang and elsewhere in its native range would be descendents from this
499 original wheat-adapted haplotype – the dominant Haplotype 1 in our study. Additional samples
500 from throughout the native distribution of RWA should be analysed to further test this
501 hypothesis.

502

503 Given the potential capacity of RWAs to invade provinces other than Xinjiang, it is interesting
504 that the most easterly site in Xinjiang where RWAs have been detected in the past is
505 Qincheng, located near the border of Gansu province (Zhang *et al.* 1999a; Du 2000). Why
506 have RWAs failed to establish in more Eastern wheat growing districts in China, when the
507 climate is predicted to be conducive (Liang *et al.* 1999)? Though a geographic barrier (eg.
508 Gobi desert) may be responsible, it is also possible that the same environmental factors are
509 limiting range expansion eastward in both China and the USA, which may be an obligate
510 ecological association with high altitudes in areas where an overwintering stage is required
511 (John Burd, personal communication).
512
513 Finally, it is important to consider that in this study we have only sampled RWAs from wheat
514 and thus, we may have examined the genetic structure of only a subsample of the RWAs in
515 the region. Without sampling on other hosts, particularly perennial native hosts, we cannot
516 discount the possibility that we have missed additional unsampled genotypes in the region.
517 In addition, this study has examined the genetic differentiation of RWAs from only a relatively
518 small part of their native range in Asia. However, our results will be critical in guiding future
519 studies of patterns of invasion not only of RWAs, but also of other invasive insect herbivores.

Acknowledgements

We are grateful to Mr. Wensheng Zhang, Dr. Runzhi Zhang, Dr. Hongbin Liang, Mr. Decheng Ma, Mr. Yonglin Wen, and Mr. Bingren Du for cooperating in field sampling for two years. We thank Dr. Peter Mather, Dr. David Hurwood, Mr Vincent Chand and Dr. Chuan Ma for molecular analysis and discussion. We also thank three anonymous reviewers for their constructive comments. The authors acknowledge the support of the Australian Government's Cooperative Research Centre for National Plant Biosecurity, CSIRO, Queensland University of Technology, Chinese Academy of Sciences, and the National Basic Research Program of China (No.2012CB114105).

References

Abbot P (2011) A closer look at the spatial architecture of aphid clones. *Molecular Ecology*, **20**, 4587-4589.

An CB, Tang LY, Barton L *et al.* (2005) Climate change and cultural response around 4000 cal yr BP in the western part of Chinese Loess Plateau. *Quaternary Research*, **63**, 347-352.

Arnaud-Haond S, Belkhir K (2007) Genclone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, **7**, 15-17.

Basky Z (2002) Biotypic variation in Russian wheat aphid (*Diuraphis noxia* Kurdjumov Homoptera: Aphididae) between Hungary and South Africa. *Cereal Research Communications*, **30**, 133-139.

Beerli P (2008) MIGRATE version 3.0: a maximum likelihood and Bayesian estimator of gene flow using the coalescent. See http://popgen.scs.edu/migrate.html.

Bellwood P (2001) Early agriculturalist population diasporas? Farming, languages, and genes. *Annual Review of Anthropology*, **20**, 181-207.

Belkhir K, Borsa P, Chikhi L *et al.* (1996-2004) GENETIX 4.05, Logiciel sous Windows pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bilgin R (2007) Kgtests: a simple Excel Macro program to detect signatures of population expansion using microsatellites. *Molecular Ecology Notes*, **7**, 416-417.

Bonizzoni M, Guglielmino CR, Smallridge CJ *et al.* (2004) On the origin of medfly invasion and expansion in Australia. *Molecular Ecology*, **13**, 3845-3855.

Bright C (1999) Invasive species: pathogens of globalisation. *Foreign Policy*, **116**, 50-64.

Burd JD, Porter DR, Puterka GJ *et al.* (2006) Biotypic variation among North American Russian wheat aphid population. *Journal of Economic Entomology,* **99**, 1862-1866.

Chapuis MP, Loiseau A, Michalakis Y *et al.* (2009) Outbreaks, gene flow and effective population size in the migratory locust, *Locusta migratoria*: a regional-scale comparative survey. *Molecular Ecology*, **18**, 792-800.

Charaabi K, Carletto J, Chavigny P *et al.* (2008) Genotypic diversity of the cotton-melon aphid *Aphis gossypii* (Glover) in Tunisia is structured by host plants. *Bulletin of Entomological Research*, **98**, 333-341.

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1660.

Clua A, Castro AM, Ramos S *et al.* (2004) The biological characteristics and distribution of the greenbug, *Schizaphis graminum*, and Russian wheat aphid, *Diuraphis noxia* (Hemiptera: Aphididae) in Argentina and Chile. *European Journal of Entomology*, **101**, 193-198.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

583 Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a
584 user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713-2719.
585
586 Delmotte F, Leterme N, Gauthier JP *et al.* (2002) Genetic architecture of sexual and asexual
587 populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers.
588 *Molecular Ecology*, **11**, 711-723.
589
590 Diamond J (2002) Evolution, consequences and future of plant and animal domestication.
591 *Nature*, **418**, 700-707.
592
593 DiLeo M, Row JR, Lougheed SC (2010) Discordant patterns of population structure for two
594 co-distributed snake species across a fragmented Ontario landscape. *Diversity and*
595 *Distributions*, **16**, 571-581.
596
597 Dixon AFG (1998) Aphid ecology, 2nd edition. Chapman and Hall, London.
598
599 Dolatti L, Ghareyazie B, Moharramipour S *et al.* (2005) Evidence for regional diversity and
600 host adaptation in Iranian population of the Russian wheat aphid. *Entomologia experimentalis*
601 *et applicata,* **114**, 171-180.
602
603 Du BR (2000) Investigation of the distribution and suitable area of Russian wheat aphid,
604 *Diuraphis noxia* (Kurdjumov) in Xinjiang. *Xinjiang Agricultral Science,* **Z1**, 82-85.
605
606 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using
607 the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.
608

609 Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software
610 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.
611
612 Ferrari J, Godfray HC, Faulconbridge AS *et al.* (2006) Population differentiation and genetic
613 variation in host choice among pea aphids from eight host plant genera. *Evolution*, **60**,
614 1574-1584.
615
616 Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.
617
618 Fu YX (1997) Statistical methods of neutrality of mutations against population growth,
619 hitchhiking and background selection. *Genetics*, **147**, 915-925.
620
621 Goodisman MAD, Matthews RW, Crozier RH (2001) Hierarchical genetic structure of the
622 introduced wasp *Vespula germanica* in Australia. *Molecular Ecology*, **10**, 1423-1432.
623
624 Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices,
625 Version 2.9.3. Available from http:// www.unil.ch/izea/softwares/Fstat.html.
626
627 Greater Mekong Subregion Agricultural Information Network:
628 http://www.gms-ain.org/Z_Show.asp?ArticleID=1703
629
630 Haley SD, Peairs FB, Walker CB *et al.* (2004) Occurrence of a new Russian wheat aphid biotype
631 in Colorado. *Crop Science,* **44**, 1589-1592.
632
633 Hall T (2004) BioEdit v. 7.0.0. Available online at www.mbio.ncsu.edu/BioEdit/.
634
635 Harrison JS, Mondor EB (2011) Evidence for an invasive aphid “superclone”: Extremely low

636 genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States.
637 *PLoS One*, **6**, e17524.
638
639 Jimoh MA, Botha CEJ, Edwards O, Bradley G (2011) Population growth rate and relative
640 virulence of the two South African biotypes of Russian wheat aphid*, Diuraphis noxia*, and bird
641 cherry-oat aphid, *Rhopalosiphum padi*, on resistant and non-resistant barley. *Entomologia*
642 *Experimentalis et Applicata*, **138**, 12-20.
643
644 Jyoti JL, Qureshi JA, Michaud JP *et al.* (2006) Virulence of two Russian wheat aphid to eight
645 wheat cultivars at two temperature. *Crop Science*, **46**, 774-780.
646
647 Kiriac I, Gruber F, Poprawski TJ *et al.* (1990) Occurrence of sexual morphs of Russian wheat
648 aphid, *Diuraphis noxia* (Homoptera: Aphididae), in several locations in the Soviet Union and
649 the Northwestern United States. *Proceedings of the Entomological Society of Washington*, **92**,
650 544-547.
651
652 Kovalev OV, Poprawski TJ, Stekolshchinov AV *et al.* (1991) *Diuraphis Aizenberg* (Hom.,
653 Aphididae): key to apterous viviparous females and a review of Russian language literature
654 on the natural history of *Diuraphis noxia* (Kudjumov, 1913). *Journal of Applied Entomology*,
655 **112**, 425-436.
656
657 Li XQ, Dodson J, Zhou XY *et al.* (2007) Early cultivated wheat and broadening of agriculture
658 in Neolithic China. *The Holocene*, **17**, 555-560.
659
660 Liang HB, Zhang RZ, Zhang GX (1999) Prediction of suitable areas for Russian wheat aphid
661 in China. *Acta Entomologica Sinica*, **42S**, 55-61.
662

663 Librado P, Rozas J (2009) DnaSP v. 5: a software for comprehensive analysis of DNA
664 polymorphism data. *Bioinformatics*, **25**, 1451-1452.
665
666 Liu X, Marshall JL, Stary P *et al.* (2010) Global Phylogenetics of *Diuraphis noxia* (Hemiptera:
667 Aphididae), an invasive aphid species: evidence for multiple invasions into north America. *Journal*
668 *of Economic Entomology,* **103**, 958-965.
669
670 Llewellyn KS, Loxdale HD, Harrington R *et al.* (2003) Migration and genetic structure of the
671 grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed
672 using microsatellites. *Molecular Ecology*, **12**, 21-34.
673
674 Loxdale HD, Hardie J, Halber S *et al.* (1993) The relative importance of short- and long-range
675 movement of flying aphids. *Biological Review*, **68**, 291-311.
676
677 Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for
678 likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655.
679
680 Martinez-Torres D, Moya A, Hebert PDN, Simon JC (1997) Geographic distribution and
681 seasonal variation of mitochondrial DNA haplotypes in the aphid *Rhopalosiphum padi*
682 (Hemiptera: Aphididae). *Bulletin of Entomological Research*, **87**, 161-167.
683
684 Michel AP, Zhang W, Jung KJ *et al.* (2009) Population genetic structure of *Aphis glycines.*
685 *Environmental Entomology*, **38**, 1301-1311.
686
687 Motro U, Thomson G (1982) On heterozygosity and the effective size of populations subject to
688 size changes. *Evolution*, **36**, 1059-1066.
689

690 Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial
691 molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**,
692 1659-1672.
693
694 Peccoud J, Ollivier A, Plantegenest M, Simon JC (2009) A continuum of genetic divergence
695 from sympatric host races to species in the pea aphid complex. *Proceedings of National*
696 *Academy of Science USA*, **16**, 7495-7500.
697
698 Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent
699 reduction in the effective population size using allele frequency data. *Journal of Heredity*, **90**,
700 502-503.
701
702 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
703 multilocus genotype data. *Genetics*, **1555**, 945-959.
704
705 Puterka GJ, Burd JD, Burton RL (1992) Biotypic variation in a worldwide collection of
706 Russisan wheat aphid (Hemiptera: Aphididae). *Journal of Economic Entomology*, **85**,
707 1497-1506.
708
709 Pysek P, Richardson DM, Pergl J *et al.* (2008) Geographical and taxonomic biases in invasion
710 ecology. *Trends in Ecology and Evolution*, **23**, 237-244.
711
712 Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against
713 population growth. *Molecular Biology and Evolution*, **19**, 2092-2100.
714
715 Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for
716 exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Reich DE*,* Goldstein DB *(*1998*)* Genetic evidence for a Paleolithic human population expansion in Africa*. Proceedings of the National Academy of Sciences USA*, **95**, 8119-8123.

Reich DE, Feldman MW, Goldstein DB (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions. *Molecular Biology and Evolution*, **16**, 453-466.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Richerson PJ, Boyd R, Bettinger RL (2001) Was agriculture impossible during the Pleistocene but mandatory during the Holocene? a climate change hypothesis. *American Antiquity*, **66**, 287-411.

Roderick GK (1996) Geographic structure of insect populations: Gene flow, phylogeography, and their uses. *Annual Review of Entomology*, **41**, 325-352.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Resources*, **4**, 137-138.

Ross KG, Krieger MJB, Shoemaker DD (2003) Alternative genetic foundations for a key social polymorphism in fire ants. *Genetics*, **165**, 1853-1867.

Ross KG, Krieger MJB, Keller L *et al.* (2007) Genetic variation and structure in native populations of the fire ant *Solenopsis invicta*: evolutionary and demographic implications. *Biological Journal of the Linnean Society,* **92**, 541-560.

744 Ross KG, Shoemaker DD (2008) Estimation of the number of founders of an invasive pest
745 insect population: the fire ant *Solenopsis invicta* in the USA. *Proceedings of Royal Society B.*
746 **275**, 2231-2240.
747
748 Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for
749 Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.
750
751 Scott JM (2007) Native and exotic plant species exhibit similar population dynamics during
752 Succession. *Ecology*, **88**, 1098-1104.
753
754 Shufran KA, Burd JD, Anstead JA, Lushai G (2000) Mitochondrial DNA sequence divergence
755 among greenbug (Homoptera: Aphididae) biotypes: evidence for host-adapted races. *Insect*
756 *Molecular Biology*, **9**, 179-184.
757
758 Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence
759 variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single
760 introduction into the United States. *Journal of the Kansas Entomological Society,* **80**, 319-326.
761
762 Shufran KA, Payton TL (2009) Limited genetic variation within and between Russian wheat
763 aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology,*
764 **102**, 440-445.
765
766 Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of
767 mitochondrial gene sequences and a compilation of conserved polymerase chain reaction
768 primers. *Annals of the Entomological Society of America*, **87**, 651-701.
769
770 Simon C, Buckley TR, Frati F *et al.* (2006) Incorporating molecular evolution into phylogenetic

771 analysis, and a new compilation of conserved Polymerase Chain Reaction primers for animal
772 mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 545-579.
773
774 Simon JC, Baumann S, Sunnucks P *et al.* (1999) Reproductive mode and population genetic
775 structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite
776 markers. *Molecular Ecology*, **8**, 531-545.
777
778 Smith CM, Belay T, Stuffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera:
779 Aphididae) populations virulent to the *Dn4* resistance gene. *Journal of Economic Entomology*,
780 **97**, 1112-1117.
781
782 Stary P (1999) Distribution and ecology of the Russian wheat aphid, *Diuraphis noxia* (Kurdj.),
783 expanded to Central Europe (Hom.: Aphididae). *Journal of Pest Science*, **72**, 25-30.
784
785 Sunnucks P, Hales DF (1996) Numerous Transposed Sequences of Mitochondrial
786 Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae). *Molecular*
787 *Biology and Evolution*, **13**, 510-524.
788
789 Sunnucks P, De Barro PJ, Lushai G *et al.* (1997) Genetic structure of an aphid studied using
790 microsatellites: cyclic parthenogenesis, differentiated lineages, and host specialization.
791 *Molecular Ecology*, **6**, 1059-1073.
792
793 Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA
794 polymorphism. *Genetics*, **123**, 585-595.
795
796 Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics
797 analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Thornton CP, Schurr TG (2004) Genes, language, and culture: an example from the Tarim Basin. *Oxford Journal of Archaeology*, **23**, 83-106.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Venette RC, Ragsdale DW (2004) Assessing the invasion by soybean aphid (Homoptera: Aphididae): Where will it end? *Annals of the Entomological Society of America,* **97**, 219-226.

Vorburger C (2006) Temporal dynamics of genotypic diversity reveal strong clonal selection in the aphid *Myzus persicae*. *Journal of Evolutionary Biology*, **19**, 97-107.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Wilson ACC, Sunnucks P, Hales DF (1997) A microsatellite linked to the X-chromosome of *Sitobion* aphids reveals random inheritence of the X-chromosome in males. *Genetical Research, Cambridge*, **69**, 233-236.

Wilson ACC, Massonnet B, Simon JC *et al.* (2004) Cross-species amplification of microsatellite loci in aphids: assessment and application. *Molecular Ecology Notes*, **4**, 104-109.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177-1191.

825

826 Zhang RZ, Liang HB, Zhang GX (1999a) Research on Russian wheat aphid, *Diuraphis noxia*
827 Kurdjumov, in China. *Acta Entomologica Sinica,* **42S**, 11-17.

828

829 Zhang RZ, Liang HB, Zhang GX (1999b) Russian wheat aphid, *Diuraphis noxia*, and its
830 research progress worldwide. *Acta Entomologica Sinica,* **42S**, 130-140.

831

832 Zhang Y, Wang L, Wu K *et al.* (2008) Flight performance of the soybean aphid, *Aphis glycines*
833 (Hemiptera: Aphididae) under different temperature and humidity regimens. *Environmental*
834 *Entomology*, **37**, 301-306.

Data Accessibility

Mitochondrial sequences: Genbank accessions JN204386 - JN204421

Microsatellite sequences: Genbank accessions JN204377 -JN204385

Sample locations: uploaded as online supporting material.

Microsatellite data: DRYAD doi:10.5061/dryad.42sh717m

Figure Legends

Fig.1 Topographical map of northwestern China, Xinjiang, with the sample localities represented by black dots.

Fig. 2 Estimated mitochondrial DNA network with 95% plausible set of haplotype connections. Each haplotype (1-18) is shown as a circle or square. The size of the circle or square relates to the number of individuals sampled (scale shown at base of figure). Small black circles represent putative haplotypes that were not sampled (not labeled). Lines between circles represent a single base pair mutation.

Fig. 3 Structure bar plot of Chinese RWAs sampled in 2009 (A, *k*=3) and 2010 (B, *k*=4). The 2010 data are also presented following removal of the three southern populations and reanalysis (*k*=2). Each individual is shown as a vertical bar representing ancestry.

Fig.4 Three-dimensional factorial correspondence analysis of Chinese RWAs sampled in 2010. The circles indicate populations that cluster according to geography.

Table 1. Indices of genetic diversity for the13 sites sampled in 2009. Ho, observed heterozygosity; He, expected heterozygosity; Hs, gene diversity; Na, numbers of alleles; Ar, allelic richness based on 9 samples per population; MLGs, number of multilocus genotypes; #within, number of MLGs shared within a population; # among, number of MLGs shared among populations; GGD, index of global genotypic diversity (MLGs/N); Fis, the inbreeding index, the asterisks indicate significance after Bonferroni correction at 0.05 level. TCA, TachengA; TCB, TachengB; TL, Toli; EM, Emin; YM, Yumin; HF, Hobuksa; AL, Altay; FH, Fuhai; HB, Haba; BR, Berqin; UR, Urumqi; QT, Qitai; ML, Mori.

2009	**TCA**	**TCB**	**TL**	**EM**	**YM**	**BR**	**UR**	**QT**	**ML**	**HF**	**AL**	**FH**	**HB**
N	49	31	50	50	31	44	16	50	42	40	10	6	50
Ho	0.62	0.61	0.62	0.62	0.64	0.67	0.65	0.60	0.62	0.59	0.81	0.78	0.61
He	0.65	0.65	0.67	0.66	0.68	0.67	0.64	0.61	0.64	0.64	0.76	0.70	0.68
Hs	0.65	0.65	0.67	0.66	0.68	0.67	0.58	0.61	0.64	0.64	0.68	0.62	0.68
Na	9.6	8.1	9.7	9.7	8.2	8.8	4.9	8	9.4	6.8	5.5	4.5	11.1
Ar	5.42	5.47	5.7	5.68	5.63	5.56	4.23	4.75	5.49	4.73	5.30	-	6.08
MLGs	48	31	49	36	30	36	15	44	42	25	9	5	48
#within	2	0	1	6	1	6	1	5	0	6	1	1	2
#among	0	0	0	0	0	0	0	0	0	0	0	0	0
GGD	0.96	1	0.98	0.82	0.97	0.82	0.94	0.88	1	0.63	0.9	0.83	0.96
***F*is**	0.036	0.065	0.081	0.06	0.059	0.007	-0.002	0.008	0.037	0.072	-0.077	-0.12	0.097*

Table 2. Indices of genetic diversity for the13 sites sampled in 2010. The abbreviations are the same as indicated in Table 1. QP, Qapqia; MS, Manas; WQ, Wuqia; CL, Cele; PS, Pishan.

2010	TCA	TCB	TL	EM	YM	BR	UR	QT	ML	QP	MS	WQ	CL	PS
N	41	22	50	31	50	11	50	50	50	53	50	52	52	9
Ho	0.61	0.76	0.64	0.71	0.62	0.74	0.62	0.57	0. 60	0.65	0.62	0.59	0.60	0.59
He	0.63	0.73	0.65	0.71	0.67	0.60	0.60	0.64	0.59	0.65	0.65	0.71	0.42	0.52
Hs	0.63	0.65	0.65	0.64	0.67	0.47	0.60	0.64	0.59	0.65	0.59	0.71	0.33	0.47
Na	7.5	6.7	10	7.9	10.5	3.5	7.7	8.7	8.6	9.8	7.9	8.1	3.9	2.7
Ar	4.99	5.27	5.57	5.382	5.85	3.3	4.51	5.29	5.00	5.50	4.76	5.4	2.52	2.7
MLGs	30	30	49	30	50	5	34	48	46	51	43	52	41	7
#within	4	3	1	1	0	2	7	2	2	2	5	0	8	1
#among	0	0	0	0	0	0	0	0	0	0	0	0	1	1
GGD	0.73	0.77	0.98	0.97	1	0.45	0.68	0.96	0.92	0.96	0.86	1	0.79	0.78
***F*is**	0.03	-0.044	0.004	0.011	0.079*	-0.26	-0.032	0.12*	-0.004	0.003	0.046	0.17*	-0.43*	-0.15

Table 3. *F*st values and significance of pairwise comparisons among (A) 2009 and (B) 2010 populations. The abbreviated names were the same as the localities in table 1. Bold values indicate significance after Bonferroni correction at 0.05 level. The grey cells highlight the *F*st Between southern and northern populations.

(A) 2009	EM	TCA	TCB	UR	HB	TL	QT	BR	HF	YM	ML	AL
Emin	-											
TachengA	0.0061	-										
TachengB	**0.0251**	**0.0318**	-									
Urumqi	**0.0917**	**0.1045**	**0.0770**	-								
Haba	0.0089	0.0086	**0.0336**	**0.0778**	-							
Toli	**0.0186**	**0.0200**	**0.0206**	**0.0710**	**0.0143**	-						
Qitai	**0.0444**	**0.0401**	**0.0466**	**0.1071**	**0.0366**	**0.0141**	-					
Berqin	**0.0338**	**0.0354**	**0.0264**	**0.0774**	**0.0275**	0.0134	**0.0223**	-				
Hobuksa	**0.0552**	**0.0639**	**0.0626**	**0.0868**	**0.0490**	**0.0575**	**0.0808**	**0.0569**	-			
Yumin	0.0091	0.0058	**0.0265**	**0.0981**	0.0113	0.0055	**0.0270**	**0.0237**	**0.0553**	-		
Mori	**0.0210**	**0.0234**	**0.0299**	**0.0943**	**0.0196**	0.0096	0.0057	**0.0167**	**0.0533**	0.0088	-	
Altay	0.0248	0.0204	**0.0525**	0.1129	0.0214	0.0271	**0.0665**	0.0422	**0.0768**	0.0195	**0.049**	-

(B) 2010	WQ	CL	PS	QP	MS	UR	QT	ML	YM	TL	TCA	TCB	EM	BR
Wuqia	-													
Cele	**0.2689**	-												
Pishan	**0.1304**	**0.1203**	-											
Qapqia	**0.0989**	**0.2214**	**0.1394**	-										
Manas	**0.1192**	**0.2525**	**0.1539**	**0.0682**	-									
Urumqi	**0.1085**	**0.2772**	**0.1648**	**0.0671**	**0.0653**	-								
Qitai	**0.1141**	**0.2673**	**0.1829**	**0.0294**	**0.0654**	**0.0527**	-							
Mori	**0.1346**	**0.2363**	**0.1812**	**0.0403**	**0.0735**	**0.0664**	**0.0167**	-						
Yumin	**0.0843**	**0.2097**	**0.1211**	**0.0212**	**0.0452**	**0.0440**	**0.0163**	**0.0262**	-					
Toli	**0.1116**	**0.2637**	**0.1715**	**0.0363**	**0.0479**	**0.0500**	**0.0125**	**0.0344**	0.0081	-				
TachengA	**0.0957**	**0.2375**	**0.1157**	**0.0440**	**0.0640**	**0.0374**	**0.0422**	**0.0676**	**0.0210**	**0.0393**	-			
TachengB	**0.1117**	**0.3004**	**0.1541**	**0.0743**	**0.0809**	**0.0666**	**0.0647**	**0.0906**	**0.0520**	**0.0615**	**0.0465**	-		
Emin	**0.0793**	**0.2794**	**0.1422**	**0.0307**	**0.0427**	0.0220	**0.0301**	**0.0517**	**0.0177**	**0.0214**	0.0219	**0.0407**	-	
Berqin	**0.1761**	**0.3768**	**0.2402**	**0.1210**	**0.1204**	**0.1233**	**0.1362**	**0.1620**	**0.1068**	**0.1156**	**0.1187**	**0.1521**	**0.1003**	-

Table 4. Tests for Chinese RWA demographic fluctuation under bottleneck or expansion calculated using BOTTLENECK and KGTEST. *:p<0.05; **p<0.01. Dash (-) indicates that the test was not performed because the sample size was too low.

	2009	Ermin	TachengA	TachengB	Urumqi	Haba	Toli	Qitai	Berqin	Hobaksa	Yumin	Mori	Altay	Fuhai
Bottleneck	TPM	0.4316	0.2324	0.1934	0.8203	0.1602	0.5566	0.1055	0.4316	1.0000	0.4316	0.01855*	0.7344	-
	Mode	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	-
kgtest	k	5	8*	6	4	6	6	6	5	5	6	7	5	7
	g	0.9348	0.8031	0.7707	2.5153	0.8610	0.8151	1.1397	0.8839	1.4450	0.9100	1.1238	0.7539	0.8094

	2010	Wuqia	Cele	Pishan	Qapqia	Manas	Urumqi	Qitai	Mori	Yumin	Toli	TachengA	TachengB	Ermin	Berqin
Bottleneck	TPM	0.9219	1.0000	-	0.01855*	0.1289	0.2754	0.3223	0.1309	0.0840	0.00488**	0.2324	0.5703	0.4258	0.8438
	Mode	L-shaped	L-shaped	-	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped	L-shaped
kgtest	k	6	6	5	8*	9**	6	6	6	8*	8*	6	3	6	7
	g	0.5462	3.3974	1.8589	0.8515	1.1492	1.0882	1.3770	1.3197	0.7431	0.7840	1.0950	1.1598	0.7812	2.8129

Author Information Box

This work forms part of B.Z.'s PhD thesis on *D. noxia* population genetics. B.Z. conducted the molecular laboratory research under the guidance of L.K. in CAS, and genetic analyses with the help of S.J.F. in QUT. O.R.E. has research interests in invertebrate genomics and in particular, aphid genetics and aphid-plant interactions. S.J.F. is a population geneticist whose research integrates field-based population studies with molecular techniques to deliver ecological management outcomes. L.K. is an entomologist studying the ecogenomics of the migratory locust and the plasticity of phase transition and gene expression modulation.

Supporting information

Appendix S1 Sampling information for RWAs in Xinjiang.

Appendix S2 Primer details and indices of genetic variation for the ten microsatellite loci used in this study.

Appendix S3 Mitochondrial genetic diversity in each population.

Appendix S4 AMOVA for RWAs sampled in 2010 and analyzed in three groups: Wuqia, Cele and Pishan, and northern populations.

Appendix S5 Pairwise *F*st and Exact G test for each site sampled in 2009 and 2010. AMOVA analysis of 9 groups for temporal comparison.

Appendix S6 FCA of Chinese RWAs from northern populations sampled in 2009 and 2010.

Appendix S7 Gene flow patterns of RWAs in far eastern ranges based on long-term estimates of gene flow.