# Evolutionary genetics and human assisted movement of a globally invasive pest (Russian wheat aphid: *Diuraphis noxia*).

**Bo Zhang**

M. Sc.

This dissertation is submitted as a requirement for the degree of

**Doctor of Philosophy**

Earth, Environment and Biological Sciences
Science and Engineering Faculty
Queensland University of Technology
Brisbane, Australia

July 2012

## Abstract

This PhD study has examined the population genetics of the Russian wheat aphid (RWA, *Diuraphis noxia*), one of the world's most invasive agricultural pests, throughout its native and introduced global range.

Firstly, this study investigated the geographic distribution of genetic diversity within and among RWA populations in western China. Analysis of mitochondrial data from 18 sites provided evidence for the long-term existence and expansion of RWAs in western China. The results refute the hypothesis that RWA is an exotic species only present in China since 1975. The estimated date of RWA expansion throughout western China coincides with the debut of wheat domestication and cultivation practices in western Asia in the Holocene. It is concluded that western China represents the limit of the far eastern native range of this species. Analysis of microsatellite data indicated high contemporary gene flow among northern populations in western China, while clear geographic isolation between northern and southern populations was identified across the Tianshan mountain range and extensive desert regions.

Secondly, this study analyzed the worldwide pathway of invasion using both microsatellite and endosymbiont genetic data. Individual RWAs were obtained from native populations in Central Asia and the Middle East and invasive populations in Africa and the Americas. Results indicated two pathways of RWA invasion from 1) Syria in the Middle East to North Africa and 2) Turkey to South Africa, Mexico and then North and South America. Very little clone diversity was identified among invasive populations suggesting that a limited founder event occurred together with predominantly asexual reproduction and rapid population expansion. The most likely explanation for the rapid spread (within two years) from South Africa to the New World is by human movement, probably as a result of the transfer of wheat breeding material. Furthermore, the mitochondrial data revealed the presence of a universal haplotype and it is proposed that this haplotype is

representative of a wheat associated super-clone that has gained dominance worldwide as a result of the widespread planting of domesticated wheat.

Finally, this study examined salivary gland gene diversity to determine whether a functional basis for RWA invasiveness could be identified. Peroxidase DNA sequence data were obtained for a selection of worldwide RWA samples. Results demonstrated that most native populations were polymorphic while invasive populations were monomorphic, supporting previous conclusions relating to demographic founder effects in invasive populations. Purifying selection most likely explains the existence of a universal allele present in Middle Eastern populations, while balancing selection was evident in East Asian populations. Selection acting on the peroxidase gene may provide an allele-dependent advantage linked to the successful establishment of RWAs on wheat, and ultimately their invasion potential.

In conclusion, this study is the most comprehensive molecular genetic investigation of RWA population genetics undertaken to date and provides significant insights into the source and pathway of global invasion and the potential existence of a wheat-adapted genotype that has colonised major wheat growing countries worldwide except for Australia. This research has major biosecurity implications for Australia's grain industry.

**Key words**: Russian wheat aphid, population genetics, native range, invasive pathways, genetic isolation, demography, salivary gland genes, selection

# Table of Contents

# List of Tables

# List of Figures

# Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

QUT Verified Signature

Bo Zhang

October 2012

## Acknowledgements

This thesis is dedicated to my father, Wensheng Zhang, who helped me to sample in the remote mountains and desert gobi region of Xinjiang.

I thank Dr. Susan Fuller, Dr. Owain Edwards and Dr. Le Kang for giving me this wonderful opportunity to pursue my doctoral studies. What I have learnt has transformed me and is all because of them. Without their help and support over the past three years, I would not have become a scholar with professional understanding and awareness in scientific research. Thanks for their excellent mentoring and support. They have not only taught me important scientific skills and professionalism, but they have also helped me to shape my career for the future. It has been a great pleasure to work in close association with them. I would also like to thank my parents for their spirit and practical support. My mum prepared all my sampling equipment and food before going to the field, while my father assisted me with field work. His commitment was such that he even traveled to some sites himself to look for Russian wheat aphids after I had returned to the laboratory. He sent me samples afterwards with detailed information of sampling date, host, altitude, coordinates and injury rate. How professional you are, my dear father!

I still remember the first time arriving in Brisbane, where the inspiring garden city shocked my heart. Not only is Brisbane beautiful, but QUT is such a friendly work place. I would like to thank Dr. Peter Mather, Dr. David Hurwood, Dr. Peter Prentis, Dr. Stephen Cameron and Dr. Tony Clarke who guided me in different aspects of my research. In particular, Dave taught me to be a critical thinker, and I have learnt that this is a true mark of a real scientist. I would also like to thank Dr. Eleanor Adamson, Dr. Matt Krosch, Litticia Bryant, Hyungtaek Jung, Vincent Chand and all the other HDR students studying in the QUT Molecular Genetics Research Facility. Your smiles every day helped me enormously.

I worked for two years in Dr. Le Kang’s laboratory, the Eco-genomics and Adaptation Group in the Institute of Zoology, Chinese Academy of Sciences. The faculty staff and students provided helpful technical assistance and academic advice. Thanks for their supportive hands.

I would also like to thank all the Russian wheat aphid scientists from around the world who sent samples to China to be used in this study. Finally, I am most grateful to the CRC Plant Biosecuity, QUT and CAS for financially supporting my project.

# List of Publications

B. ZHANG, O. R. EDWARDS, L. KANG, S. J. FULLER. (2012) Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction? *Molecular Ecology*, **21**: 2130–2144.

B. ZHANG, O. R. EDWARDS, L. KANG, S. J. FULLER. Worldwide invasion genetics of the Russian wheat aphid (*Diuraphis noxia*). In preparation for submission to Molecular Ecology.

B. ZHANG, O. R. EDWARDS, L. KANG, S. J. FULLER. Evidence for selection acting on salivary gland peroxidase in Russian wheat aphids (*Diuraphis noxia*). In preparation for submission to Bulletin of Entomological Research.

B. ZHANG, C. MA, O. R. EDWARDS, L. KANG, S. J. FULLER. Large repetitive sequences outside the control region in aphid mitochondrial genomes. In preparation for submission to PLoS ONE.

# Chapter 1

# General introduction and literature review

## 1. 1 Biological Invasion

Biological invasion is regarded as a worldwide environmental problem (Everett 2000; Mooney & Cleland 2001; Levine 2008; Liebhold & Tobin 2008). With the increase in human population numbers over the decades, an increase in human movement has been accompanied by the movement of a variety of plants and animals entering into new ecosystems beyond their own native ranges (Kolar & Lodge 2001). Irrespective of whether the introduction is intentional or accidental, the introduction of an alien species influences and accelerates the ecological change of new habitat (Mooney & Cleland 2001). Biological invasions have been recognized as an important factor influencing global change (Bright 1999), particularly biodiversity loss, as the global environment becomes increasingly compromised as a result of environmental degradation. As invasions can primarily change ecosystem processes, related fields of experts have been called on to develop a global strategy for dealing with detrimental non-indigenous species (Everett 2000).

The process of invasion of non-indigenous species can be separated into three stages: arrival, establishment, and spread (Liebhold & Tobin 2006). A newly introduced exotic species may initially distribute relatively small numbers of propagules to remote locations (Doren *et al.* 2009). To successfully establish, the primary step is for the exotic species to withstand the conditions (humidity, temperature, predators, etc.) in each location. Once established, as the species reproduces over time, additional propagules are released and new sites colonized. As new propagules are produced and distributed, more propagules can be released over an ever-increasing region, providing a greater opportunity for those propagules to encounter favorable conditions and to create greater spatial connectivity between populations (Mooney & Cleland 2001; Lockwood *et al.* 2005).

The spread of an invasion may not always be continuous, but rather it may be stratified. In other words, part of the population may jump in front, and this is often a relatively isolated individual, which individually, or through its descendents ultimately coalesces with the rest of the population (Liebhold &

Tobin 2006). Continuous spread results from dispersal following population growth, however, many observed cases of spread fail to follow the theoretical model of diffusion. Such a model is too simplistic for most invasive species, particularly because it does not take into account long-distance dispersal. Stratified dispersal is defined as the combination of short-range and long-range diffusion, and it is the fundamental process in exotic species dispersal. Invasive species that have exhibited this form of dispersal include the gypsy moth (Whitmire & Tobin 2006), soybean aphid (Venette & Ragsdale 2004), argentine ant (Suarez *et al.* 2001), and emerald ash borer (Muirhead *et al.* 2006).

Most pathways of introduction have resulted from increased worldwide trade. With increased trade, there has been an increase in the volume and speed of transportation by mobile vectors that move exotic species into new ecosystems (Everett 2000; Karatayev *et al.* 2007; Ascunce *et al.* 2011). It is critical to control new invaders in the arrival stage; however, this stage is often ignored (Puth & Post 2005). Invasion is generally recognized as being complete once the pest population has grown to a considerable level and is unlikely to go extinct in the environment.

### *1.1.1 Genetic basis of biological invasion*

Most invasive species have the ability to respond quickly and to adapt to new and/or challenging environmental conditions (Mooney & Cleland 2001). Changing selective regimes may affect an invasive species through increased selection for adapted genotypes, or through relaxed selection for defense because of the absence of co-evolved natural enemies. Thus, rapid phenotypic response often occurs and has been reported in a large number of studies (Grosholz 2002; Hanfling & Kollmann 2002; Richards *et al.* 2012). For example, among introduced populations of the lizard, *Anolis sagrei*, phenotypic variation (body size, toe pad-lamella number, and body shape) was attributed to differential admixture of various source populations and adaptation to the new environment (Kolbe *et al.* 2007). Changes can also result in the evolution of highly competitive, but less well defended genotypes. *Lythrum salicaria*, a perennial plant showed greater biomass in introduced

individuals than their ancestral genotypes and were less resistant to, or tolerant of herbivores (Hanfling & Kollmann 2002). This case highlights that increased competitive capability may not only result from the growth or defense tradeoff, but also from weak co-adaptation between indigenous and exotic species.

Populations of invasive species often exhibit low genetic diversity, and therefore a high degree of phenotypic plasticity might be an advantage when dealing with new environments and is thought to increase invasion success (Frankham 2005; Kolbe *et al.* 2004; Lindholm *et al.* 2005). A number of studies have shown that a species with low genetic diversity in its introduced range adapts better to new environments than the same species in its native range (Dlugosch & Parker 2008; Mergeay *et al.* 2006; Rollins *et al.* 2009; Shoemaker *et al.* 2006). However, the opposite has also been found where exotic species with low genetic variability are limited in their invasion success because they are constrained in their ability to adapt to the new environment (Roman & Darling 2007).

If the invasive population is founded by a limited number of individuals and there is a low rate of reproduction resulting in a bottleneck (Nei *et al.* 1975), the genetic background will be further constrained as additive genetic variance is eroded (Lee 2002; Sakai *et al.* 2001). Such a loss of additive genetic variance will slow evolutionary responses to selection and limit the adaptive evolution of fitness-related traits. In Australia, invasive guppies, *Poecilia reticulata*, were found to have low genetic variability (limited mtDNA diversity and low microsatellite allelic diversity) as a result of having gone through a bottleneck (Lindholm *et al.* 2005). They found that the low genetic diversity was consistent with the release of additive genetic variation by dominance and epistasis following inbreeding, and with disruptive and negative frequency-dependent selection on fitness traits.

Genetic studies on invasive species can provide valuable information on how frequently introductions are accompanied by severe genetic bottlenecks, whether bottlenecks constrain adaptive evolution in invaders, and whether

contemporary gene flow among introductions represents the key to invasion success.

*1.1.2 Molecular markers used in invasion genetics research*

Invasion genetics research can provide critical information on the source and invasion pathway of an exotic species (Bonizzoni *et al.* 2004). Reliable records of the precise origins of introduced populations do not exist for many invasive species. In this case, it is useful to survey molecular variation among individuals from within the native and invasive range to identify the source of the introduction. The accuracy of this approach will be determined by the intensity of sampling in introduced and source areas, by the resolution of the molecular markers involved, and by the scale of differentiation across the potential source area. Modern genetic marker types include enzyme variants (allozymes), microsatellites (SSR), amplified fragment length polymorphisms (AFLP) and randomly amplified polymorphic DNA (RAPD). markers All of these have been well utilized in genetic studies of invasive species. For example, microsatellites have been investigated in invasive populations of the common starling in Australia to provide data on gene flow and source populations of infestation (Rollins *et al.* 2009). Mitochondrial DNA sequences of *Dendrobaena octaedra* indicated that non-native earthworms were introduced via multiple pathways (Cameron *et al.* 2008). Both mtDNA and microsatellite analyses have been used to investigate multiple invasions of the European green crab to North America and South Africa, followed by secondary invasions to the north-eastern Pacific, Tasmania, and Argentina (Darling *et al.* 2008). Studies of the introduced wasp, *Vespula germanica*, in Australia have revealed a clear picture of their invasion pathway (Goodisman *et al.* 2001). Research on the garden ant, *Lasius neglectus,* has suggested that introduced populations in Europe have most likely arisen from only a very few independent introductions from the native range, and new infestations have come from introductions from other invasive populations (Ugelvig *et al.* 2008).

Researchers are now armed with a range of innovative and reliable molecular tools that can be used to improve our understanding of the

population genetic structure of invasive species in their native and introduced ranges and importantly, these data can elucidate the occurrence of single vs. multiple introductions providing critical information on the mechanisms by which an invasive species may be introduced and spread. Invasion genetics data can also be used to predict the array of evolutionary responses and impacts that may result, as well as the future distribution of an invasive species.

## 1.2 Aphids: model organisms for invasion genetics research

Aphids are small insects that feed on plant phloem sap (Brisson & Stern 2006; international aphid genomics consortium 2010; Stern 2008). Many species have become important invasive economic pests because of their rapid population growth potential and resultant negative effect on plant growth and crop yield. Aphids exhibit a range of characteristics including high fecundity and phenotypic diversity that has likely aided their colonization of new environments, facilitating successful worldwide invasion, and making them an attractive study organism for invasion genetics research.

### *1.2.1 Phenotypic plasticity*

Aphids exhibit polyphenism, a special case of phenotypic plasticity in which one genotype can produce several phenotypes without intermediate forms (Nijhout 1999). A common form of aphid polyphenism is the switch between viviparous parthenogenesis and sexual reproduction with eggs, depending on environmental cues. The major advantage of sexual reproduction in aphids is to resist negative ecological conditions and for asexual reproduction, to increase rapidly population numbers (colonies) (Miura *et al.* 2003). Reproductive polyphenism allows aphids to combine the immediate advantage of clonal growth conferred by parthenogenesis with the long-term advantages of recombination conferred by sexual reproduction (Simon *et al.* 2002). Thus, aphids provide an opportunity for biologists to study the evolution of sex, adaptation, and coexistence of both modes of reproduction (Doncaster *et al.* 2000; Thomas *et al.* 2012).

Aphids also display a range of phenotypic morphs including winged or wingless morphs, nymphs or adults, asexual or sexual reproductive morphs, and even specialized soldier nymphs (Abbot *et al.* 2001; Brisson & Stern 2006; Stern & Foster 1996; Strassmann & Queller 2001). Seasonal forms may have different host preferences (Tosh *et al.* 2002), as well as different physiological and ecological properties. The winged morph is specialized for migration and plays a major role in host selection by aphids (Powell *et al.* 2006). Additionally, environmental cues (e.g., crowding, deteriorating plants, natural enemies, and abiotic factors such as temperature and photoperiod) may induce the production of winged aphids in response to habitat and resource deterioration.

### *1.2.2 Aphid-host plant interactions*

Aphids interact with their hosts in a number of ways, but the most pervasive interaction comes from the action of withdrawing and injecting saliva when penetrating plant leaves (Goggin 2007; Miles 1999). The proteinaceous and watery saliva that aphids secrete has been found to contain various enzymes that can undergo translocation in the host and have toxic effects on plants (Burd 2002; Madhusudhan & Miles 1998; Miles 1999; Tjallingii 2006). The calcium-binding proteins in aphid saliva can reverse phloem occlusion, which allows aphids to feed at one site for hours at a time. Aphid saliva prevents sieve tube plugging in the host plant (Will *et al.* 2007). Research on pea aphids has shown that the knockdown of transcript C002, which is abundant in the salivary gland, contributes to the reduction of contact time between aphids and host sap, and has even led to aphid lethality (Mutti *et al.* 2006 & 2008). Additionally, aphid-host interactions can have impacts on the physiology of both the hosts and the aphids. For example, aphids can modify host morphology resulting in the formation of foliar galls, and they can also affect the nutritional quality of the host (Botha & Matsiliza 2006; Girousse *et al.* 2005; Wool 2004).

### *1.2.3 Fitness and host-based differentiation*

It is widely accepted that aphid fitness is determined by the capacity of the aphid to obtain nutrients (Powell *et al.* 2006), which is influenced by plant

nutritional quality, aphid nutritional requirements, and defensive mechanisms (Awmack & Leather 2002). Therefore, while aphids may make contact with a range of plant hosts that contain nutritionally suitable substances, they may not settle, feed, or reproduce on these hosts (Caillaud & Via 2000). Host-based differentiation studies on aphids have found that plant preference and genetic divergence have resulted from long-term population development and mating on the same host (Hawthorne & Via 2001; Kirkpatrick & Ravigne 2002; Servedio 2001). It has been shown that aphid populations often consist of a number of genetically divergent host-associated races (Via 1999). Peccoud *et al.* (2008) found that pea aphids show striking associations with particular host plants rather than sampling locations, and phenotypic analyses confirmed their strong host specialization. Simon (2003) found distinct host races of pea aphid on pea, clover and alfalfa in France, and that there was a very strong association between host races of pea aphid and their symbiotic microbiota.

### *1.2.4 Aphid endosymbionts*

Over 150 million years ago, aphids evolved a mutualistic association with the bacteria *Buchnera aphidicola*, which live within specialized, polyploid aphid cells called bacteriocytes (Tagu *et al.* 2008). The genus *Buchnera* has one of the smallest known bacterial genomes (0.45 to 0.67 Mb), due to massive gene losses typical of endosymbiotic lineages (Moran and Wernegreen 2000, Tagu *et al.* 2008). Since the endosymbiont genome has lost many genes required for autonomous survival, it holds a strongly mutualistic relationship with aphids (Burke & Moran 2011). The genes that encode biosynthesis of some essential amino acids are still present suggesting that the symbiotic interaction is based mainly on nutritional exchange between *Buchnera* and the aphid (Baumann 2005). Aphids cannot survive without the endosymbiont, which is essential for nutrition and is transmitted from parent to offspring (Moran & Wernegreen 2000; Wernegreen 2002).

Endosymbiont genome evolution has been shown to reflect the aphid host genome, but at an accelerated rate due to the rapid generation time of the bacteriocyte relative to its host (Burke & Moran 2011; Clark *et al.* 1999;

Peccoud *et al.* 2009a). Therefore partial fragments of the endosymbiont genome can be examined to investigate the genetic relationships among aphid populations (Peccoud *et al.* 2008; Simon *et al.* 2003).

*1.2.5 Aphid population genetics*

Molecular markers have been used in a number of studies to trace the origin and historical pathway of spread of aphid pests. Peccoud *et al.* (2008) analyzed microsatellite and endosymbiont genetic diversity to conclude that the invasion of the pea aphid, *Acyrthosiphon pisum,* in Chile resulted from multiple introductions of highly host-specialized, asexual clones. Multiple routes of invasion of the tobacco aphid (*Myzus persicae nicotianae*) into North America were revealed using microsatellite markers (Zepeda-Paulo *et al.* 2010). Conversely, due to limited polymorphism found for microsatellite loci, mitochondrial genes, and RAPD markers, Shufran & Payton (2009) concluded that the Russian wheat aphid (*Diuraphis noxia*) in North America was the result of a single introduction. However, this conclusion has recently been challenged as AFLP markers have revealed that multiple introductions are likely to have occurred (Liu *et al.* 2010).

To understand and forecast population increases and movement to other crop fields, knowledge of the genetic structure of pest insect populations is required. In most aphid species, significant genetic differentiation and population structure occurs for two reasons. First, as a result of adaptation and specialization on different host plants, aphids may become isolated and divergent host associated races may develop (Charaabi *et al.* 2008; Ferrari *et al.* 2006; Peccoud *et al.* 2009a; Via 2001). Second, as a result of weak flying ability, migration and gene flow is limited and populations may become genetically differentiated (Loxdale *et al.* 1993; Thomas *et al.* 2012). However, large-scale dispersal, presumably wind-aided, can also occur leading to limited genetic differentiation among populations and continued expansion of aphids in both their native and invasive ranges (Dolatti *et al.* 2005; Michel *et al.* 2009; Shufran & Payton 2009). The rapid spread of the soybean aphid (*Aphis glycines*) across North America in a short period after its arrival (Venette & Ragsdale 2004) has been attributed to large-scale dispersal.

Michel *et al.* (2009) revealed low genetic diversity and genetic differentiation among invasive populations indicating that from a small colonizing soybean aphid population in North America, there was rapid population growth and widespread dispersal.

Population genetic studies have revealed that aphids generally have low mitochondrial DNA variation (Peccoud *et al.* 2009a; Shufran *et al.* 2007; Shufran & Payton 2009) and phylogenetic studies based on barcoding sequences have also shown limited genetic divergence among Aphidinae species (Lee *et al.* 2011; Wang *et al.* 2011). This may be the result of an interaction between inherited endosymbionts and mitochondria (Hurst & Jiggins 2005), however research is required to investigate this hypothesis in aphids. Interestingly, endosymbiont markers are rarely used in aphid population genetics research despite being successfully used to differentiate a continuum of pea aphid host races from divergent populations to almost complete speciation (Lozier *et al.* 2007; Peccoud *et al.* 2009a).

### *1.2.6 Research context of the current study*

The Russian wheat aphid (*Diuraphis noxia*) was selected as our research species because of its capacity for rapid population growth (Burd *et al.* 2006; Jyoti *et al.* 2006; Morrison & Peairs 1998; Smith *et al.* 2004) and its fast and successful invasion throughout all major grain growing regions of the world (except Australia) during last thirty years. The global invasion of Russian wheat aphids provides an ideal opportunity to examine some of the fundamental questions relating to the evolutionary biology and genetics of invasive species. In particular, this PhD study aims to improve our understanding of the evolutionary processes and rapid responses that characterise a recently introduced invasive species. By examining neutral and functional gene diversity within native and invasive populations, the effect that selective pressures (such as virulence interactions with host) have on invasiveness can be evaluated. Finally, this study will investigate the low variation that typifies aphid mitochondrial genomes and determine the utility of endosymbiont markers for invasion genetics research.

## 1.3 Russian wheat aphids

The Russian wheat aphid (RWA), *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae), is one of the world's most important pests of grain (barley, *Hordeum vulgare*, and wheat, *Triticum aestivum* L.). Since its introduction to the United States of America in 1986, management of this pest has focused on the development of resistant wheat cultivars, while chemical control has had less efficient and economic results. In the USA, estimated losses were in excess of $800 million in the period 1987 to 1993 (Morrison & Peairs 1998). The infested plants display white, yellow, or purple longitudinal streaks on leaves and stems, develop rolled leaves, and bent heads (Jyoti *et al.* 2006). Various management approaches have been developed to alleviate damage by RWAs, but chemical and biological controls have not worked well (Noma *et al.* 2005).

### *1.3.1 Worldwide distribution of Russian wheat aphids*

RWAs are native to central-western Asia and current evidence favors the center of origin to be in the Iranian-Turkestanian mountain range and extending to southern Russia, the Middle East, and central Asia (Kovalev *et al.* 1991). The earliest documented record of RWA crop damage comes from the Ukraine in the early 1900s, however RWAs gained recognition as an emerging global pest during the 1970's and 1980's as a result of their rapid spread through major grain producing areas in Europe, Africa and the Americas (Kovalev *et al.* 1991; Smith *et al.* 2004; Stary 1999). RWAs have been recognized as an invasive pest in China since they were first detected in 1976 (Li & He 1990). A map of the worldwide distribution of RWAs can be found in Figure 1.1, the historical detection dates for different countries in Table 1.1.

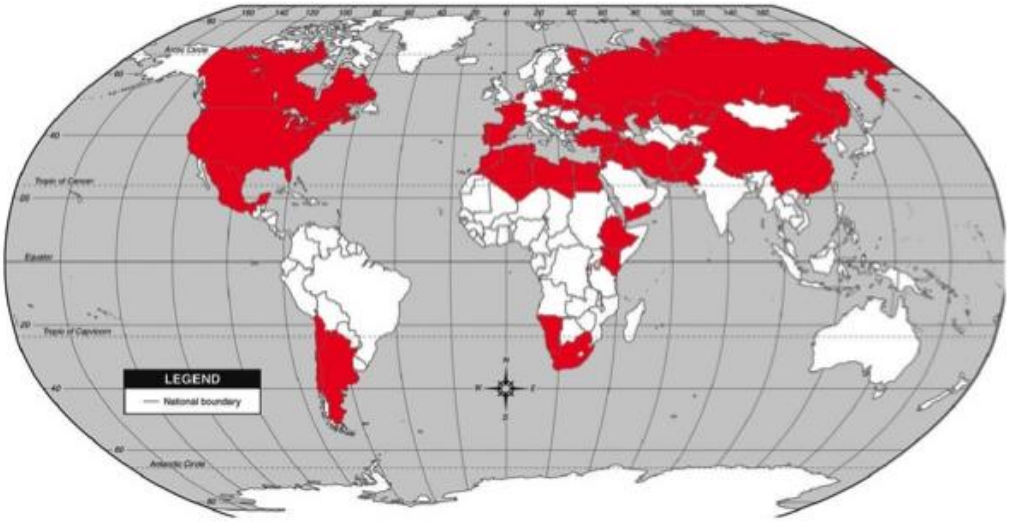



**Figure 1.1** Map of the worldwide distribution of Russian wheat aphids.

**Table 1.1** Year of first detection of Russian wheat aphid (from Zhang *et al.* 1999).

| **Continent** | **Country and record year** |
|---|---|
| Asia | Kazakhstan (1919), China (1975), Yemen (1981), Iraq (1983), Nepal (1986), Iran (1988), Pakistan (1989), Jordan (1989), Kirghizstan (1993), Syria (1994) |
| Europe | Ukraine (1900), Russia (1915), Georgia (1916), Spain (1947), Portugal (1947), Belgium (1963), Bulgaria (1981), Czech (1989), Yugoslavia (1989), Poland (1989), France (1989), Turkey (1989) |
| Africa | Algeria (1938), Morocco (1938), Egypt (1957), Libya (1962), South Africa (1978), Namibia (1978), Ethiopia (1984), Burundi (1989), Tunisia (1990) |
| North America | Mexico (1980), United States (1986), Canada (1988) |
| South America | Argentina (1989), Chile (1988) |

According to Zhang *et al.* (1999) the history of invasion of RWAs throughout the world can be divided into three stages.

- Emergence in the early 20th Century. RWA only occurred in Ukraine, Russia, Georgia, and Kazakhstan during this period.
- Spread up until the 1970s. This was a period of gradual colonisation, where RWAs slowly spread to many European countries. During this period, the RWA was recognized as an emerging pest.
- Jeopardization after the 1980s. During this period, RWAs spread rapidly throughout the main grain production areas of Europe, Asia, North America, and South America and caused severe crop damage.

*1.3.2 Russian wheat aphid biology*

RWAs exhibit two patterns of life cycle; cyclical parthenogenesis (holocycly) and obligate parthenogenesis (anholocycly). RWAs are reported to have an anholocyclic life cycle in southern Africa, North America, South America, France, and Turkey, while they are holocyclic in the USSR, Hungary, Spain and China (Zhang *et al.* 2001). The difference in life cycle pattern depends on the temperature of the habitat and does not appear to be due to whether it is an invasive or native population. For example, the RWA is native to Iran, but an obligate parthenogen in this country (Dolatti *et al.* 2005).

Typically, RWAs reproduce most of the year by viviparous parthenogenesis, a major factor in their destructive potential. In areas where RWAs are anholocyclic, obligate parthenogenesis occurs in asexual lines that have apomictic clonal generations throughout the whole year without sexual forms and mating. Wei *et al.* (1994) describes the holocyclic annual life cycle of RWAs as follows. Viviparous pregnant females are produced in late spring and early summer and each parthenogenetic female produces hundreds of genetically identical embryos. The progeny are either alate (winged) or apterous (wingless) virginopara. In early summer, RWA populations peak and large numbers of winged aphids migrate, often to an alternate host while the wheat crop is harvested. After oversummering on wheat volunteers or

native grasses, they move to winter wheat in early autumn. Induced by short photoperiod and low temperatures in autumn, sexual male and female aphids are produced. Sexual reproduction occurs and cold-resistant diapausing eggs are produced that survive the freezing conditions in winter.

*1.3.3 Russian wheat aphid virulence*

A range of RWA biotypes have been described and although they are morphologically identical, these biotypes display differential preference and fitness on specific host plants (Shufran *et al.* 2007). RWA biotypes are characterized according to an aphid's ability to damage wheat with resistance genes, and are determined using leaf chlorosis damage ratings and an index of leaf rolling (Burd *et al.* 2006; Haley *et al.* 2004; Puterka *et al.* 2007; Shufran *et al.* 2007). Five biotypes have been discovered in the USA and named as RWA1 (originally collected in Texas), RWA2 (damages crops with the *Dn4* gene), and RWA3, RWA4 and RWA5 (damages wheat crops with the *Dn1-Dn9* resistance genes) (Burd *et al.* 2006). RWA biotype RWA2 in North America is characterized as having a rapid growth potential and wide spread distribution (Puterka *et al.* 2007). In South Africa, at least three distinct biotypes have been reported including RWASA1, RWASA2, and unannotated Clone1 (Tolmay *et al.* 2007). As is the case in North America, RWAs from Hungary, Russia, Syria, Chile, Czech Republic, and Ethiopia have all been identified to resist the *Dn4* gene (Basky *et al.* 2002; Puterka *et al.* 1992; Smith *et al.* 2004).

The injection of RWA salivary enzymes into plant tissues is toxic to plants and has a negative impact on the membranes of chloroplasts (Miles 1999). Protein fractions extracted from RWAs and injected into wheat seedlings have induced plant stunting and injury symptoms (Lapitan *et al.* 2007). Damage symptoms were compared between two wheat cultivars with and without the *Dn7* resistance gene. The *Dn7*-containing cultivar was resistant to aphid injury and the extent of plant growth and higher levels of defense-related enzyme activity suggested that this cultivar was in better defensive condition. These results are consistent with previous work on wheat responses to RWA infestation, and suggest a possible interaction between

*Dn7* and a protein or proteins within the aphid. It may be a salivary elicitor, but no further research has been undertaken to purify or locate the protein/proteins responsible. Unfortunately, the degree of relatedness between the two wheat cultivars was not specified in this report; therefore, it was questionable whether other genes besides *Dn7* may be involved in the differential plant reactions to aphid extracts. Nevertheless, Lapitan *et al.*'s (2007) study of the protein offers the prospect of identifying specific virulence factors from highly destructive RWA biotypes. Although RWA1 was first introduced to the USA in 1986 (Shufran *et al.* 2007), a second biotype (RWA2) emerged in 2003 (Burd *et al.* 2006). It is assumed that the emergence and prevalence of RWA2 in the USA results from plant resistance, and it is possible that an aphid saliva-plant interaction may be responsible. Cui *et al.* (2012) found that transcripts of putative virulence genes in salivary glands exhibit high diversity in RWA1 and RWA2. However, it remains unknown whether these diverse transcripts could affect protein function, or whether it could be detected in plant sap and tissue, or whether such transcripts could influence plant immunity and metabolism.

### *1.3.4 Russian wheat aphid invasion genetics*

Knowledge of the genetic background of the founding native population of an invasive species will provide an improved ability to predict the array of evolutionary responses and impacts that may result, as well as the future distribution of the invasive species (Liebhold & Tobin 2008). However, only a limited number of population genetic studies on RWAs have been undertaken, and these have focused on introduced populations. Very little is known about this species in its native range (Shufran *et al.* 2007; Shufran & Payton 2009).

An important consideration is that RWAs in their winged form possess the ability to migrate or disperse. Logistically, it is difficult to directly monitor flight of RWAs in the field, and levels of immigration and emigration among populations cannot be determined using traditional ecological methods. If RWAs possess a high migration potential, have they naturally colonized surrounding countries (eg. in a step-wise fashion) or has their expansion been human mediated? Genetic methods, combined with innovative

statistical analysis techniques, can be used to answer such questions concerning the pathway of invasion and mechanisms driving the colonization process worldwide.

## 1.4 PhD research objectives

The Russian wheat aphid is one of the greatest biosecurity threats to the grains industry in Australia. The Australian grains industry is developing a RWA response plan based on pre-emptive breeding (Australian CRC Plant Biosecuirity 2008-2012), but RWAs are known to develop virulent biotypes that can overcome host plant resistance. It is likely that more genetically diverse populations of RWA exist in native areas and that these areas may harbor an abundant source of virulent clones. With increasing global trade, there is a corresponding increase in the likelihood of RWA incursions from areas where this species is native, potentially increasing the biosecurity risk of this species to the Australian grains industry. Despite this threat, there has been very little genetic research on native populations of RWA, and limited examination of the worldwide introduced populations. To address this knowledge gap, this PhD has analyzed the population genetics of native and introduced RWA populations to determine the invasion pathway of this species and to examine whether there is a functional basis for invasiveness.

To achieve this aim, it was first necessary to obtain a comprehensive selection of samples from within the native range of RWAs. Political barriers prevented sampling in countries such as Iran, Iraq, Afghanistan, Kazakhstan, and other parts of the former USSR. However, in collaboration with the Chinese Academy of Sciences in Beijing, it was possible to sample RWA populations in Western China. While some scientists suggest that RWAs are native to central Asia (Kovalev *et al.* 1991), others report that RWAs are an introduced pest only present in China since the 1970s (Zhang *et al.* 1999). It is important to consider that existing historical data provides an indication of the date of detection of RWAs in different countries, but this may not be the actual time that RWAs colonized each country. An examination of molecular genetic variation can solve the dispute as to whether Chinese RWAs are a

native or invasive pest (in native populations we would expect high genetic and clonal diversity). It is also difficult to trace the worldwide invasion pathway of RWAs, based on largely anecdotal historical data. Thus, this PhD study also examined the genealogical and evolutionary relationships among individuals and traced the pathway of colonization worldwide.

The reason for the successful invasion of RWAs worldwide is still an unanswered question. Essential factors influencing the successful introduction and establishment of RWAs may be related to its ability to adapt to new environments; however, the dispersal and stable spread of this species has been intrinsically linked with the presence of compatible host plants. The fundamental connection between an aphid and its host plant lies in the interaction between the injection of aphid saliva and the plant response. Based on the Jones–Dangl zigzag model (Jones and Dangl 2006), the salivary elicitor should be compatible with the plant immunity system so that the elicitor escapes from plant recognition and thus does not affect further injection. If variation exists in the salivary elicitors and is present in the founding population of a new invasion, aphids may have a greater chance of successfully colonizing new plants and rapidly increasing in number. Additionally, if variation in salivary gland genes can be aligned with the RWA invasion pathway, it may indicate that there is a functional genetic basis to the global invasiveness of RWAs.

Finally, this PhD study has compared the mitochondrial genomes of a range of aphid species, including the RWA, to characterize the particular genome features that are unique to aphids and may be responsible for low mitochondrial DNA variation.

### *1.4.1 Specific objectives of the study*

Studies on worldwide populations of RWA have not been undertaken, and therefore the invasion genetics and evolutionary biology of this species is not well understood.

The specific objectives of this PhD study are:

- To characterize the population genetics of RWA in its native range.
- To compare genetic diversity within and among native and introduced populations of RWA.
- To determine the pathway of invasion of RWAs throughout the world and the potential mechanisms enhancing global invasion.
- To examine whether natural selection has acted on salivary gland genes and has influenced the global invasion of RWA.
- To investigate the low mitogenome variation of RWA.

## 1.5 References

Abbot P, Withgott JH, Moran NA (2001) Genetic conflict and conditional altruism in social aphid colonies. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12068-12071.

Ascunce MS, Yang CC, Oakey J et al. (2011) Global invasion history of the fire ant Solenopsis invicta. Science, 331, 1066-1068.

Awmack CS, Leather SR (2002) Host plant quality and fecundity in herbivorous insects. *Annual Review of Entomology*, **47**, 817-844.

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.

Baumann P (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.

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

Botha CEJ, Matsiliza B (2006) Reduction in transport in wheat (*Triticum aestivum*) is caused by sustained phloem feeding by the Russian wheat aphid (*Diuraphis noxia*). *South African Journal of Botany*, **70**, 249-254.

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

Brisson JA, Stern DL (2006) The pea aphid, *Acyrthosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bioessays*, **28**, 747-755.

Burd JD (2002) Physiological modification of the host feeding site by cereal aphids (Homoptera: Aphididae). *Journal of Economic Entomology*, **95**, 463-468.

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.

Burke GR, Moran NA (2011) Massive genomic decay in *serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biology and Evolution*, **3**, 195-208.

Caillaud MC, Via S (2000) Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *American Naturalist*, **156**, 606-621.

Cameron EK, Bayne EM, Coltman DW (2008) Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. *Molecular Ecology*, **17**, 1189-1197.

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.

Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Molecular Biology and Evolution*, **16**, 1586-1598.

Cui F, Smith CM, Reese J *et al.* (2012) Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. Insect Science, DOI: 10.1111/j.1744-7917.2011.01487.x

Darling JA, Bagley MJ, Roman J *et al.* (2008) Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, **17**, 4992-5007.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431-449.

Dolatti L, Ghareyazie B, Moharramipour S *et al.* (2005) Evidence for regional diversity and host adaptation in Iranian population of the Russian wheat aphid. *Entomologia Experimentalis et Applicata*, **114**, 171-180.

Doncaster CP, Pound GE, Cox SJ (2000) The ecological cost of sex. *Nature*, **404**, 281-285.

Doren RF, Richards JH, Volin JC (2009) A conceptual ecological model to facilitate understanding the role of invasive species in large-scale ecosystem restoration. *Ecological Indicators*, **9**, S150-160.

Everett RA (2000) Patterns and pathways of biological invasions. *Trends in Ecology and Evolution*, **15**, 177-178.

Ferrari J, Godfray HCJ, Faulconbridge AS *et al.* (2006) Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution*, **60**, 1574-1584.

Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131-140.

Girousse C, Moulia B, Silk W *et al.* (2005) Aphid infestation causes different changes in carbon and nitrogen allocation in alfalfa stems as well as different inhibitions of longitudinal and radial expansion. *Plant Physiology*, **137**, 1474-1484.

Goggin FL (2007) Plant-aphid interactions: molecular and ecological perspectives. *Current Opinion in Plant Biology*, **10**, 399-408.

Goodisman MAD, Matthews RW, Crozier RH (2001) Hierarchical genetic structure of the introduced wasp *Vespula germanica* in Australia. *Molecular Ecology*, **10**, 1423-1432.

Grosholz E (2002) Ecological and evolutionary consequences of coastal invasions. *Trends in Ecology and Evolution*, **17**, 22-27.

Haley SD, Peairs FB, Walker CB *et al.* (2004) Occurrence of a new Russian wheat aphid biotype in Colorado. *Crop Science*, **44**, 1589-1592.

Hanfling B, Kollmann J (2002) An evolutionary perspective of biological invasions. *Trends in Ecology and Evolution*, **17**, 545-546.

Hawthorne DJ, Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, **412**, 904-907.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 1525-1534.

International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PloS Biology*, **8**, e1000313.

Jyoti JL, Qureshi JA, Michaud JP, Martin TJ (2006) Virulence of two Russian wheat aphid biotypes to eight wheat cultivars at two temperatures. *Crop Science*, **46**, 774-780.

Karatayev AY, Padilla DK, Minchin D *et al.* (2007) Changes in global economies and trade: the potential spread of exotic freshwater bivalves. *Biological Invasions*, **9**, 161-180.

Kirkpatrick M, Ravigne V (2002) Speciation by natural and sexual selection: models and experiments. *American Naturalist*, **159**, S22-35.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution*, **16**, 199-204.

Kolbe JJ, Glor RE, Schettino LRG *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177-181.

Kolbe JJ, Larson A, Losos JB (2007) Differential admixture shapes morphological variation among invasive populations of the lizard *Anolis sagrei*. *Molecular Ecology*, **16**, 1579-1591.

Kovalev OV, Poprawski TJ, Stekolshchinov AV *et al.* (1991) *Diuraphis Aizenberg* (Hom., Aphididae): key to apterous viviparous females and a review of Russian language literature on the natural history of *Diuraphis noxia* (Kudjumov, 1913). *Journal of Applied Entomology*, **112**, 425-436.

Lapitan NLV, Li YC, Peng JH, Botha AM (2007) Fractionated extracts of Russian wheat aphid eliciting defense responses in wheat. *Journal of Economic Entomology*, **100**, 990-999.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution*, **17**, 386-391.

Lee W, Kim H, Lim J *et al.* (2011) Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. *Molecular Ecology Resources*, **11**, 32-37.

Levine JM (2008) Biological invasions. *Current Biology*, **18**, R57-60.

Li GY, He FD (1990) Preliminary research on *Diuraphis noxia. Xinjiang Agricultral Science*, **1**, 22-23.

Liebhold AM, Tobin PC (2006) Growth of newly established alien populations: comparison of North American gypsy moth colonies with invasion theory. *Population Ecology*, **48**, 253-262.

Liebhold AM, Tobin PC (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology*, **53**, 387-408.

Lindholm AK, Breden F, Alexander HJ *et al.* (2005) Invasion success and genetic diversity of introduced populations of guppies *Poecilia reticulata* in Australia. *Molecular Ecology*, **14**, 3671-3682.

Liu X, Marshall JL, Stary P *et al.* (2010) Global Phylogenetics of *Diuraphis noxia* (Hemiptera: Aphididae), an invasive aphid species: evidence for multiple invasions into North America. *Journal of Economic Entomology,* **103**, 958-965.

Lockwood JA, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, **20**, 223-228.

Loxdale HD, Hardie J, Halbert S *et al.* (1993) The relative importance of short-range and long-range movement of flying aphids. *Biological Reviews of the Cambridge Philosophical Society*, **68**, 291-311.

Lozier JD, Roderick GK, Mills NJ (2007) Genetic evidence from mitochondrial, nuclear, and endosymbiont markers for the evolution of host pant associated species in the aphid genus *Hyalopterus* (Hemiptera: Aphididae). *Evolution* **61**, 1353-1367.

Madhusudhan VV, Miles PW (1998) Mobility of salivary components as a possible reason for differences in the responses of alfalfa to the spotted

alfalfa aphid and pea aphid. *Entomologia Experimentalis et Applicata*, **86**, 25-39.

Mergeay J, Verschuren D, De Meester L (2006) Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 2839-2844.

Michel AP, Zhang W, Jung JK *et al.* (2009) Population genetic structure of *Aphis glycines*. *Environmental Entomology*, **38**, 1301-1311.

Miles PW (1999) Aphid saliva. *Biological Reviews of the Cambridge Philosophical Society*, **74**, 41-85.

Miura T, Braendle C, Shingleton A *et al.* (2003) A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrthosiphon pisum* (Hemiptera : Aphidoidea). *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, **295B**, 59-81.

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5446-5451.

Moran NA, Wernegreen JJ (2000) Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology and Evolution*, **15**, 321-326.

Morrison WP, Peairs FB (1998) Response model concept and economic impact. In, *A response model for an introduced pest – the Russian wheat aphid*. Ed. By Quisenberry SS, Peairs FB, Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD, 1-11.

Muirhead JR, Leung B, van Overdijk C *et al.* (2006) Modelling local and long-distance dispersal of invasive emerald ash borer *Agrilus planipennis* (Coleoptera) in North America. *Diversity and Distributions*, **12**, 71-79.

Mutti NS, Park Y, Reese JC, Reeck GR (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum. Journal of Insect Science*, **6**, 38.

Mutti NS, Louis J, Pappan LK, *et al.* (2008) A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a host plant. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9965-9969.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

Nijhout HF (1999) Control mechanisms of polyphenic development in insects - In polyphenic development, environmental factors alter same aspects of development in an orderly and predictable way. *Bioscience*, **49**, 181-192.

Noma T, Brewer MJ, Pike KS, Gaimari SD (2005) Hymenopteran parasitoids and dipteran predators of *Diuraphis noxia* in the west-central Great Plains of North America: Species records and geographic range. *Biocontrol*, **50**, 97-111.

Peccoud J, Figueroa CC, Silva AX *et al.* (2008) Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Molecular Ecology*, **17**, 4608-4618.

Peccoud J, Ollivier A, Plantegenest M, Simon JC (2009a) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7495-7500.

Peccoud J, Simon JC, McLaughlin HJ, Moran NA (2009b) Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving

endosymbionts. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 16315-16320.

Powell G, Tosh CR, Hardie J (2006) Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annual Review of Entomology*, **51**, 309-330.

Puterka GJ, Burd JD, Burton RL (1992) Biotypic variation in a worldwide collection of Russisan wheat aphid (Hemiptera: Aphididae). *Journal of Economic Entomology*, **85**, 1497-1506.

Puterka GJ, Burd JD, Porter D *et al.* (2007) Distribution and diversity of Russian wheat aphid (Hemiptera: Aphididae) biotypes in North America. *Journal of Economic Entomology*, **100**, 1679-1684.

Puth LM, Post DM (2005) Studying invasion: have we missed the boat? *Ecology Letters*, **8**, 715-721.

Richards CL, Schrey AW, Pigliucci M (2012) Invasion of diverse habitats by few Japanese knotweed genotypes is correlated with epigenetic differentiation. *Ecology Letters*, doi: 10.1111/j.1461-0248.2012.01824.x

Rollins LA, Woolnough AP, Wilton AN *et al.* (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology*, **18**, 1560-1573.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology and Evolution*, **22**, 454-464.

Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305-332.

Servedio MR (2001) Beyond reinforcement: The evolution of premating isolation by direct selection on preferences and postmating, prezygotic incompatibilities. *Evolution*, **55**, 1909-1920.

Shoemaker DD, DeHeer CJ, Krieger MJB, Ross KG (2006) Population genetics of the invasive fire ant *Solenopsis invicta* (Hymenoptera: Formicidae) in the United States. *Annals of the Entomological Society of America*, **99**, 1213-1233.

Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single introduction into the United States. *Journal of the Kansas Entomological Society*, **80**, 319-326.

Shufran KA, Payton TL (2009) Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology*, **102**, 440-445.

Simon JC, Carre S, Boutin M *et al.* (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 1703-1712.

Simon JC, Rispe C, Sunnucks P (2002) Ecology and evolution of sex in aphids. *Trends in Ecology and Evolution*, **17**, 34-39.

Smith CM, Belay T, Stauffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) populations virulent to the Dn4 resistance gene. *Journal of Economic Entomology*, **97**, 1112-1117.

Stary P (1999) Distribution and ecology of the Russian wheat aphid, *Diuraphis noxia* (Kurdj.), expanded to central Europe (Hom.: Aphididae). *Journal of Pest Science*, **72**, 25-30.

Stern DL (2008) Aphids. *Current Biology*, **18**, R504-505.

Stern DL, Foster WA (1996) The evolution of soldiers in aphids. *Biological Reviews of the Cambridge Philosophical Society*, **71**, 27-79.

Strassmann JE, Queller DC (2001) Selfish responses by clone invaders. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11839-11841.

Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 1095-1100.

Tagu D, Klingler JP, Moya A, Simon JC (2008) Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Molecular Plant-Microbe Interactions*, **21**, 701-708.

Thomas S, Boissot N, Vanlerberghe-Masutti F (2012) What do spring migrants reveal about sex and host selection in the melon aphid? *BMC Evolutionary Biology*, **12**, 47.

Tjallingii WF (2006) Salivary secretions by aphids interacting with proteins of phloem wound responses. *Journal of Experimental Botany*, **57**, 739-745.

Tolmay VL, Lindeque RC, Prinsloo GJ (2007) Preliminary evidence of a resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia* (Kurdiumov) (Homoptera: Aphididae), in South Africa. *African Entomology*, **15**, 228-230.

Tosh CR, Powell G, Hardie J (2002) Maternal reproductive decisions are independent of feeding in the black bean aphid, *Aphis fabae*. *Journal of Insect Physiology*, **48**, 619-629.

Ugelvig LV, Drijfhout FP, Kronauer DJC *et al.* (2008) The introduction history of invasive garden ants in Europe: Integrating genetic, chemical and behavioural approaches. *BMC Biology*, **6**.

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.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446-1457.

Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution*, **16**, 381-390.

Wang JF, Jiang LY, Qiao GX (2011) Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). *Zookeys*, **122**, 1-17.

Wei ZM, Wen YL, Yan P *et al.* (1994) Prevention research of wheat pest *Diuraphis noxia. Tacheng Technology,* **1**, 6-10.

Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. *Nature Reviews Genetics*, **3**, 850-861.

Whitmire SL, Tobin PC (2006) Persistence of invading gypsy moth populations in the United States. *Oecologia*, **147**, 230-237.

Will T, Tjallingii WF, Thonnessen A, van Bel AJE (2007) Molecular sabotage of plant defense by aphid saliva. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10536-10541.

Zepeda-Paulo FA, Simon JC, Ramirez CC *et al.* (2010) The invasion route for an insect pest species: the tobacco aphid in the New World. *Molecular Ecology*, **19**, 4738-4752.

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

Zhang RZ, Liang HB, Ren L *et al.* (2001) Induced life cycle transition from holocycle to anholocycly of the Russian wheat aphid (Homoptera, Aphididae). *Science China Life Sciences,* **44**, 1-7.

# Chapter 2

# Population genetics of the Russian wheat aphid in China

This chapter has been published in the journal *Molecular Ecology* and gives an account of the population genetics of RWAs in China. In this paper the genetic diversity and population structure of RWA was investigated using ten microsatellite loci and two mitochondrial DNA genes to identify whether RWAs in China represent a native expansion or a recent introduction.

# Statement of Joint Authorship

B. ZHANG, O. R. EDWARDS, L. KANG, S. J. FULLER. (2012) Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction? *Molecular Ecology*, **21**: 2130–2144.

**Bo Zhang**

Designed and developed experimental protocol, carried out field and laboratory research, analyzed data and wrote manuscript.

**Owain R. Edwards**

Co-supervisor of the study. Project initiation and design. Assisted in the interpretation of data and revision of manuscript.

**Le Kang**

Co-supervisor of this study. Coordinated field sampling and laboratory research in China. Revised manuscript.

**Susan J. Fuller**

Principal supervisor of this study. Project initiation and design. Assisted in the interpretation of data and revision of manuscript.

# 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

## 2.1 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.

## 2.2 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 (Burd *et al.* 2006; Jyoti *et al.* 2006; Smith *et al.* 2004). 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 20$^{th}$ 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; Smith *et al.* 2004; Stary 1999).

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 (Basky 2002; Burd *et al.* 2006; Haley *et al.* 2004; Puterka *et al.* 1992). 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 (Clua *et al.* 2004; Kiriac *et al.* 1990).

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 (Basky 2002; Jimoh *et al.* 2011; Puterka *et al.* 1992; Smith *et al.* 2004), 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 (Charaabi *et al.* 2008; Ferrari *et al.* 2006; 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.

## 2.3 Materials and Methods

*2.3.1 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 (Figure 2.1 & Appendix S2.1). 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.

*2.3.2 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.

*2.3.3 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 with1000 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 1999) 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).

*2.3.4 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 generalized 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.

*2.3.5 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.

## 2.4 Results

*2.4.1 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 S2.2) 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 2.1). In contrast, the lowest allelic number was found in Pishan with 2.7, and lowest allelic richness in Cele with 2.51 (Table 2.2). 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 2.1 &2.2). 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.

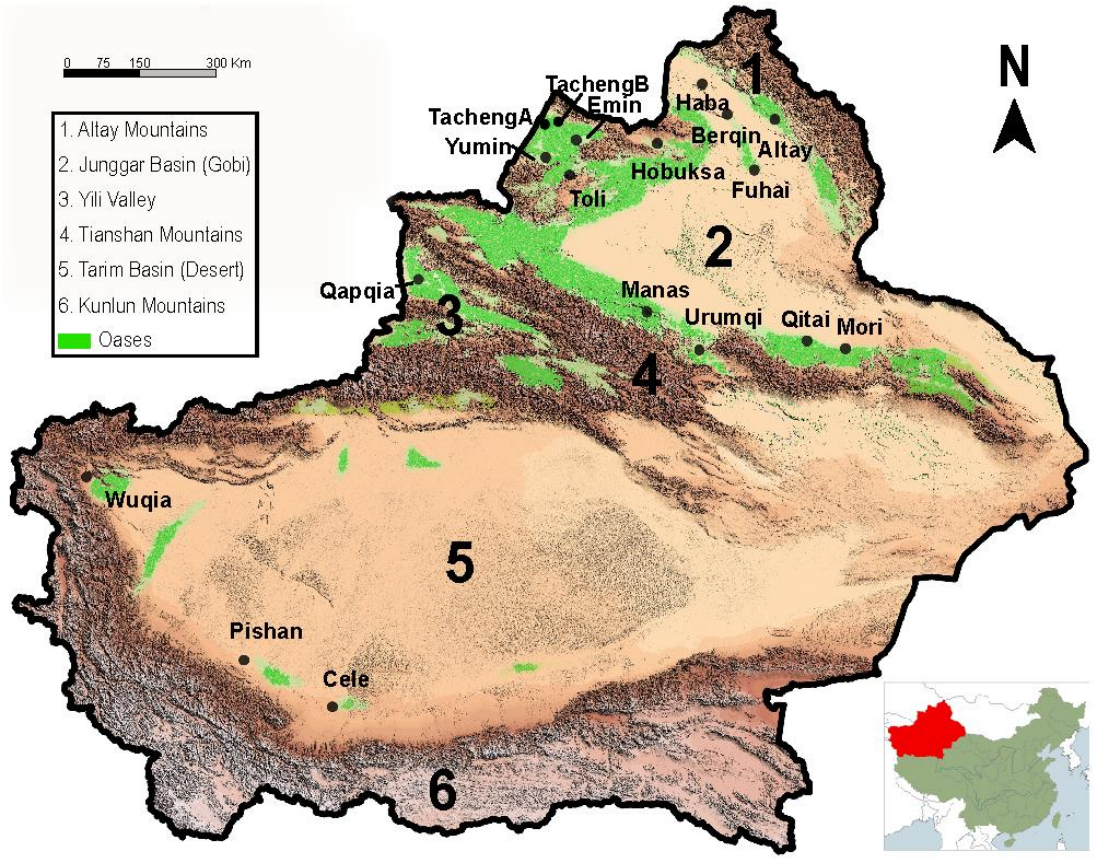



**Figure 2.1** Topographical map of nor thwestern China, Xinjiang, with the sample localities represented by black dots.

**Table 2.1** Indices of genetic diversity for the13 sites sampled in 2009.

| 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* |

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.

**Table2.2** Indices of genetic diversity for the14 sites sampled in 2010.

| 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 |

The abbreviations are the same as indicated in Table 2.1. QP, Qapqia; MS, Manas; WQ, Wuqia; CL, Cele; PS, Pishan.

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 (Figure2.2). 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 S2.3). 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).

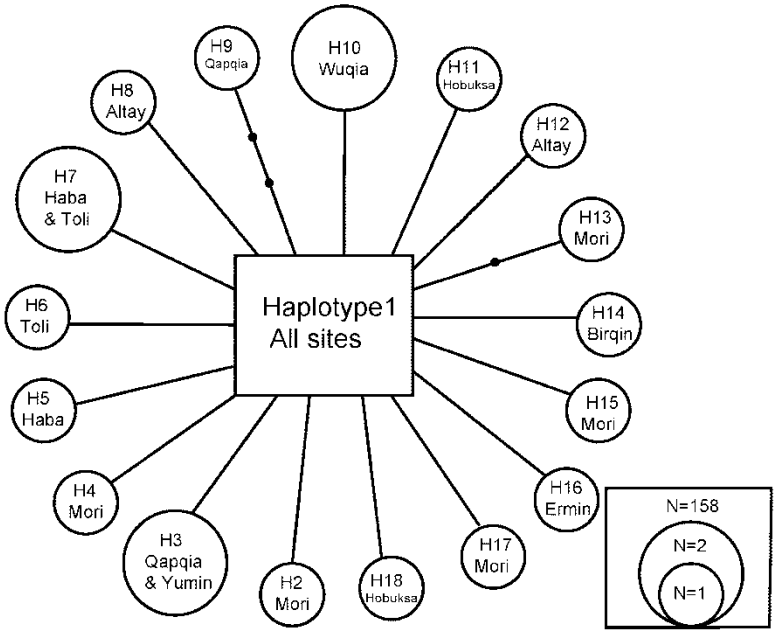



**Figure 2.2** Estimated mitochondrial DNA network with 95% plausible set of haplotype connections. Each haplotype (1-18) is shown as a circle or square. The size of the circle or square relates to the number of individuals sampled (scale shown at base of figure). Small black circles represent putative haplotypes that were not sampled (not labeled). Lines between circles represent a single base pair mutation.

The TCS network revealed a star-like pattern centered on the widely distributed Hap1 (Figure 2.2). 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.

*2.4.2 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 (0.0055 to 0.1129), but majority of pairwise comparisons were significant (Table 2.3A). In 2010, the majority of *F*st were significant among northern sites again, however, a much higher level of differentiation was detected between northern and southern sites (Table 2.3B). 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 S2.4).

**Table 2.3** *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 2.1. 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** | - |

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 Ermin (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.

*2.4.3 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 (Figure 2.3A, Appendix S2.6A). 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 (Figure2.3B). The FCA analysis also identified the three southern sites as distinct from the northern sites, with Pishan genetically intermediate between Cele and Wuqia (Figure2.4). 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 (Figure 3B, Appendix S2.6B).

Evolutionary scenario testing using DIYABC revealed higher posterior probabilities for simultaneous expansion from the three northern sites analyzed (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 S2.7), 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 (Figure 2.2). 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).

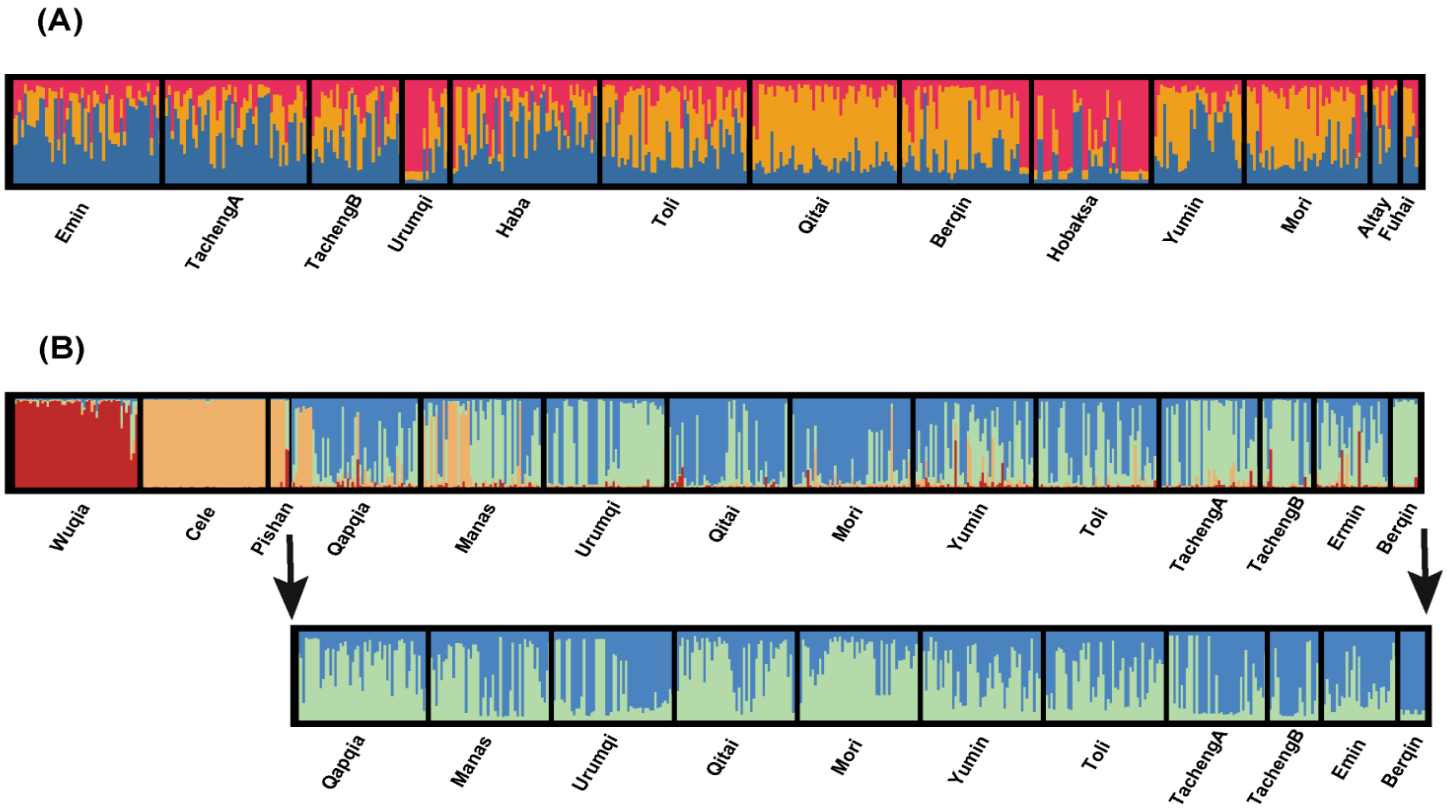



**Figure2.3** Structure bar plot of Chinese RWAs sampled in 2009 (A, *k*=3) and 2010 (B, *k*=4). The 2010 data are also presented following removal of the three southern populations and reanalysis (*k*=2). Each individual is shown as a vertical bar representing ancestry.

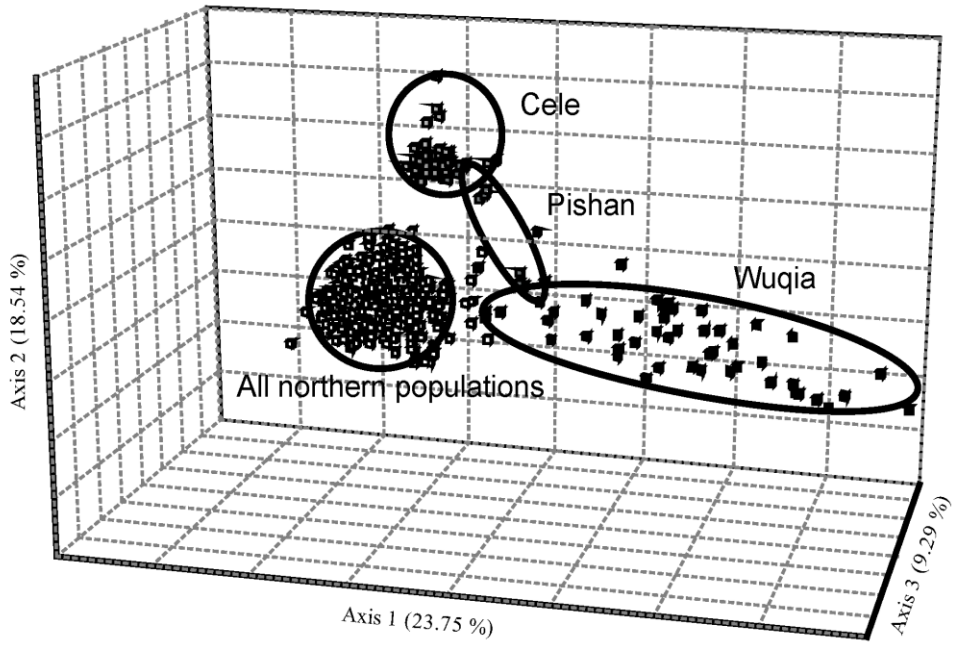



**Figure2.4** Three-dimensional factorial correspondence analysis of Chinese RWAs sampled in 2010. The circles indicate populations that cluster according to geography.

*2.4.4 RWA Population Demographic History*

Population demographic history examined using Bottleneck and Kgtest displayed little evidence for past population fluctuation (Table 2.4). Significant heterozygote deficits were only detected at three sites. Therefore, the reduction in allele number within populations was probably due to founder events rather than rapid decline in population size. Likewise, the L-shaped mode of allele frequency distribution suggests a long-term stable population size. Furthermore, the *k* test was not significant for most sites indicating that the allele length distribution was not significantly different from a binomial distribution and that the population size has been steady. The *g* tests were also not significant providing further evidence of stable population size.

However, when considering all 18 sites as one population, the *k* test indicated significant population expansion. Although the *g* test value was not significant, it was less than one, thus supporting the conclusion of past population expansion in western China.

The mtDNA data also provided evidence of rapid demographic expansion, with the universal haplotype at the center of a star-like cluster formed by the 17 rare haplotypes (Figure2.2). Furthermore, the pairwise mismatch distribution was unimodal, with a strong peak evident at zero, which steeply declined from zero to one base pair. The goodness-of-fit tests were not significant [*p*(SSD)=0.52 and *p*(Harpending's RI)=0.68], and evidence for highly significant population expansion was detected in the R2 statistic (R2=0.08347, *p*=0.002), Tajima's *D* (*D*=-2.39352, *p*<0.01), and Fu's *Fs* (*Fs*=-28.395, *p*<0.0001). The estimated generation time since expansion for Chinese populations was approximately 3,200 years, based on ô value of 0.146 and 1.77%/MY as mutation rate based on the rate given by Papadopoulou *et al.* (2010) for beetle mtDNA.

**Table 2.4** 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 | Hobaksa | Altay | Fuhai | Haba | Berqin | Urumqi | Qitai | Mori | Yumin | Toli | TachengA | TachengB | Emin |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bottleneck | TPM | 1.000 | 0.734 | - | 0.160 | 0.432 | 0.820 | 0.105 | 0.01855* | 0.432 | 0.557 | 0.232 | 0.193 | 0.432 |
| | Shape Mode | L | L | - | L | L | L | L | L | L | L | L | L | L |
| kgtest | K | 5 | 5 | 7 | 6 | 5 | 4 | 6 | 7 | 6 | 6 | 8* | 6 | 5 |
| | G | 1.445 | 0.754 | 0.809 | 0.861 | 0.884 | 2.515 | 1.140 | 1.124 | 0.910 | 0.815 | 0.803 | 0.771 | 0.935 |

| | 2010 | Wuqia | Cele | Pishan | Qapqia | Berqin | Urumqi | Qitai | Mori | Yumin | Toli | TachengA | TachengB | Emin | Manas |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bottleneck | TPM | 0.922 | 1.000 | - | 0.01855* | 0.844 | 0.275 | 0.322 | 0.131 | 0.084 | 0.00488** | 0.232 | 0.570 | 0.426 | 0.129 |
| | Shape Mode | L | L | - | L | L | L | L | L | L | L | L | L | L | L |
| kgtest | K | 6 | 6 | 5 | 8* | 7 | 6 | 6 | 6 | 8* | 8* | 6 | 3 | 6 | 9** |
| | G | 0.546 | 3.397 | 1.859 | 0.851 | 2.813 | 1.088 | 1.377 | 1.320 | 0.743 | 0.784 | 1.095 | 1.160 | 0.781 | 1.149 |

## 2.5 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.

### *2.5.1 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).

*2.5.2 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 behavioral 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 centers 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).

*2.5.3 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 human domestication of wheat.

Our microsatellite data also revealed a signal of population expansion when all sites were combined. Most sites displayed a very slight growth trend, indicating long-term co-evolution of the RWA with its host in natural habitats. Thus, our data are consistent with the theory that long-term effective population size should be in general, closer to the actual size during the

remission period than that in the initial expansion and growth period (Motro & Thomson 1982). In addition, high gene flow among populations of RWA in the north during the expansion and growth period probably enhanced the homogenizing effect, as has been found during an outbreak event of the migratory locust, *Locusta migratoria* (Chapuis *et al.* 2009).

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

Given the potential capacity of RWAs to invade provinces other than Xinjiang, it is interesting that the most easterly site in Xinjiang where RWAs have been detected in the past is Qincheng, located near the border of Gansu province (Du 2000; Zhang *et al.* 1999a). Why have RWAs failed to establish in more Eastern wheat growing districts in China, when the climate is predicted to be conducive (Liang *et al.* 1999)? Though a geographic barrier (eg. Gobi desert) may be responsible, it is also possible that the same environmental factors are limiting range expansion eastward in both China and the USA, which may

be an obligate ecological association with high altitudes in areas where an overwintering stage is required (John Burd, personal communication).

Finally, it is important to consider that in this study we have only sampled RWAs from wheat and thus, we may have examined the genetic structure of only a subsample of the RWAs in the region. Without sampling on other hosts, particularly perennial native hosts, we cannot discount the possibility that we have missed additional unsampled genotypes in the region. In addition, this study has examined the genetic differentiation of RWAs from only a relatively small part of their native range in Asia. However, our results will be critical in guiding future studies of patterns of invasion not only of RWAs, but also of other invasive insect herbivores.

## 2.6 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).

## 2.7 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.

Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713-2719.

Delmotte F, Leterme N, Gauthier JP *et al.* (2002) Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology*, **11**, 711-723.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700-707.

DiLeo M, Row JR, Lougheed SC (2010) Discordant patterns of population structure for two co-distributed snake species across a fragmented Ontario landscape. *Diversity and Distributions*, **16**, 571-581.

Dixon AFG (1998) Aphid ecology, 2nd edition. Chapman and Hall, London.

Dolatti L, Ghareyazie B, Moharramipour S *et al.* (2005) Evidence for regional diversity and host adaptation in Iranian population of the Russian wheat aphid. *Entomologia Experimentalis et Applicata,* **114**, 171-180.

Du BR (2000) Investigation of the distribution and suitable area of Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) in Xinjiang. *Xinjiang Agricultral Science,* **Z1**, 82-85.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.

Ferrari J, Godfray HC, Faulconbridge AS *et al.* (2006) Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution*, **60**, 1574-1584.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.

Fu YX (1997) Statistical methods of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.

Goodisman MAD, Matthews RW, Crozier RH (2001) Hierarchical genetic structure of the introduced wasp *Vespula germanica* in Australia. *Molecular Ecology*, **10**, 1423-1432.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices, Version 2.9.3. Available from http:// www.unil.ch/izea/softwares/Fstat.html.

Greater Mekong Subregion Agricultural Information Network: http://www.gms-ain.org/Z_Show.asp?ArticleID=1703.

Haley SD, Peairs FB, Walker CB *et al.* (2004) Occurrence of a new Russian wheat aphid biotype in Colorado. *Crop Science,* **44**, 1589-1592.

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

Harrison JS, Mondor EB (2011) Evidence for an invasive aphid “superclone”: Extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States. *PLoS One*, **6**, e17524.

Jimoh MA, Botha CEJ, Edwards O, Bradley G (2011) Population growth rate and relative virulence of the two South African biotypes of Russian wheat aphid*, Diuraphis noxia*, and bird cherry-oat aphid, *Rhopalosiphum padi*, on resistant and non-resistant barley. *Entomologia Experimentalis et Applicata*, **138**, 12-20.

Jyoti JL, Qureshi JA, Michaud JP *et al.* (2006) Virulence of two Russian wheat aphid to eight wheat cultivars at two temperature. *Crop Science*, **46**, 774-780.

Kiriac I, Gruber F, Poprawski TJ *et al.* (1990) Occurrence of sexual morphs of Russian wheat aphid, *Diuraphis noxia* (Homoptera: Aphididae), in several locations in the Soviet Union and the Northwestern United States. *Proceedings of the Entomological Society of Washington*, **92**, 544-547.

Kovalev OV, Poprawski TJ, Stekolshchinov AV *et al.* (1991) *Diuraphis Aizenberg* (Hom., Aphididae): key to apterous viviparous females and a review of Russian language literature on the natural history of *Diuraphis noxia* (Kudjumov, 1913). *Journal of Applied Entomology*, **112**, 425-436.

Li XQ, Dodson J, Zhou XY *et al.* (2007) Early cultivated wheat and broadening of agriculture in Neolithic China. *The Holocene*, **17**, 555-560.

Liang HB, Zhang RZ, Zhang GX (1999) Prediction of suitable areas for Russian wheat aphid in China. *Acta Entomologica Sinica*, **42S**, 55-61.

Librado P, Rozas J (2009) DnaSP v. 5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* **25**, 1451-1452.

Liu X, Marshall JL, Stary P*et al.* (2010) Global Phylogenetics of *Diuraphis noxia* (Hemiptera: Aphididae), an invasive aphid species: evidence for multiple invasions into northAmerica. *Journal of Economic Entomology,* **103**, 958-965.

Llewellyn KS, Loxdale HD, Harrington R *et al.* (2003) Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Molecular Ecology*, **12**, 21-34.

Loxdale HD, Hardie J, Halber S *et al.* (1993) The relative importance of short-

and long-range movement of flying aphids. *Biological Review*, **68**, 291-311.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655.

Martinez-Torres D, Moya A, Hebert PDN, Simon JC (1997) Geographic distribution and seasonal variation of mitochondrial DNA haplotypes in the aphid *Rhopalosiphum padi* (Hemiptera: Aphididae). *Bulletin of Entomological Research*, **87**, 161-167.

Michel AP, Zhang W, Jung KJ *et al.* (2009) Population genetic structure of *Aphis glycines. Environmental Entomology*, **38**, 1301-1311.

Motro U, Thomson G (1982) On heterozygosity and the effective size of populations subject to size changes. *Evolution*, **36**, 1059-1066.

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659-1672.

Peccoud J, Ollivier A, Plantegenest M, Simon JC (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of National Academy of Science USA,* **16**, 7495-7500.

Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reduction in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502-503.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **1555**, 945-959.

Puterka GJ, Burd JD, Burton RL (1992) Biotypic variation in a worldwide collection of Russisan wheat aphid (Hemiptera: Aphididae).*Journal of*

*Economic Entomology*, **85**, 1497-1506.

Pysek P, Richardson DM, Pergl J *et al.* (2008) Geographical and taxonomic biases in invasion ecology. *Trends in Ecology and Evolution*, **23**, 237-244.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092-2100.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for 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.

Ross KG, Shoemaker DD (2008) Estimation of the number of founders of an invasive pest insect population: the fire ant *Solenopsis invicta* in the USA. *Proceedings of Royal Society B,* **275**, 2231-2240.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Scott JM (2007) Native and exotic plant species exhibit similar population dynamics during Succession. *Ecology*, **88**, 1098-1104.

Shufran KA, Burd JD, Anstead JA, Lushai G (2000) Mitochondrial DNA sequence divergence among greenbug (Homoptera: Aphididae) biotypes: evidence for host-adapted races. *Insect Molecular Biology*, **9**, 179-184.

Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single introduction into the United States. *Journal of the Kansas Entomological Society,* **80**, 319-326.

Shufran KA, Payton TL (2009) Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology,* **102**, 440-445.

Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological*

*Society of America*, **87**, 651-701.

Simon C, Buckley TR, Frati F *et al.* (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved Polymerase Chain Reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 545-579.

Simon JC, Baumann S, Sunnucks P *et al.* (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, **8**, 531-545.

Smith CM, Belay T, Stuffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) populations virulent to the *Dn4* resistance gene. *Journal of Economic Entomology*, **97**, 1112-1117.

Stary P (1999) Distribution and ecology of the Russian wheat aphid, *Diuraphis noxia* (Kurdj.), expanded to Central Europe (Hemiptera: Aphididae). *Journal of Pest Science*, **72**, 25-30.

Sunnucks P, Hales DF (1996) Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510-524.

Sunnucks P, De Barro PJ, Lushai G *et al.* (1997) Genetic structure of an aphid studied using microsatellites: cyclic parthenogenesis, differentiated lineages, and host specialization. *Molecular Ecology*, **6**, 1059-1073.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics 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.

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

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

130-140.

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

## 2.8 Data Accessibility

Mitochondrial sequences: Genbank accessions JN204386 - JN204421.
Microsatellite sequences: Genbank accessions JN204377 -JN204385.
Sample locations: Uploaded as online supporting material.
Microsatellite data and mitochondrial haplotypes: DRYAD
doi:10.5061/dryad.42sh717m.

## 2.9 Appendices

**Appendix S2.1** Sampling information for RWAs in Xinjiang.

| Locality | Abbr. | Coordinates | Year | Altitude (m) | #Colonies per year |
|---|---|---|---|---|---|
| Tacheng A | TCA | 46°50.684N 82°53.842E | 2009, 2010 | 646 | 50, 41 |
| Tacheng B | TCB | 46°52.338N 83°09.170E | 2009, 2010 | 758 | 33, 22 |
| Yumin | YM | 46°06.312N 82°45.744E | 2009, 2010 | 1070 | 31, 50 |
| Toli | TL | 46°02.563N 83°44.165E | 2009, 2010 | 843 | 50, 50 |
| Emin | EM | 46°39.649N 84°01.108E | 2009, 2010 | 656 | 47, 31 |
| Hobuksa | HF | 46°29.905N 86°00.225E | 2009 | 786 | 42 |
| Altay | AL | 47°48.617N 87°56.656E | 2009 | 778 | 13 |
| Fuhai | FH | 46°41.047N 88°02.698E | 2009 | 591 | 6 |
| Berqin | BR | 47°52.029N 87°01.318E | 2009, 2010 | 525 | 44, 11 |
| Haba | HB | 48°09.127N 86°25.901E | 2009 | 601 | 54 |
| Manas | MS | 43°57.621N 86°18.402E | 2010 | 1113 | 50 |
| Urumqi | UR | 43°27.752N 87°29.103E | 2009, 2010 | 1600 | 16, 50 |
| Qitai | QT | 43°39.207N 89°44.469E | 2009, 2010 | 1556 | 50, 50 |
| Mori | ML | 43°46.767N 90°10.618E | 2009, 2010 | 1234 | 44, 50 |

| | | | | | |
|---|---|---|---|---|---|
| Wuqia | WQ | 39°42.533N<br>75°10.611E | 2010 | 2151 | 50 |
| Pishan | PS | 37°02.731N<br>78°54.233E | 2010 | 1904 | 9 |
| Cele | CL | 36°17.280N<br>81°15.683E | 2010 | 2300 | 50 |
| Qapqia | QP | 43°35.892N<br>81°10.075E | 2010 | 1178 | 50 |

**Appendix S2.2** Primer details and indices of genetic variation for the ten microsatellite loci used in this study.

| Locus | Primers (5'-3') F/R | Ta (℃) | Repeat motif | Size ranges (bp) | Na | Ho | Hs | Fis | F(Null) |
|---|---|---|---|---|---|---|---|---|---|
| Dn1 | GCAGGGTTACCAATGTTTC<br>TGAGTAGCAGGTAATTCAGGAG | 53 | $(CAA)_n$ | 218-242 | 9 | 0.050 | 0.067 | 0.209* | 0.19 |
| DnE1 | ATACTATGCGTCCGTCGTCC<br>GCTGGACTTGTTGATGGTGA | 57 | $(CAA)_n$ | 130-154 | 9 | 0.355 | 0.368 | -0.041 | 0.02 |
| Dn5 | AATCGCACCCTGGGCAAC<br>GTGGGATTCTAAACTGAGGGC | 60 | $(CAA)_n$ | 205-277 | 22 | 0.849 | 0.819 | 0.006 | 0.02 |
| Dn6 | TGATCGGCTCCATAAAAAC<br>GTAGCAAGTTTGACCCTAAA | 52 | $(GTT)_n$ | 331-451 | 31 | 0.639 | 0.697 | 0.087* | 0.09 |
| Dn13 | AGATTCTGCCGTATGTGATTC<br>CGCAGCCAACAAGCTATTA | 55 | $(GT)_n$ | 158-258 | 33 | 0.712 | 0.749 | 0.087* | 0.09 |
| Dn16 | GTCCTCGTGGATACTCATCAT<br>AATCGGTGTCAGGTTTCG | 53 | $(TG)_n$ | 116-178 | 26 | 0.726 | 0.741 | 0.023* | 0.07 |
| Dn22 | ACGGATTTAACGCAAATTTTA<br>CGAATGTAATGCGATGTTGC | 55 | $(CA)_n$ | 176-242 | 29 | 0.820 | 0.803 | -0.018 | 0.02 |
| Dn25 | GCGTGATCCGAGGTCTTT<br>GACGATTAGGGAGAAGTGAA | 55 | $(AC)_n$ | 102-124 | 12 | 0.670 | 0.690 | 0.048* | 0.06 |
| Dn27 | TTCTGTGGTAGTGGTCCCG<br>GACCACTCACCCTATCTCAC | 55 | $(GT)_n$ | 180-206 | 11 | 0.694 | 0.638 | -0.027 | 0.01 |
| $Sa4\Sigma^1$ | GTGACGTATACGCGATGCG<br>GACGTCGATATTAGCCTAGCC | 55 | $(AC)_5TT(AC)_{16}$ | 150-176 | 17 | 0.631 | 0.634 | 0.033 | 0.04 |

**Appendix S2.3** Mitochondrial genetic diversity in each population.

*p<0.05; **p<0.01; ***p<0.001

| | N | Number of Haplotypes | Nucleotide Diversity | Theta_S | Tajima's D | Fs |
|---|---|---|---|---|---|---|
| Altay | 9 | 3 | 0.4444 | 0.7359 | -1.3624 | -1.0811* |
| Berqin | 10 | 2 | 0.2 | 0.3535 | -1.1117 | -0.3393 |
| Emin | 10 | 2 | 0.2 | 0.3535 | -1.1117 | -0.3393 |
| Fuhai | 10 | 1 | 0 | 0 | 0 | 0 |
| Haba | 10 | 3 | 0.4 | 0.707 | -1.4009 | -1.1639* |
| Hobusksa | 10 | 3 | 0.4 | 0.707 | -1.4009 | -1.1639* |
| TachengA | 10 | 1 | 0 | 0 | 0 | 0 |
| TachengB | 10 | 1 | 0 | 0 | 0 | 0 |
| Yumin | 10 | 2 | 0.2 | 0.3535 | -1.1117 | -0.3393 |
| Toli | 10 | 3 | 0.4 | 0.707 | -1.4009 | -1.1639* |
| Mori | 10 | 6 | 1.2 | 1.7674 | -1.7411* | 3.2939** |
| Qitai | 10 | 1 | 0 | 0 | 0 | 0 |
| Urumqi | 10 | 1 | 0 | 0 | 0 | 0 |
| Manas | 10 | 1 | 0 | 0 | 0 | 0 |
| Pishan | 8 | 1 | 0 | 0 | 0 | 0 |
| Wuqia | 11 | 2 | 0.3273 | 0.3414 | -0.1 | 0.3563 |
| Cele | 10 | 1 | 0 | 0 | 0 | 0 |
| Qapqia | 10 | 3 | 0.6 | 1.0605 | -1.5622* | -0.4586 |
| **Global** | 178 | 18 | 0.196 | 2.968 | 2.3935** | -28.395*** |

**Appendix S2.4** AMOVA for RWAs sampled in 2010 and analyzed in three groups: Wuqia, Cele and Pishan, and northern populations.

***: p<0.0001

| Source of Variation | df | Variance components | Percentage of Variation | Fixation Indices |
|---|---|---|---|---|
| Among regions | 2 | 0.45170 | 12.42 | FCT = 0.12416*** |
| Among sites within regions | 11 | 0.16111 | 4.43 | FSC = 0.05056*** |
| Within sites | 1128 | 3.02531 | 83.16 | FST = 0.16844*** |

## Appendix S2.5

**(A**) Pairwise Fst and Exact G test of allele frequency differentiation for each site sampled in 2009 and 2010. Significant values (following Bonferroni correction) are given in bold.

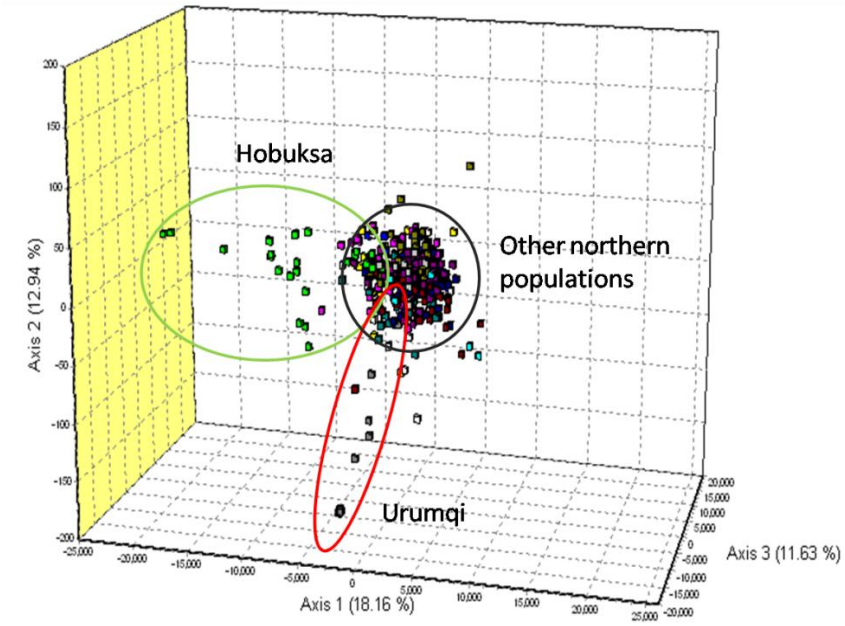
| Site | Fst | Exact G probability |
|---|---|---|
| Ermin | 0.003 | 0.05154 |
| Tacheng A | **0.015** | **0.00001** |
| Tacheng B | **0.060** | **0.00001** |
| Urumqi | **0.123** | **0.00001** |
| Toli | **0.012** | **0.00009** |
| Qitai | 0.008 | **0.00001** |
| Berqin | **0.120** | **0.00001** |
| Yumin | 0.009 | **0.00002** |
| Mori | **0.017** | **0.00001** |

**(B)** AMOVA analysis in 9 groups (each group is a site sampled in 2009 and 2010)

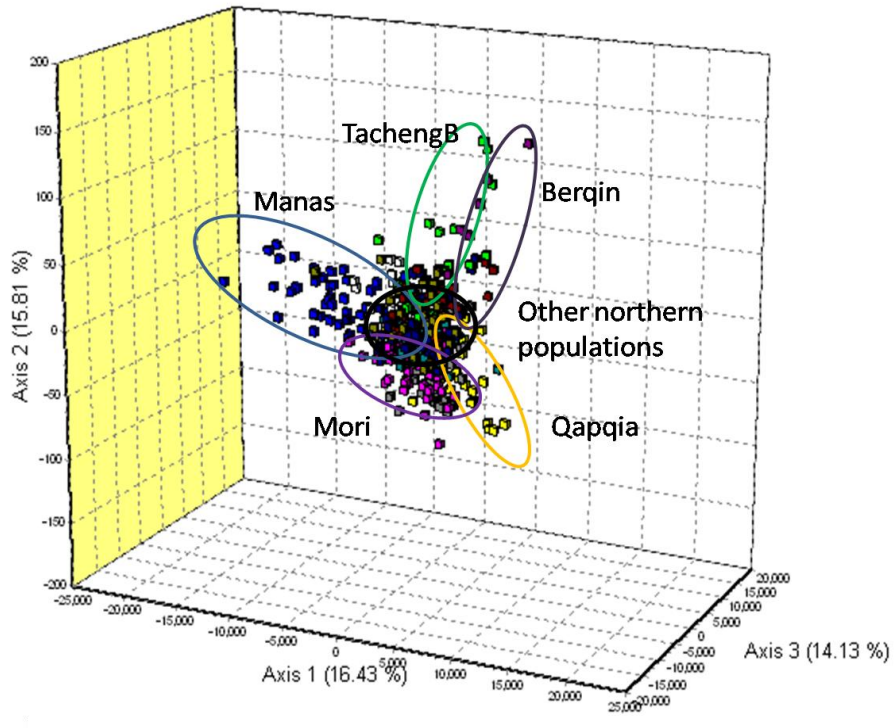
| Source of Variation | df | Variance components | Percentage of Variation | Fixation Indices |
|---|---|---|---|---|
| Among sites within years | 8 | 0.02246 | 0.68 | FCT: **0.00679** |
| Between years within sites | 9 | 0.09525 | 2.88 | FSC: **0.02897** |
| Among individuals within years | 700 | 0.09773 | 2.95 | FIS: **0.03061** |
| Among individuals within sites | 718 | 3.09471 | 93.49 | FIT: **0.06509** |

**Appendix S2.6** FCA of Chinese RWAs from northern populations sampled in 2009 (A) and 2010 (B).

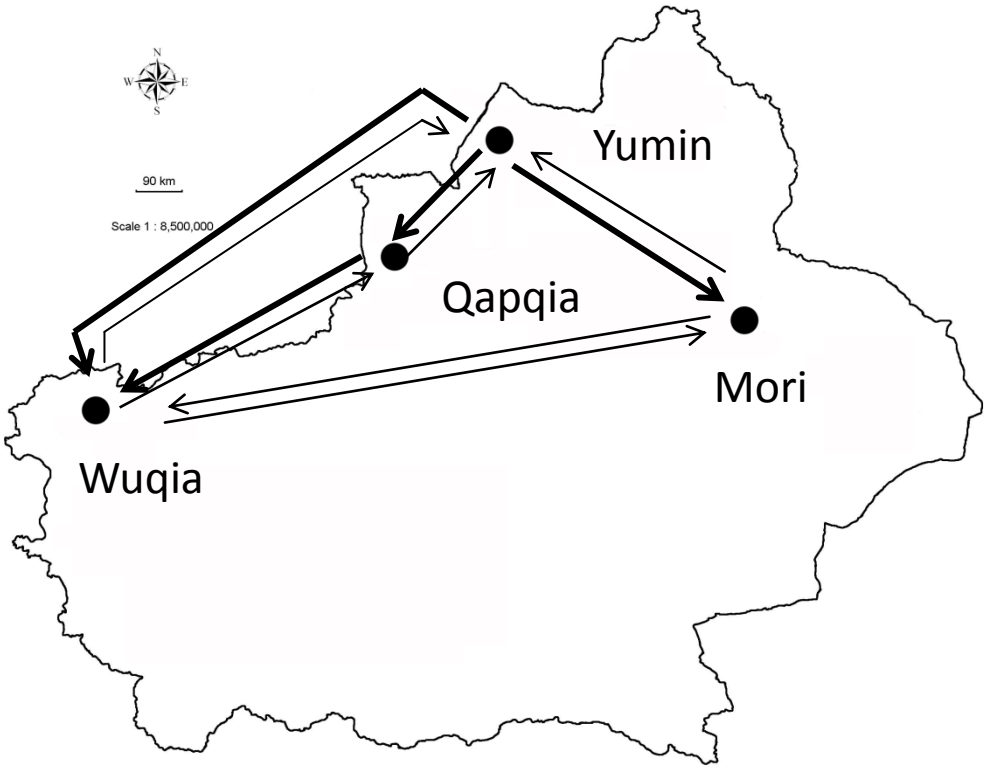
**(A)**




**(B)**



**Appendix S2.7** Gene flow patterns of RWAs in far eastern ranges based on long-term estimates of gene flow.




Arrows indicate direction of gene flow between each population pair, and the relative thickness of each arrow represents the amount of directional gene flow.

# Chapter 3

# Worldwide invasion genetics of RWA

Chapter 2 examined the population genetics of RWAs in their native range. Chapter 3 extends this research by including both native and introduced RWA populations in an analysis of the worldwide invasion genetics of RWAs. Three sources of genetic data were examined; microsatellite, mitochondrial and endosymbiont DNA. The results indicate the invasion pathways of RWAs to the New World, and the most likely source sites of these invasions.

# Statement of Joint Authorship

B. ZHANG, O. R. EDWARDS, L. KANG, S. J. FULLER. Worldwide invasion genetics of the Russian wheat aphid (*Diuraphis noxia*).

**Bo Zhang**

Designed and developed experimental protocol, carried out field and laboratory research, analyzed data and wrote manuscript.

**Owain R. Edwards**

Co-supervisor of the study. Project initiation and design. Assisted in the interpretation of data and revision of manuscript.

**Le Kang**

Co-supervisor of this study. Coordinated field sampling and laboratory research in China. Revised manuscript.

**Susan J. Fuller**

Principal supervisor of this study. Project initiation and design. Assisted in the interpretation of data and revision of manuscript.

# Worldwide invasion genetics of the Russian wheat aphid (*Diuraphis noxia*)

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

[1]Science & Engineering Faculty, 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]CSIRO Ecosystem Sciences, Centre for Environment and Life Sciences, Underwood Avenue, Floreat, WA 6014, Australia; [4]Cooperative Research Centre for National Plant Biosecurity, LPO Box 5012, Bruce, ACT 2617, Australia.

**Keywords**: *Diuraphis noxia*, microsatellite, mitochondrial, endosymbiont, genetic diversity, genetic structure, invasion route

**Running title**: Invasion genetics of *Diuraphis noxia*

***Correspondence**:

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

## 3.1 Abstract

We investigated the population genetics, demographic history and pathway of invasion of the Russian wheat aphid (RWA) from its native range in East Asia, the Middle East and Europe, to South Africa and the Americas. We screened microsatellite markers, mitochondrial DNA, and endosymbiont genes in 504 RWAs from eighteen sites worldwide. Higher genetic and clonal diversity was found in native populations compared with invasive populations. Strong genetic structure was identified among native populations, including Ethiopia. Following pathway analysis of microsatellite and endosymbiont data, we conclude that Turkey and Syria were the most likely sources of invasion to Kenya and South Africa, respectively. Furthermore we found that one clone transferred between South Africa and the Americas was responsible for the New World invasion. The comparison of native and invasive populations suggested two separate invasion events of RWA from the Middle East to Kenya and South Africa, most likely via human-aided transportation of contaminated materials. In addition, we did not detect any host-based differentiation in worldwide RWA populations. This study has provided valuable insights into the factors that may have facilitated the recent global invasion by this damaging pest. Finally, endosymbiont DNA was found to be a high resolution population genetic marker, extremely useful for studies of recent invasion over a relatively short evolutionary history timeframe.

## 3.2 Introduction

Biological invasion is regarded as a worldwide environmental problem (Everett 2000; Levine 2008; Liebhold & Tobin 2008; Mooney & Cleland 2001). Concomitant with increasing human population densities, increasing human movement has facilitated the entry of a variety of plants and animals into new ecosystems beyond their native ranges (Kolar & Lodge 2001). Irrespective of whether the introduction is intentional or accidental, the invasion of an alien species influences and accelerates the ecological evolution of its new habitat (Mooney & Cleland 2001).

Invasive species have the ability to respond quickly and to adapt to new and/or challenging environmental conditions and genetic diversity can positively or negatively influence the success of a invading species (Dlugosch & Parker 2008; Mergeay *et al.* 2006; Rollins *et al.* 2009; Shoemaker *et al.* 2006). Invasive species exhibit a high degree of phenotypic plasticity and rapid phenotypic response has been found to enhance invasion success (Frankham 2005; Kolbe *et al.* 2004; Lindholm *et al.* 2005). However, the opposite has also been found where exotic species with low genetic variability are limited in their invasion success because they are constrained in their ability to adapt to a new environment (Roman & Darling 2005). If the invasive population is founded by a limited number of individuals and mating occurs between related individuals, the genetic background will be further constrained as additive genetic variance is eroded (Lindholm *et al.* 2005). Such a loss of additive genetic variance will slow evolutionary responses to selection and will limit the adaptive evolution of fitness-related traits.

Comparative studies that examine the genetic diversity of a species in both its introduced and native range allow inferences to be made on the pathway and source of an introduction. Research on the garden ant, *Lasius neglectus*, has suggested that the fourteen introduced populations in Europe have most likely arisen from only a very few independent introductions from the native range, and new infestations were typically started through introductions from other invasive populations (Ugelvig *et al.* 2008). Importantly, comparative

studies throughout an invasive species distributional range can provide valuable information on the chronological order of colonization (Rollins *et al.* 2009), particularly when historical records of first incursion are sparse or incomplete.

Knowledge of the genetic background of the founding population of an invasive species will provide an improved ability to predict the array of evolutionary responses and impacts that may result, as well as the future distribution of the invasive species. Researchers are now armed with a range of innovative and reliable molecular tools that can be used to improve our understanding of the population genetic structure of invasive species in their native and introduced ranges. Importantly, we can elucidate the occurrence of single vs. multiple introductions providing critical information on the mechanisms by which invasive species are introduced and spread, enabling pre-emptive action to prevent future incursions into new ranges. Critical data can be obtained on how frequently introductions are accompanied by severe genetic bottlenecks, whether bottlenecks constrain adaptive evolution in invaders, and whether contemporary gene flow among introductions represents the key to invasion success.

In this study we have used a molecular approach to examine the global pattern of invasion of the Russian wheat aphid (RWA), *Diuraphis noxia* Kurdjumov, one of the world's most invasive cereal (wheat and barley) crop pests (Burd *et al.* 2006; Jyoti *et al.* 2006; Smith *et al.* 2004). Current opinion favors the center of origin of the native distribution of RWAs to be in the Iranian-Turkestanian mountain range and extending to southern Russia, the Middle East, and central Asia (Kovalev *et al.* 1991). The earliest documented record of RWA crop damage comes from the Ukraine in the early 1900s, however they gained recognition as an emerging global pest during the 1970's and 1980's as a result of their rapid spread through major grain producing areas in Europe, Africa and the Americas (Kovalev *et al.* 1991; Stary 1999; Smith *et al.* 2004). Although RWAs have been recognized as an invasive pest in China since they were first detected in 1976, recent

population genetic studies have shown that western China forms part of this species native range (Zhang *et al.* 2012).

While there is no direct evidence that RWA invasiveness is related to some biological, genetic and/or ecological response, the dispersal and stable spread of this species is likely to be linked with the presence of compatible host plants and mode of reproduction. RWAs exhibit both cyclical (holocyclic) and obligate (anholocyclic) parthenogenesis depending on environmental conditions (in particular minimum winter temperature). Native populations are generally, but not exclusively holocyclic. In Iran, for example, both modes of reproduction have been observed (Dolatti *et al.* 2005). Invasive populations of RWA are primarily anholocyclic, although the appearance of sexual females and eggs has been reported recently in North America and Argentina (Clua *et al.* 2004; Kiriac *et al.* 1990). Parthenogenetic reproduction enables rapid population growth and is likely to be a key factor influencing invasion success, particularly in countries experiencing mild winters where cereal crops are planted all year round or native grasslands are in close proximity providing an alternative overwintering host.

Most of the population genetics research on RWAs has focused on characterizing levels of genetic diversity in invasive populations in North America, South America, and South Africa (Shufran *et al.* 2007; Shufran & Payton 2009). While some comparative studies of this species in its native and introduced ranges have been undertaken, they have not clearly resolved the pattern or pathway of worldwide invasion (Liu *et al.* 2010; Puterka *et al.* 1992; Stary 1999), primarily due to the limited polymorphism of the markers. This is a common problem in pathway analysis: maternally-inherited markers are ideal for tracking lineages, but there is often insufficient genetic variability in mitochondrial genes. In this study, we identify a solution to this problem by targeting fast-evolving regions of DNA in the aphid endosymbiont, *Buchnera aphidicola*. *B. aphidicola* has an obligatory mutualistic association with its host and is responsible for biosynthesizing up to 90% of the aphid's essential amino acids (Douglas 2006). Endosymbiont genome evolution has been shown to reflect the aphid host genome, but at an accelerated rate due to the

rapid generation time of the bacteriocyte relative to its host (Burke & Moran 2011; Clark *et al.* 1999; Peccoud *et al.* 2009a & b).

Here we report results of the most comprehensive documentation of the invasion genetics of RWAs worldwide. This involved characterizing the levels and patterns of genetic diversity within and among native and introduced RWA populations using data from four genomes (nuclear and mitochondrial RWA, plasmid and genomic endosymbiont). We provide strong evidence for native demographic population expansion in the Middle East and eastern Asia, and we have reconstructed the pathway of invasion of RWAs from their native distribution in the Old World to their invasive distribution in the New World.

## 3.3 Material and Methods

*3.3.1 Aphid sampling*

Russian wheat aphids were sampled from wheat, rye, and barley crops and native grasses from eighteen sites worldwide (Table 3.1 & Appendix S3.1). In total, 504 parthenogenetic, wingless female aphids were collected and preserved in 100% ethanol until DNA extraction.

*3.3.2 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 10 microsatellite loci, including nine loci developed from RWAs (Zhang *et al.* 2012), and one cross-species locus, Sa4Σ, developed from *Sitobion* aphids (Simon *et al.* 1999). Microsatellite loci were amplified in a total volume of 10μl and PCR cycling conditions followed Zhang *et al.* (2012). 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.

Two mitochondrial DNA fragments were sequenced: 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 Zhang *et al.* (2012).

Three endosymbiont genes from the *Buchnera aphidicola* genomes of RWA were sequenced: two from plasmids, TrpEG pseudogene (Wernegreen & Moran2000) and LeuBC, as well as one from genomic DNA, ddlB pseudogene. The PCR cycling parameters were as follows: initial

denaturation 3 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 53°C (LeuBC) or 55°C (ddlB and TrpEG), 3min at 72°C; the final elongation of 10 min at 72°C. The PCR products were visualized by electrophoresis in a 1.5% agarose gel and purified using an ABgene Ultra PCR Clean-Up Kit (Thermo Scientific) according to manufacturer's directions, before being sequenced on a 3730xl DNA Analyser (Applied Biosystems). The three genes concatenated gave 3,258 bps of edited sequence.

*3.3.3 Genetic Diversity*

Microsatellite data were examined using Micro-Checker v2.2.3 (Van Oosterhout *et al.* 2004) for any evidence of large allele dropout or scoring errors. Each individual RWA was classified according to multilocus genotype (MLG) in GenClone v2.0 (Arnaud-Haond & Belkhir 2007). Gross genotypic diversity (GGD) was calculated as G/N, with G equal to the number of MLGs, and N equal to the sample size (Llewellyn *et al.* 2003). The dataset was pruned to include only one copy of each MLG and then genetic diversity estimates were calculated using FSTAT v2.9.3. (Goudet 2001). Observed and expected heterozygosity (Ho & He), number of alleles (Na), allelic richness (Ar), and the *f* estimator of *F*is and significance values (Weir & Cockerham 1984) were calculated. Genepop v4.0 was used to calculate allele frequencies, Hardy-Weinberg equilibrium (HWE), and linkage disequilibrium tests (Raymond & Rousset 1995; Rousset 2008) and significance was assessed following Bonferroni correction (Rice 1989).

Mitochondrial and endosymbiont DNA sequences were aligned and edited using BioEdit v7.0.0 (Hall 1999) and MEGA v4.1 (Tamura *et al.* 2007). DnaSP v5 was used to calculate the number and frequencies of haplotypes (Librado& Rozas 2009), and phylogeographic networks were inferred using TCS (Clement *et al.* 2000) and NETWORK 4.6 (Bandelt *et al.* 1999). To infer deviations from neutrality and to detect demographic changes or selection (Fu & Li 1993), Tajima's *D* (Tajima 1989) and Fu's *Fs* (Fu 1997) were calculated from the mitochondrial DNA data using Arlequin v3.5.1.2 (Excoffier *et al.* 2005).

*3.3.4 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. Three clustering methods were used to identify population structure. Firstly, using Structure v2.1 (Pritchard *et al.* 2000) an admixture model was implemented assuming independent allelic frequencies, with 100,000 iterations of Markov Chain Monte-Carlo after a 20,000 burn-in period. Ten independent runs of each *K* (inferred clusters) from 1 to 20 were evaluated. The most likely *K* value was determined using the Evanno method (Evanno *et al.* 2005) and then Distruct v1.1 (Rosenberg 2004) was used to display the bar plot. Secondly, the three-dimensional spatial distribution of genetic variation for each individual was examined using factorial correspondence analysis (FCA) implemented in Genetix v4.05 (Belkhir *et al.* 1996-2004). Finally, to confirm population clusters and to differentiate the variation component among populations, an analysis of molecular variance (AMOVA) was conducted in Arlequin.

We also measured the relatedness of each population based on microsatellite allele shared distance using Population v1.2.3 (http://bioinformatics.org/~tryphon/populations/). The neighbor-joining tree was calculated based on Cavalli-Sforza's chord distance (Dc) from ten microsatellite loci, and 1000 bootstrap replications were performed to test the robustness of the nodes.

Five evolutionary scenarios of expansion and gene flow among sites were examined using DIYABC v 0.7 (Cornuet *et al.* 2008). The posterior distributions of one evolutionary scenario of simultaneous expansion from the Middle East to the Americas and South Africa and four scenarios of stepwise expansion (including potential admixture events) were evaluated. 15000 sets of simulated data that were closest to the observed microsatellite data were compared using logistic regression. The following assumptions were made; a stable effective population size ($N_e$), a transitory bottleneck (db=5) and a generalised stepwise model (GSM) of mutation. 500000 simulated datasets were produced for each scenario.

## 3.4 Results

*3.4.1 Genetic diversity (nDNA)*

From a total of 504 RWAs screened for ten microsatellite loci, 370 MLGs were detected across the eighteen sites. Only one MLG was shared among sites ($MLG_{AM}$). $MLG_{AM}$ was detected at all four sites in the Americas (Table 3.1). A second MLG was detected in the USA, but this MLG only differed from $MLG_{AM}$ by one repeat (one allele out of 20 was different). Interestingly, although none of the MLGs present in Asia, the Middle East and Africa were identical to $MLG_{AM}$, one MLG in Turkey was different to $MLG_{AM}$ by two alleles at one locus and one MLG in South Africa differed from $MLG_{AM}$ by only one repeat in one allele. Similarly, the other three MLGs from South Africa were all derivatives of $MLG_{AM}$, differing by two alleles, each only one repeat different. Highest genetic diversity was found in Asian, Middle Eastern and Ethiopian samples, with the highest MLG diversity found in found in Asia, Ethiopia and Syria (GGD=1.00), the highest allelic richness in Ethiopia (4.23) and the highest number of alleles (11.1) and gene diversity (0.71) in China. Low genetic diversity was found in all the putative introduced populations (Table 3.1).

HWE tests involving 190 pairwise comparisons (locus by population) revealed some significant departures, however, no consistent pattern was detected for any population or locus. Both deficit and excess observed heterozygosity were found in native populations, while in the putative invasive populations, only excess observed heterozygosity was evident. No linkage disequilibrium was observed for any pair of loci after Bonferroni correction.

*3.4.2 Genetic diversity (mtDNA)*

Concatenated, 1272bp of mitochondrial DNA was obtained from 155individuals from eighteen sites worldwide. Haplotype diversity (Hd) was 0.24. Thirteen haplotypes were identified, with one common haplotype found at all sites shared by 135 individuals. The frequency of the common haplotype was 87.1% and 12 rare haplotypes were found at low frequencies

(0.6-4.5%). Apart from the common haplotype, one other haplotype was shared among sites (Hap10 was found at Yumin and Qapqia in China), and two haplotypes were shared within a site (Hap2 was found in seven individuals from Iran, and Hap12 was found in two individuals from Wuqia in China) (Appendix S3.2). The highest nucleotide diversity was found in China (0.48), while the lowest nucleotide diversity (0) was found in introduced populations.

*3.4.3 Genetic differentiation (nDNA)*

Population genetic differentiation was analyzed using pairwise *F*st. The majority of pairwise *F*st comparisons among sites from Asia, the Middle East and Ethiopia were significant (64 out of 66), with *F*st values ranging from 0.008 to 0.427 (Appendix S3.3). Among introduced populations, *F*st values were around zero and not significant.

The Neighbor-joining tree based on genetic distance (Dc) revealed two clear groups, one including introduced populations, and the other including the twelve native populations plus the North African sites of Ethiopia and Kenya (Figure 3.1). The native populations were further separated into two geographical sub-groups including sites from the Middle East and East Asia.

**Table3.1** Indices of genetic diversity for the 17 sites with a sample size (N) greater than four.

| Region | Site | N | MLGs | #within | #among | GGD | Ho | Na | Ar | Hs | Fis |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **East Asia** | **Qapqia (Ch)** | 53 | 51 | 2 | 0 | 0.96 | 0.65 | 9.8 | 3.77 | 0.65 | 0.01 |
| | **Yumin (Ch)** | 31 | 30 | 1 | 0 | 0.97 | 0.67 | 8.2 | 3.90 | 0.68 | 0.07 |
| | **Mori (Ch)** | 42 | 42 | 0 | 0 | 1.00 | 0.64 | 9.4 | 3.73 | 0.64 | 0.04 |
| | **Haba (Ch)** | 50 | 48 | 2 | 0 | 0.96 | 0.69 | 11.1 | 4.04 | 0.69 | 0.098* |
| | **Cele (Ch)** | 52 | 41 | 9 | 0 | 0.79 | 0.43 | 3.9 | 2.07 | 0.34 | -0.413* |
| | **Wuqia (Ch)** | 52 | 52 | 0 | 0 | 1.00 | 0.71 | 8.1 | 3.93 | 0.71 | 0.17* |
| | **Tajikistan** | 30 | 25 | 4 | 0 | 0.83 | 0.70 | 5.9 | 3.59 | 0.70 | -0.151* |
| **Middle East** | **Turkey** | 49 | 26 | 6 | 0 | 0.53 | 0.73 | 6.6 | 3.52 | 0.65 | -0.08 |
| | **Iran** | 18 | 12 | 3 | 0 | 0.67 | 0.75 | 5.6 | 3.23 | 0.67 | -0.177* |
| | **Syria2000** | 10 | 10 | 0 | 0 | 1.00 | 0.73 | 4.5 | 3.48 | 0.65 | -0.06 |
| | **Syria2011** | 30 | 7 | 2 | 0 | 0.23 | 0.64 | 3.1 | 2.03 | 0.49 | -0.533* |
| **North Africa** | **Ethiopia** | 17 | 17 | 0 | 0 | 1.00 | 0.76 | 7.4 | 4.23 | 0.69 | 0.13 |
| | **Kenya** | 12 | 1 | 1 | 0 | 0.08 | 1.00 | 1.8 | 1.80 | — | — |
| **South Africa** | **Sth Africa** | 33 | 4 | 4 | 0 | 0.12 | 0.62 | 2.1 | 1.92 | 0.45 | -0.794* |
| **Americas** | **Mexico** | 6 | 1 | 1 | 1 | 0.17 | 1.00 | 1.8 | 1.80 | — | — |
| | **US** | 8 | 2 | 1 | 1 | 0.25 | 0.69 | 1.9 | 1.85 | 0.43 | -0.882* |
| | **Chile** | 5 | 1 | 1 | 1 | 0.20 | 1.00 | 1.8 | 1.80 | — | — |
| | **Argentina** | 4 | 1 | 1 | 1 | 0.25 | 1.00 | 1.8 | 1.80 | — | — |

Ch, China; 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); Ho, observed heterozygosity; Hs, gene diversity; Na, numbers of alleles; Ar, allelic richness; Fis, the inbreeding index. Asterisks indicate significance after Bonferroni correction at 0.05 level, - indicates no genetic diversity and grey shading highlights putative invasive populations.

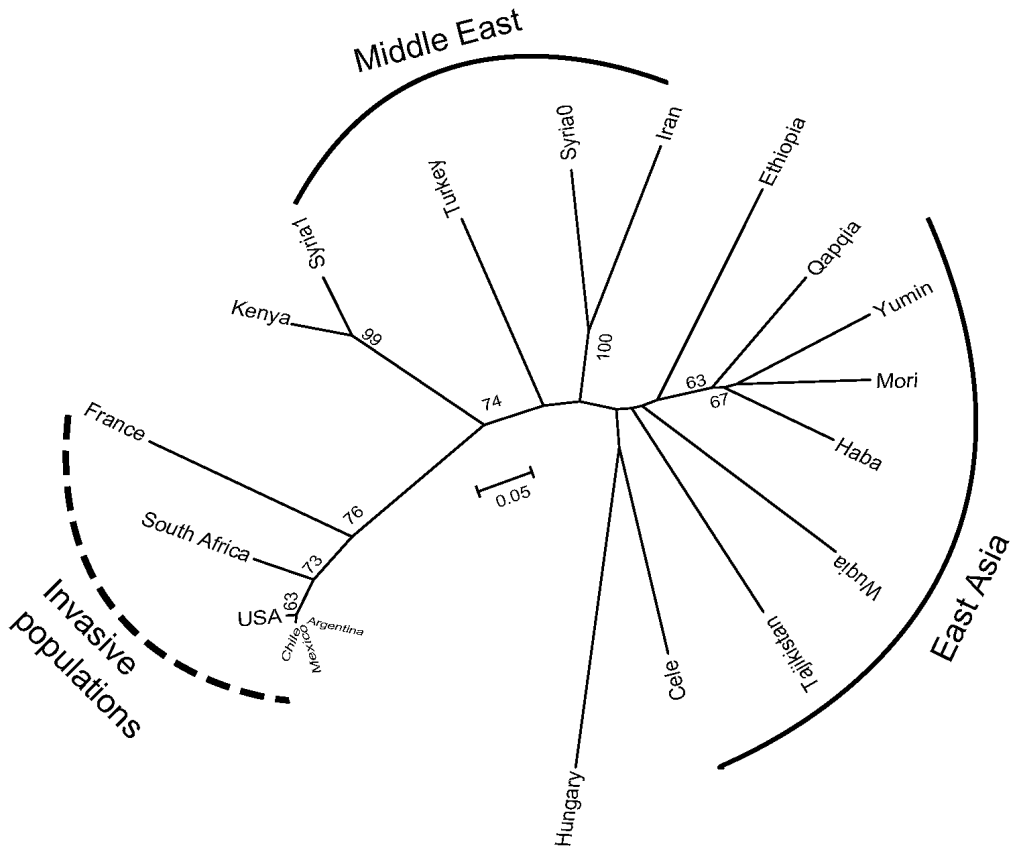



**Figure 3.1** Neighbor-joining population tree based on the chord distance of Cavalli-Sforza & Edwards (Dc). Bootstrap values were calculated using 1000 replications and are given as percentages, with only values greater than 50% shown. Two samples from Syria are labeled Syria0, sampled in 2000; Syria1, sampled in 2011.

The Structure analysis revealed the presence of clear population clusters (Figure 3.2). When we adopted k=5, three clusters from East Asia, one from the Middle East, and one including all introduced populations were separated (Figure 3.2A). Ethiopia was more closely related to the East Asian populations than to the Middle Eastern populations. Tajikistan (in central Asia) is geographically the closest site to Wuqia (in China) and the Structure analysis assigned these sites to the same cluster. The Middle East populations presented a strong pattern of mixed ancestry. When further increasing k to 7, Tajikistan became distinct from Wuqia, and the cluster of Kenya and Syria (sampled in 2011) became isolated from the entire invasive population cluster (Figure 3.2B).

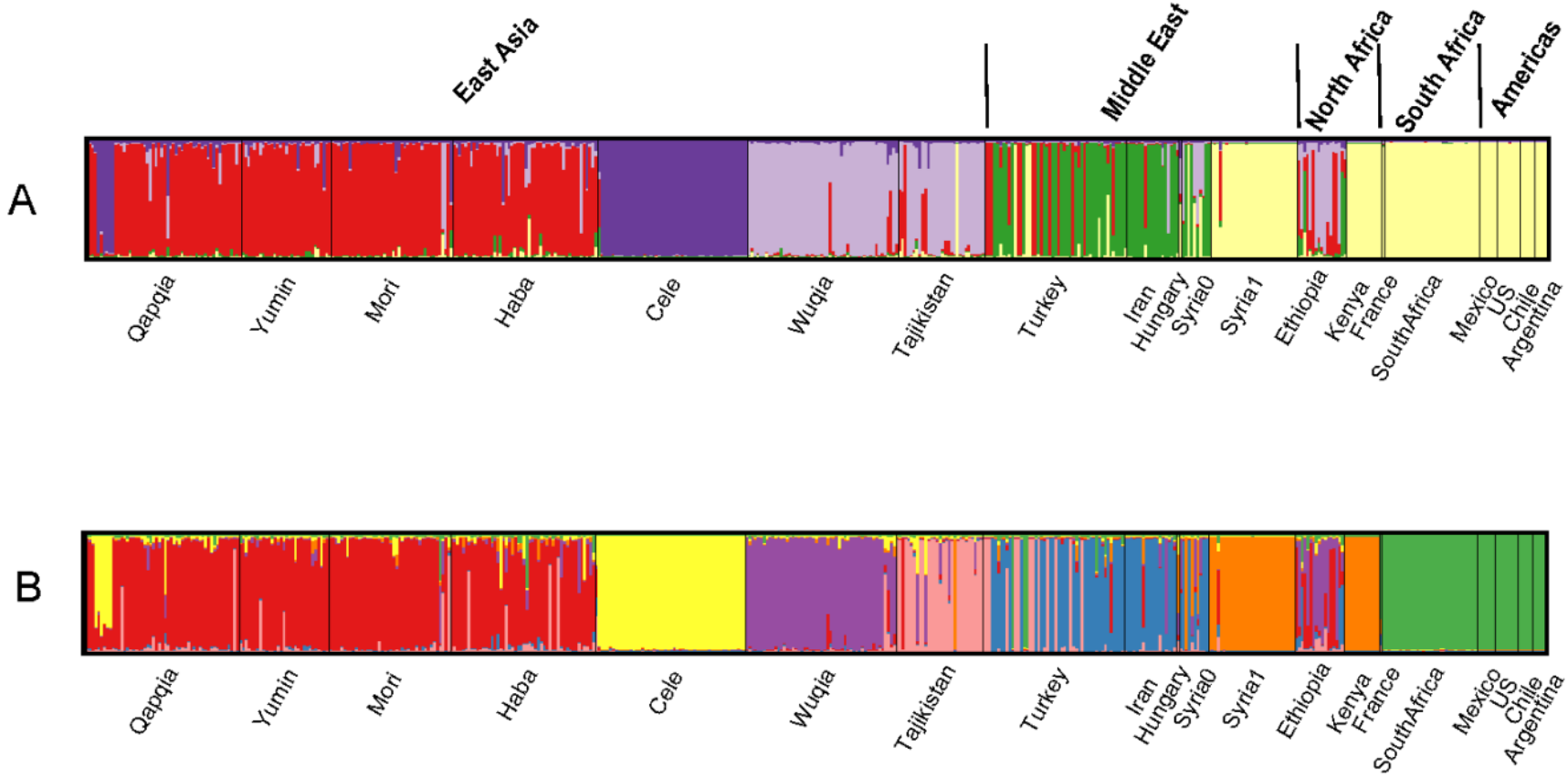



**Figure 3.2** Structure clusters of RWA worldwide. (A) *k*=5; (B) *k*=7. Each individual is shown as a vertical bar representing ancestry.

An AMOVA analysis was conducted and separated sites into seven groups. The proportion of variance among groups (15.91%) was larger than that found among sites within groups (4.79%), and the fixation index (Fct=0.159) was significant, indicating extremely restricted gene flow among these groups (Appendix S3.4).

Factorial correspondence analysis depicted a similar pattern (Appendix S3.5). Accounting for 35% of the variation, the FCA revealed distinct clusters corresponding to geographical isolation in the native range (Middle East and East Asia) and introduced populations. Interestingly, Kenya samples were located between the invasive cluster and the Asian cluster.

*3.4.4 Genetic differentiation (endosymbiont DNA)*

The concatenated sequences of three endosymbiont genes of *Buchnera* showed 108 polymorphic sites resulting in 76 haplotypes among worldwide RWA populations. These polymorphic sites included 43 variable sites differing by a single base pair and 65 parsimony informative sites. Haplotype diversity (Hd) was 0.89, and nucleotide diversity (Pi) was 0.002. The haplotype network revealed two distinct clusters separated by a 21bp insertion in the plasmid psuedogene trpEG (Figure 3.3). This insertion appeared to be the result of a slipped-strand mispairing during DNA replication, as the inserted sequence was identical to the 21bps located either side of the insertion. Haplotypes containing the insertion were found in all individuals in the invasive range except Kenya. The Middle Eastern sites of Turkey, Iran and Syria and one Asian site (Tajikistan) contained haplotypes with and without the insertion, while all the Chinese populations did not have this insertion.

In the network, Chinese haplotypes were central to all of the native populations (Figure 3.3). Two haplotypes connected the introduced and native branches of the network. One of the haplotypes was shared between Syria and Kenya, and the other was detected in Tajikistan, Iran, South Africa,

and Mexico. Most of the haplotypes from the invasive clade (USA, Argentina and Chile) were tip haplotypes.

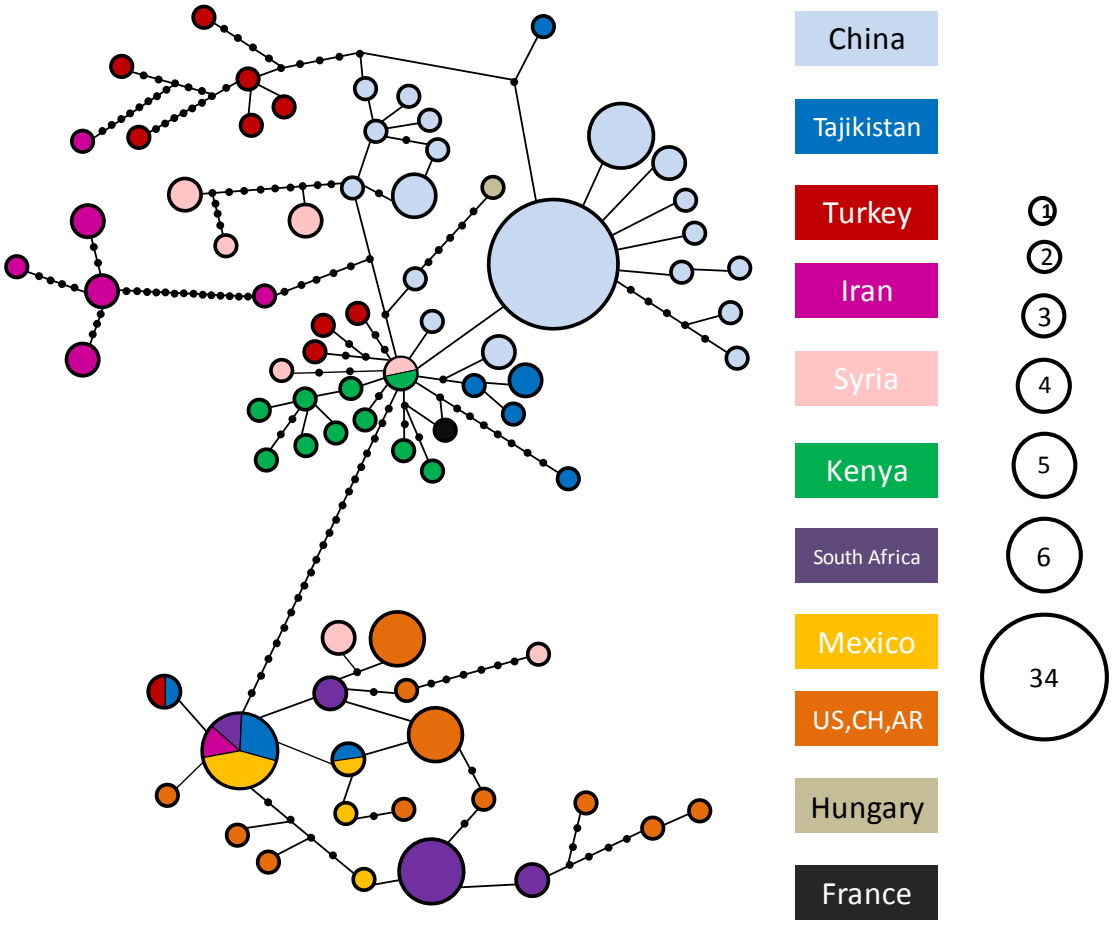



**Figure 3.3** Network of endosymbiont haplotypes found in worldwide RWAs. The color of each circle corresponds to different sites, and the size of the circles represents the number of individuals sharing the haplotype.

*3.4.5 Demographic expansion in native populations*

The TCS network revealed a star-like pattern for the mtDNA data centered on the widely distributed Hap1, providing evidence of rapid demographic expansion (Appendix S3.2). The pairwise mismatch distribution was unimodal, with a strong peak evident at zero, which steeply declined from zero to one base pair. The goodness-of-fit tests were not significant

[*p*(SSD)=0.56 and *p*(Harpending's RI)=0.61], and evidence for highly significant population expansion was detected in the R2 statistic (R2=0.08609, *p*<0.0001), Tajima's *D* (*D*=-2.296, *p*<0.0001), and Fu's *Fs* (*Fs*=-17.662, *p*<0.0001).

A similar star-cluster pattern was discovered in native populations from the endosymbiont DNA network (Figure 3.3). Analysis of the mismatch distribution for the three concatenated endosymbiont genes showed observed values fitting well with the expected simulation, and the goodness-of-fit tests were not significant [*p*(SSD)=0.737 and *p*(Harpending's RI)=0.833]. When separately analyzing only the protein-coding gene of the plasmid, LeuBC, a clear pattern of radiation from one common haplotype to the others was evident indicating rapid demographic expansion (Appendix S3.6). The R2 statistic was also highly significant (R2=0.0845, *p*<0.0001).

*3.4.6 Possible invasion routes from native to introduced ranges*

When the endosymbiont network and microsatellite population data are considered in tandem, putative pathways of RWA invasion from the native range to the introduced areas can be postulated. Firstly, our microsatellite data indicates that Ethiopia represents an extension of this species native range and shares its origin with East Asian RWA. Secondly, the microsatellite data strongly suggests that South Africa and the Americas were colonized by the same RWA clone ($MLG_{AM}$), most likely originating from the Middle East. While the endosymbiont network failed to resolve the order of invasion (South Africa to Americas or vice versa), dates of first detection suggest that South Africa was colonized first, followed by Mexico two years later. Furthermore, although while not conclusive, DIYABC analysis indicated a slightly higher posterior probability for the scenario of stepwise expansion from the Middle East to South Africa and then the Americas (0.5404, range 0.4783-0.6027), than from the Middle East to the Americas and then South Africa (0.4410, range 0.3791-0.5028) or from simultaneous expansion from the Middle East to the Americas and South Africa (0.0000, range 0.0000-0.0000). From this we can conclude that only one invasion event from the native range resulted in the colonization of South Africa and the new world

(Figure 3.4). Thirdly, we postulate that a second invasion took place from Syria to Kenya based on both microsatellite and endosymbiont DNA data.

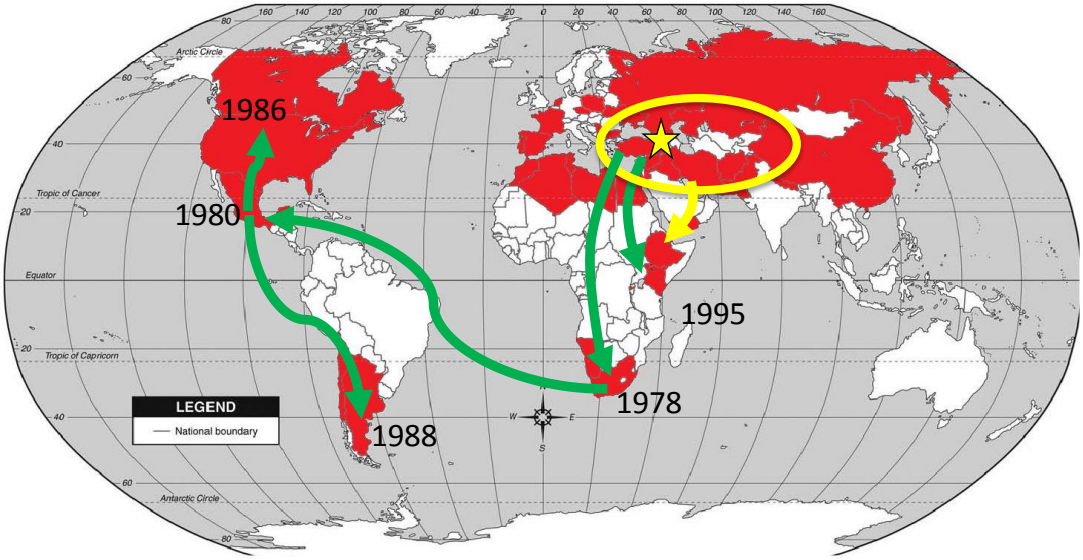



**Figure 3.4** Putative global invasion routes of Russian wheat aphid. The red indicates countries where the RWA has been reported; the yellow circle indicates the native range and arrow to Ethiopia. The yellow star in the Middle East points to the source of invasive clones. The green arrows show the independent invasion pathways from the Middle East to Africa, and single colonization of the Americas. The year of first detection is given for selected countries.

## 3.5 Discussion

Understanding the historical pathways of invasion is critical to understanding the biological and ecological factors that determine whether a species will become invasive. Genetic approaches are often used to understand invasion histories (Dlugosch & Parker 2008), but to be successful both introduced populations and potential source populations must be characterized. Furthermore, the genetics of introduced populations must be characterized at a sufficient depth to be able to discriminate between different invasion events (Rollins *et al.* 2009). Previous studies that have aimed to examine the invasion history of the Russian wheat aphid (*D. noxia*) have failed to identify sufficient genetic variation to discriminate among introduced populations (Shufran *et al.* 2007; Shufran & Payton 2009), or have not characterized native populations sufficiently to identify source populations with any level of confidence (Liu *et al.* 2010). In this study, we have taken advantage of a previously untapped source of genetic variation, that of intracellular symbiotic bacteria, to identify independent historical invasion events of the RWA. By analyzing these invasive lineages together with populations throughout the native range of this species, we have provided valuable insights into the recent global invasion of this damaging pest species.

### *3.5.1 Native population genetic structure*

Our results indicate that populations of RWAs in the Middle East and East Asia possess similar levels of mitochondrial and nuclear genetic diversity to populations from western China (Zhang *et al.* 2012). The low level of mitochondrial variation across this species native range supports the hypothesis that current RWA populations arose from a host switch to cultivated cereals at the debut of wheat domestication (Zhang *et al.* 2012). Novel mitochondrial haplotypes were discovered among the native populations examined, but the level of microsatellite variation in populations elsewhere in its native range was not as high as that observed in some parts of western China. The high genetic diversity in western China samples and the central position of Qapqai in the endosymbiont network points to this area being ancestral. Qapqai is located in the Yili Valley, a west-facing valley

bordered by the Tianshan Mountain range and opening to Kazakhstan. It is likely that Qapqai forms part of the center of origin of RWAs, which has been reported to include Kazakhstan (Kovalev *et al.* 1991; Stary 1999). The microsatellite data also showed significant partitioning of genetic variation among native populations according to geographic region, and the high genetic diversity and high number of unique MLGs within and among native populations is consistent with RWAs exhibiting predominantly holocyclic reproduction within its native range.

Our data suggests that the genetic diversity of RWAs in Syria has reduced markedly in the last decade. Structure analysis indicated that the samples from 2001 and 2011 shared a common ancestry, but in 2001 ten MLGs were identified from a sample of ten individuals compared to seven MLGs from 30 individuals in 2011. Significant heterozygote excess was detected and the number of different MLGs was substantially lower than sample size at this site. This is probably a reflection of local climatic conditions (mild winters) favouring obligate parthenogensis in Syria. Rapid reductions in population genetic diversity are common in aphids (Vorburger 2006; Abbot 2011; Harrison & Mondor 2011), and are probably the result of a combination of regional selection pressures and asexual reproduction (Simon *et al.* 1999). It may be that sexual reproduction is less common in Syria than in the more northern parts of the native range of RWAs.

The level of genetic diversity observed was not associated with the diversity of host plants targeted for collections. All samples from China were collected from wheat, whereas the samples from Turkey were collected from multiple hosts (rye, wheat, barley and oat), and there was no association of MLGs with host plant (Dolatti *et al.* 2005). Furthermore, in South Africa the same genotype was collected from rye, wheat, oats, and *Bromus* grasses.

The strong genetic structuring we observed among Asian populations may have resulted from the recent spread of a host-adapted genotype from its origin in Western China and Kazakhstan. This should not be the case, as the populations would have passed through sufficient generations to achieve a

genetic equilibrium (Lacy 1987). We also observe strong isolation by distance amongst populations in the native range, with strong evidence of gene flow among geographically proximate sites. For example in the Middle East, Turkish and Iranian RWAs formed one population cluster, and in China all northern sites clustered as a single population (Zhang *et al.* 2012).

The results of this study suggest that the native range of RWAs may include Ethiopia, which contradicts current thinking that the RWA was introduced into Ethiopia in 1972 (Zhang *et al.* 1999b). The high allelic richness and gene diversity of microsatellites in Ethiopia indicates a more historical association compared with other African countries included in this study (Kenya and South Africa), although the mtDNA data was monomorphic. Interestingly, Ethiopia is the center of origin of barley diversification worldwide (Badr *et al.* 2000), and barley is a favored host plant of RWA (Stary 1999; Zhang *et al.* 1999a). Further resolution may have been gained by examining levels and patterns of endosymbiont gene diversity in Ethiopia, but unfortunately despite exhaustive efforts, none of the three endosymbiont genes could be amplified from Ethiopian samples, precluding further conclusions from being drawn.

*3.5.2 Global invasion history of RWA*

As has been observed in earlier studies (Shufran & Payton 2009; Liu *et al.* 2010), introduced RWA populations were characterized by highly reduced genetic variability and a single mitochondrial haplotype. However, in this study we were able to distinguish two independent global invasion events using endosymbiont and microsatellite data, one invasion into Kenya and a second invasion into South Africa and the Americas. We were also able to identify putative source populations for each invasion event. The independence of these two invasion events is further supported by their separation into two branches of the endosymbiont network.

Despite being geographically close to Ethiopia, Kenyan RWA was not related to Ethiopian RWA but instead the one MLG detected in Kenya most closely resembled genotypes identified from Syria. We postulate that a single accidental introduction of RWA to Kenya occurred and most likely resulted

from human-aided transportation of contaminated material (eg. commercial wheat or products or germplasm breeding exchange).

The second invasion to South Africa and the Americas was also the likely result of a single human-aided accidental introduction. We detected only a limited number of MLGs across all populations from South Africa and the Americas, all of which could be traced back to a single clone that was shared between South Africa and Mexico. Interestingly, the microsatellite structure analysis indicated that several individuals from Turkey were identical to RWAs from South Africa and the Americas (two of the Turkish samples differed by only one repeat at two alleles from $MLG_{AM}$), and we conclude that Turkey or a related Middle Eastern population was the most likely source of this invasion.

Our data cannot distinguish with any confidence the order of colonization from the Middle East to South Africa and the Americas because of the limited clonal diversity found in this study. However, there is some evidence to suggest that South Africa was the bridge from the native range to the new world. First, the date of detection of RWA in South Africa was two years earlier than when they were first detected in Mexico (Bush *et al.* 1989). Secondly, DIYABC scenario testing of our microsatellite data found a slightly higher posterior probability for colonization of South Africa before the Americas than *vice versa*. All South African and Mexican endosymbiont haplotypes were internal in the network while haplotypes from elsewhere in the Americas were located at the tips. This is consistent with a single introduction of RWAs from South Africa to Mexico and the subsequent spread north to the USA and Canada, and south to Chile and Argentina within ten years. RWAs are more or less continuously spread from Mexico to North America, consistent with natural dispersal. However, RWAs are not continuously distributed from Mexico to Chile and Argentina and therefore it is likely that they were accidently introduced by humans from Mexico to South America.

Our conclusion of a single introduction of RWAs into the USA supports some previous work (Shufran & Payton 2009), but contradicts a recent study of RWAs in the USA using AFLP markers (Liu *et al.* 2010). This may be explained by the greater diversity of RWA biotypes examined by Liu *et al.* (2010), but there is no evidence for multiple introductions in the endosymbiont sequences from these same clones (Swanevelder *et al.* 2010).

## 3.6 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. 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).

## 3.7 References

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

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.

Badr A, Muller K, Schafer-Pregl R (2000) On the Origin and Domestication History of Barley (*Hordeum vulgare*). *Molecular Biology and Evolution*, **17**, 499-500.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies.*Molecular Biology and Evolution*, **16,** 37-48.

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.

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.

Burke GR, Moran NA (2011) Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biology and Evolution,* **3**, 195-208.

Bush L, SlosserJE, Worrall WD (1989) Variations in damage to wheat caused by Russian wheat aphid (Homoptera: Aphididae) in Texas. *Journal of Economic Entomology*, **82**, 466-471.

Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base composition. *Molecular Biology and Evolution*, **16**, 1586-1598.

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.

Dlugosch KM, Parker AM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431-449.

Douglas AE (2006) Phloem-sap feeding by animals: Problems and solutions. *Journal of Experimental Botany,* **57**, 747-754.

Dolatti L, Ghareyazie B, Moharramipour S *et al.* (2005) Evidence for regional diversity and host adaptation in Iranian population of the Russian wheat aphid. *Entomologia experimentalis et applicata,* **114**, 171-180.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.

Everett RA (2000) Patterns and pathways of biological invasions. *Trends in Ecology and Evolution,* **15**, 177-178.

Frankham R (2005) Genetics and extinction. *Biological Conservation,* **126**, 131-140.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.

Fu YX (1997) Statistical methods of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices, Version 2.9.3. www.unil.ch/izea/softwares/Fstat.html.

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

Harrison JS, Mondor EB (2011) Evidence for an invasive aphid "superclone": Extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States. *PLoS One*, **6**, e17524.

Jyoti JL, Qureshi JA, Michaud JP *et al.* (2006) Virulence of two Russian wheat aphid to eight wheat cultivars at two temperature. *Crop Science*, **46**, 774-780.

Kikuchi Y (2009) Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments*, **24**, 195-204.

Kiriac I, Gruber F, Poprawski TJ *et al.* (1990) Occurrence of sexual morphs of Russian wheat aphid, *Diuraphis noxia* (Homoptera: Aphididae), in several

locations in the Soviet Union and the Northwestern United States. *Proceedings of the Entomological Society of Washington*, **92**, 544-547.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution,* **16**, 199-204.

Kolbe JJ, Glor RE, Schettino LR *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature,* **431**, 177-181.

Kovalev OV, Poprawski TJ, Stekolshchinov AV *et al.* (1991) *Diuraphis Aizenberg* (Hom., Aphididae): key to apterous viviparous females and a review of Russian language literature on the natural history of *Diuraphis noxia* (Kudjumov, 1913). *Journal of Applied Entomology*, **112**, 425-436.

Lacy RC (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, migration, selection, and population subdivision. *Conservation Biology*, **2**, 143-158.

Levine JM (2000) Species diversity and biological invasions: relating local process to community pattern. *Science*, **288**, 852-853.

Librado P, Rozas J (2009) DnaSP v. 5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* **25**, 1451-1452.

Liebhold AM, Tobin PC (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology,* **53**, 387-408.

Lindholm AK, Breden F, Alexander HJ *et al.* (2005) Invasion success and genetic diversity of introduced populations of guppies *Poecilia reticulata* in Australia. *Molecular Ecology*, **14**, 3671-3682.

Liu X, Marshall JL, Stary P *et al.* (2010) Global Phylogenetics of *Diuraphis noxia* (Hemiptera: Aphididae), an invasive aphid species: evidence for

multiple invasions into North America. *Journal of Economic Entomology,* **103**, 958-965.

Llewellyn KS, Loxdale HD, Harrington R *et al.* (2003) Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Molecular Ecology*, **12**, 21-34.

Mergeay J, Verschuren D, Meester LD (2006) Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. *Proceedings of Royal Society B,* **273**, 2839-2844.

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of USA,* **98**, 5446-5451.

Oliver KM, Moran NA (2009) Defensive symbionts in aphids and other insects. In: White J, Torres M, editors. Defensive Mutualism in Microbial Symbiosis. 1st ed. CRC Press, Taylor & Francis Group. pp. 129–148.

Peccoud J, Figueroa CC, Sliva AX *et al.* (2008) Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Molecular Ecology*, **17**, 4608-4618.

Peccoud J, Olivier A, Plantegenest M *et al.* (2009a) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of USA,* **106**, 7495-7500.

Peccoud J, Simon JC, McLaughlin HJ *et al.* (2009b) Post-Pleistocene radiation of the pea aphid complex revealed by rapidly evolving endosymbionts. *Proceedings of the National Academy of Sciences of USA,* **106**, 16315-16320.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **1555**, 945-959.

Puterka GJ, Burd JD, Burton RL (1992) Biotypic variation in a worldwide collection of Russian wheat aphid (Hemiptera: Aphididae).*Journal of Economic Entomology*, **85**, 1497-1506.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

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

Rollins LA, Woolnough AP, Wilton AN *et al.* (2009) Invasive species can't cover their tracks: using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. *Molecular Ecology*, **18,** 1560-1573.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology and Evolution*, **22**, 454-464.

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

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.

Smith CM, Belay T, Stuffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) populations virulent to the *Dn4* resistance gene. *Journal of Economic Entomology*, **97**, 1112-1117.

Shoemaker DD, DeHeer CJ, Krieger MJB, *et al.* (2006) Population genetics of the invasive fire ant *Solenopsis invicta* (Hymenoptera: Formicidae) in the United States. *Annals of the Entomological Society of America,* **99**, 1213-1233.

Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single introduction into the United States. *Journal of the Kansas Entomological Society,* **80**, 319-326.

Shufran KA, Payton TL (2009) Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology,* **102**, 440-445.

Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651-701.

Simon C, Buckley TR, Frati F *et al.* (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved Polymerase Chain Reaction primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 545-579.

Simon JC, Baumann S, Sunnucks P *et al.* (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, **8**, 531-545.

Stary P (1999) Distribution and ecology of the Russian wheat aphid, *Diuraphis noxia* (Kurdj.), expanded to Central Europe (Hom.: Aphididae). *Journal of Pest Science*, **72**, 25-30.

Sunnucks P, Hales DF (1996) Numerous Transposed Sequences of Mitochondrial Cytochrome Oxidase I-II in Aphids of the Genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510-524.

Swanevelder ZH, Surridge AKJ, Venter E *et al.* (2010) Limited Endosymbiont Variation in Diuraphis noxia (Hemiptera: Aphididae) Biotypes from the United States and South Africa. *Journal of Economic Entomology*, **103**, 887-897.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Ugelvig LV, Drijfhout FP, Kronauer DJC, *et al.* (2008) The introduction history of invasive garden ants in Europe: Integrating genetic, chemical and behavioural approaches. *BMC Biology,* **6**, 11.

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.

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.

Wernegreen JJ, Moran NA. (2000) Decay of mutualistic potential inaphid endosymbionts through silencing of biosynthetic loci: *Buchnera* of *Diuraphis*. *Proceedings of Royal Society B*, **267**, 1423-1431.

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

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

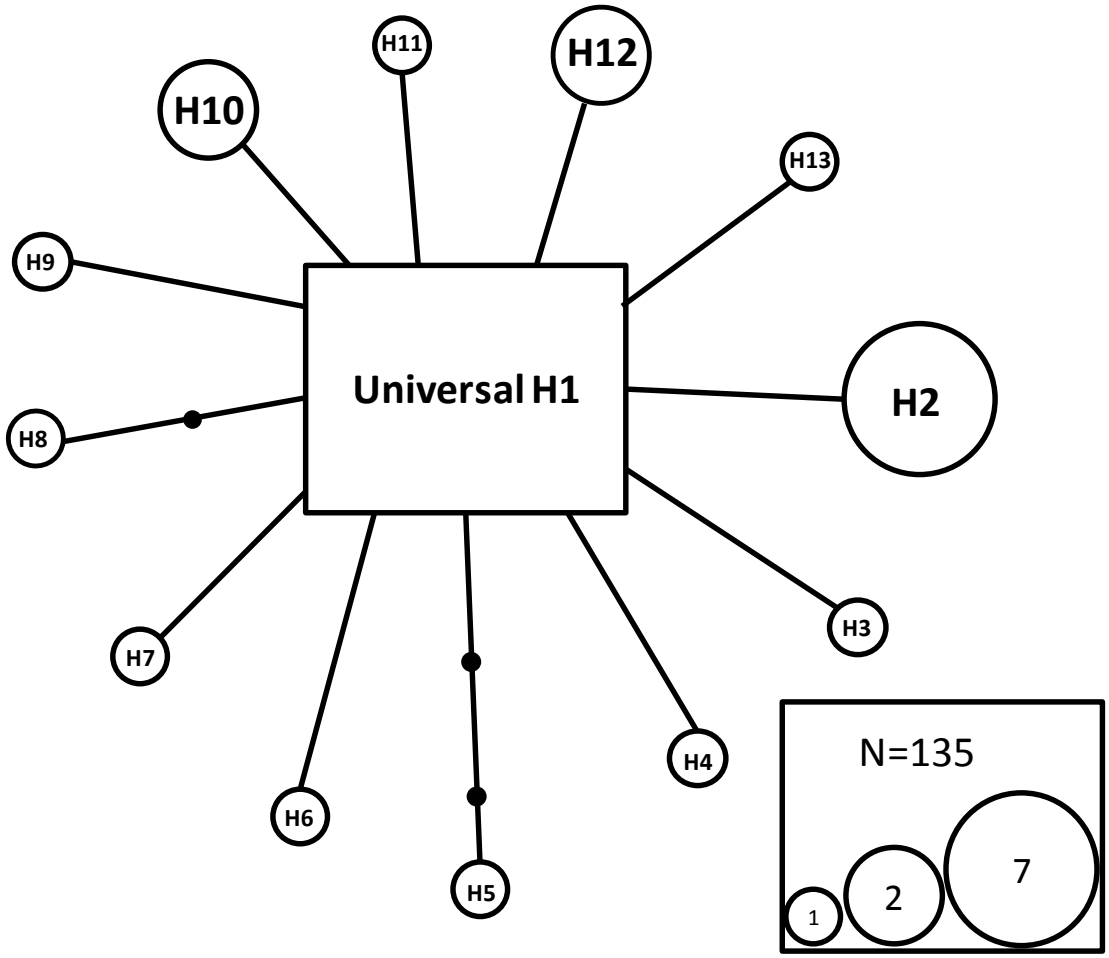
Zhang B, Edwards OR, Kang L, Fuller SJ (2012) Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction? *Molecular Ecology*, **21**, 2130-2144.

# Appendices

**Appendix S3.1**. Sampling information of RWA worldwide.

| Site (Abbr.) | No. | Coordinates | Altitude(m) | Host |
|---|---|---|---|---|
| Turkey (TK) | 49 | 38.57 23.72 | 70-1151 | Durum wheat, rye, oats, barley, volunteer wheat |
| Iran (IR) | 18 | 37.80 47.47 | 1638 | Grasses |
| Syria (SY) | 40 | 36.10 35.60 | 389-660 | Wheat |
| Ethiopia (ET) | 17 | 09.80 38.73 | 3082 | Barley |
| Kenya (KY) | 12 | 05.69 35.31 | 2154-3079 | Wheat |
| South Africa (SA) | 33 | -34.34 19.89 | 1577-1717 | Wheat, Bromus, oats, rye |
| Mexico (MX) | 6 | 19.53 -98.85 | 2273 | Spring wheat, barley |
| USA (US) | 8 | 37.38 -102.5 | 278-1214 | Wheat, barley |
| Chile (CH) | 5 | -32.83 -70.57 | 731-1000 | Barley |
| Argentina (AR) | 4 | -38.37 -60.27 | 83-765 | Bromus, wheat, barley |
| France (FR) | 1 | 43.57 3.89 | 23 | Barley |
| Hungary (HU) | 1 | 47.50 19.08 | 155 | Barley |
| Qapqia (QP) | 53 | 43.59 81.17 | 1178 | Spring wheat |
| Yumin (YM) | 31 | 46.11 82.75 | 1070 | Spring wheat |
| Haba (HB) | 50 | 48.15 86.42 | 601 | Spring wheat |
| Mori (MR) | 42 | 43.77 90.17 | 1234 | Spring wheat |
| Wuqia (WQ) | 52 | 39.71 75.17 | 2151 | Spring wheat |
| Cele (CL) | 52 | 36.29 81.25 | 2300 | Spring wheat |
| Tajikistan (TJ) | 30 | 38.57 68.71 | 803 | Unknown |

**Appendix S3.2** Network of worldwide RWA mtDNA haplotypes.



**Appendix S3.3** *F*st values and significance of pairwise comparisons among worldwide populations. Black values indicate significance after Bonferroni correction at 0.05 level. The bold highlight the *F*st of non-significance.

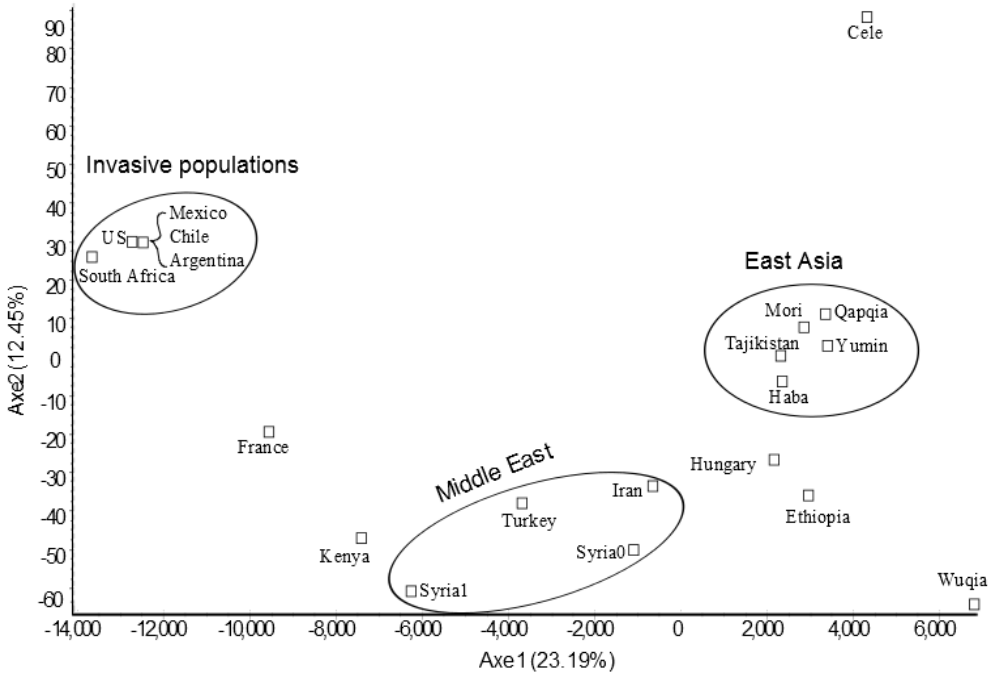
| *F* st | Qapqia | Yumin | Mori | Haba | Cele | Wuqia | Tajikistan | Turkey | Iran | Syria0 | Syria1 | Ethiopia | Kenya | Sth Africa | Mexico | US | Chile | Argentina |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Qapqia** | 0 | | | | | | | | | | | | | | | | | |
| **Yumin** | 0.02681 | 0 | | | | | | | | | | | | | | | | |
| **Mori** | 0.02843 | **0.00824** | 0 | | | | | | | | | | | | | | | |
| **Haba** | 0.03477 | 0.01035 | 0.01722 | 0 | | | | | | | | | | | | | | |
| **Cele** | 0.21294 | 0.23953 | 0.23559 | 0.22124 | 0 | | | | | | | | | | | | | |
| **Wuqia** | 0.09854 | 0.07834 | 0.10664 | 0.07274 | 0.2542 | 0 | | | | | | | | | | | | |
| **Tajikistan** | 0.10158 | 0.10451 | 0.12058 | 0.09353 | 0.29668 | 0.0953 | 0 | | | | | | | | | | | |
| **Turkey** | 0.13419 | 0.11263 | 0.11936 | 0.09118 | 0.33698 | 0.1532 | 0.14393 | 0 | | | | | | | | | | |
| **Iran** | 0.1579 | 0.14457 | 0.17385 | 0.1476 | 0.38974 | 0.1318 | 0.14638 | 0.13034 | 0 | | | | | | | | | |
| **Syria0** | 0.13311 | 0.1329 | 0.16484 | 0.13017 | 0.39488 | 0.1232 | 0.13449 | 0.11171 | **0.0329** | 0 | | | | | | | | |
| **Syria1** | 0.18184 | 0.16364 | 0.16075 | 0.1342 | 0.42657 | 0.2027 | 0.20057 | 0.10439 | 0.2001 | 0.15591 | 0 | | | | | | | |
| **Ethiopia** | 0.06522 | 0.06993 | 0.0776 | 0.05668 | 0.31432 | 0.0662 | 0.09895 | 0.11614 | 0.1437 | 0.12065 | 0.18562 | 0 | | | | | | |
| **Kenya** | **0.06004** | **0.0479** | **0.0439** | **0.00868** | **0.39192** | **0.1061** | **0.05027** | **-0.0474** | **0.1045** | **0.08502** | **0.0469** | **0.02632** | 0 | | | | | |
| **Sth Africa** | 0.06004 | 0.0479 | 0.0439 | 0.00868 | 0.39192 | 0.1061 | 0.05027 | -0.0474 | 0.1045 | 0.08502 | **0.0469** | 0.02632 | **-1** | 0 | | | | |
| **Mexico** | **0.16026** | **0.14737** | **0.14159** | **0.11601** | **0.4291** | **0.192** | **0.14905** | **0.07222** | **0.1969** | **0.18462** | **0.16079** | **0.12811** | **-0.5743** | **-0.5743** | 0 | | | |
| **US** | 0.06004 | 0.0479 | **0.0439** | **0.00868** | 0.39192 | **0.1061** | **0.05027** | **-0.0474** | **0.1045** | **0.08502** | **0.0469** | **0.02632** | **-1** | **-1** | **-0.5743** | 0 | | |
| **Chile** | **0.19259** | **0.19226** | **0.19022** | **0.16456** | **0.46548** | **0.2255** | **0.18078** | **0.12515** | **0.222** | **0.20244** | **0.2113** | **0.16628** | **-0.3565** | **-0.3565** | **-0.1409** | **-0.3565** | 0 | |
| **Argentina** | **0.06335** | **0.02855** | **0.03538** | **-0.0134** | **0.40389** | **0.0769** | **0.07527** | **-0.0428** | **0.0549** | **-0.009** | **-0.3009** | **0.05193** | **-0.3913** | **-0.3913** | **-0.0859** | **-0.3913** | **0.03966** | 0 |

**Appendix S3.4** AMOVA for RWAs worldwide based on microsatellite loci and analyzed in two groups: native and introduced populations.

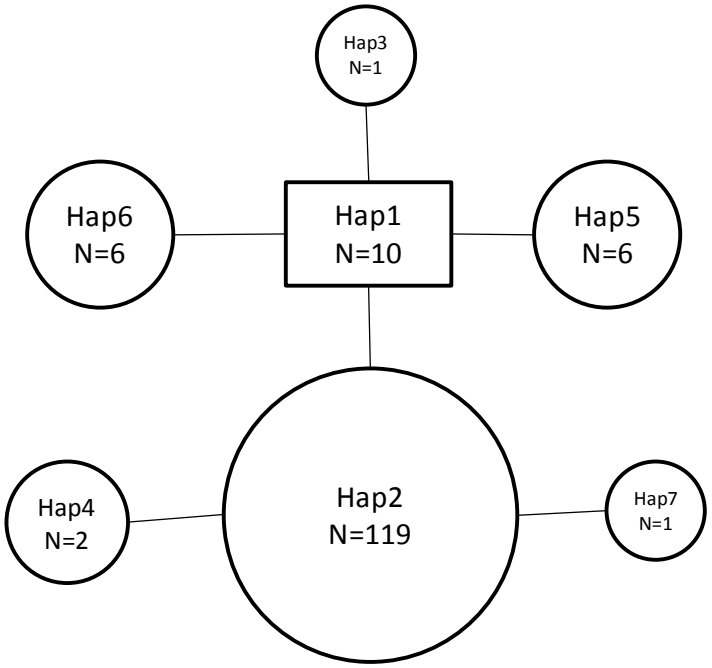
***: p<0.0001

| Source of Variation | Df | Variance components | Percentage of Variation | Fixation Indices |
|---|---|---|---|---|
| Among regions | 6 | 0.5880 | 15.91 | FCT = 0.1591*** |
| Among sites within regions | 11 | 0.1771 | 4.79 | FSC = 0.0570*** |
| Within sites | 986 | 2.9317 | 79.31 | FST = 0.2070*** |

**Appendix S3.5** Factorial correspondence analysis of worldwide RWAs. The circles indicate populations that cluster according to native geography and introduced ranges.



**Appendix S3.6** Network of worldwide RWA endosymbiont LeuBC protein-coding genes.



# Chapter 4

# Salivary gland genes in worldwide RWA

The worldwide pathway of invasion of RWAs was examined using mitochondrial, microsatellite and endosymbiont genetic markers in Chapter 3. Chapter 4 builds on this research through an investigation of salivay gland gene diversity in worldwide populations of RWAs to determine whether there may be a functional genetic basis for RWA invasiveness. Significant selection pressure on salivary gland genes was identified in native RWA populations.

# Statement of Joint Authorship

B. ZHANG, L. KANG, S. J. FULLER, O. R. EDWARDS. Evidence for selection acting on the peroxidase salivary gland gene in Russian wheat aphids (*Diuraphis noxia*)

**Bo Zhang**

Designed and developed experimental protocol. Carried out field and laboratory work, analyzed data, and wrote manuscript.

**Le Kang**

Co-supervisor of this study. Assisted in field sampling and laboratory work and was involved in manuscript discussion.

**Susan J. Fuller**

Principal supervisor of this study. Designed the project. Assisted in the interpretation of data, language editing and contributed to the structure and contents of this manucript.

**Owain R. Edwards**

Co-supervisor of the study. Designed this project. Assisted in samples collection, and data interpretation. Contributed to the structure and manucript editing.

# Evidence for selection acting on the peroxidase salivary gland gene in Russian wheat aphids (*Diuraphis noxia*)

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

[1]Science &Engineering Faculty, 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*, salivary gland gene, peroxidase, wheat domestication, balancing and purifying selection

**Running title**: Selection acting on the peroxidase gene in *Diuraphis noxia*

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

## 4.1 Abstract

Salivary secretions play a critical role in aphid-host interactions and are responsible for damage associated with aphid feeding. In this study we analyze the salivary gland genes in the Russian wheat aphid (*Diuraphis noxia*), one of the world's most invasive agricultural pests. Seventeen salivary gland transcripts in RWA were cloned in two Chinese RWA populations. The variability of transcripts among populations largely depended on single base pair mutations found in only one individual. From the five gene fragments amplified at the DNA level, peroxidase revealed the highest number of SNP sites and gene diversity. Screening of peroxidase gene diversity in worldwide RWA populations revealed seven alleles, one universal allele found in native and introduced populations, and six alleles found only in native populations. Tests for selection indicated that the peroxidase gene is not evolving in a neutral manner and comparison with microsatellite DNA supported this conclusion. Peroxidase in native populations exhibited two types of selection, balancing selection in eastern Asia and purifying selection in the Middle East. Together with the differences in peroxidase gene diversity, the diversifying selection mechanism for this gene may be important in regulating diverse host plant interactions.

## 4.2 Introduction

Aphids (Hemiptera: Aphididae) are important pests of crops worldwide, primarily causing damage to their host plants through feeding. The major food source of an aphid is phloem sap, which is sucked by the aphid through stylet mouthparts that penetrate intercellular plant tissues and tap into the phloem sieve tube (Miles 1999). Aphids inject saliva when feeding and sap-sucking activities are believed to lead to the suppression of host plant defenses, and/or the induction of changes in plant physiology that facilitate aphid feeding and improve nutritional quality (Miles 1999; Will *et al.* 2007; Mutti *et al.* 2008). As aphid feeding involves probing and saliva injection followed by ingestion, knowledge of the salivary gland and its secretary substances is critical to understanding aphid-host plant interactions.

Components of aphid saliva that are released into the phloem are known to have a considerable impact on plant growth and morphology. It has been reported that aphids can decrease growth the rate of roots and shoots, cause leaf discoloration, as well as decrease the yield of crop products (Miles 1999). The concentrations of aphid salivary constituents may be dependent on past feeding and varies over time, even within a single individual (Miles 1999). The watery saliva contains a large number of free amino acids, pectinase, cellulase and possibly other enzymes that may depolymerize complex carbohydrates, hydrolyse phenolic glycosides, and oxidase, amylase and hydrolyse sucrose (Miles 1989; Miles 1999). Host plant susceptibility to attack by aphids is thought to be intrinsically linked to the presence of salivary enzymes that complement that somehow condition the plant for successful aphid feeding. For example, greenbug (*Schizaphis graminum*) biotypes exhibited variable virulence and fitness responses on a variety of cultivars (Kerns *et al.* 1989; Ryan *et al.* 1987) and this was attributed to differences in pectinase activity (Campbell & Dreyer 1985; Dreyer & Campbell 1984). Cellulases of aphid saliva may play a complementary role to pectinases by assisting penetration of the cellulose of plant cell walls (Miles 1989). Though salivary oxidases may help to prevent accumulation of phytochemicals through their action of detoxifying or oxidizing a variety of

defensive phytochemicals (Madhusudhan & Miles 1998; Miles & Peng 1989; Miles & Oertli 1993), they may also be involved in mediating redox reactions that mediate plant cellular defences. Although different aphid species secrete a variety of components with different functions, salivary enzymes are believed to play an important role in facilitating the aphid-plant interaction (Miles 1989; Miles 1999).

Some species of aphids cause toxicosis in the host plant. Localized cell death at the aphid-feeding site, analogous to the hypersensitive response that contributes to many forms of pathogen resistance (Miles 1999), may cause necrosis or leaf discoloration and drooping (Du Toit 1987; Miles 1989). When aphids feed the injection of watery saliva has been shown to cause a wound response in phloem and xylem tissues (Jimoh *et al.* 2011; Saheed *et al.* 2007). Necrosis may be a normal immune response by the plant in defense against exotic substances that evoke a hypersensitive reaction, but necrosis may also result from an imbalance caused by the aphid saliva secretion, which may affect an essential signal pathway for tissue growth or cell differentiation.

A dynamic interplay results from the interaction between the saliva of an aphid and the host plant. Aphids affect host plant growth and morphology via saliva injection, while the plant has a range of defensive responses. The Jones–Dangl zigzag model (Jones & Dangl 2006) has been widely accepted as the most appropriate model for phloem-sucking insects (Thompson & Goggin 2006; Ma & Guttman 2008). In this model, the virulence-associated effector of a pathogen (or salivary elicitor in insects; Kaloshian & Walling 2005) is the primary weapon required for suppression of plant immunity, however plant immune systems can recognize these effectors and alter their response to the invader. The development of host plant cultivars resistant to aphids is a constant challenge, because the zigzag process of plant/pathogen co-evolution results in a dynamic range of virulent and intermediate elicitors. Proteins have been discovered in the body and saliva of aphids that may act as elicitors that affect plant metabolism (Lapitan *et al.* 2007). Therefore, the identification and genetic characterization of these

elicitors is critical to our understanding of how they influence aphid-plant interactions.

In this study we have examined salivary gland genes in one of world's most notorious grain pests, the Russian wheat aphid (RWA, *Diuraphis noxia*). RWAs are an important pest of grain (barley, *Hordeum vulgare*, and wheat, *Triticum aestivum* L.) because of their capacity for rapid population growth and severe damage (Burd *et al.* 2006; Jyoti *et al.* 2006; Morrison & Peairs 1998; Smith *et al.* 2004). Since its introduction into the USA in 1986, management of this pest has focused on the development of resistant wheat cultivars, while chemical control has had less efficient and economic results. In the USA during the period 1987 to 1993, estimated losses were reported to be in excess of $800 million (Morrison & Peairs 1998). The infested plants display white, yellow, or purple longitudinal streaks on the leaves and stems, develop rolled leaves and heads develop a hook-shaped bend (Jyoti *et al.* 2006). Lapitan *et al.* (2007) identified a protein fraction isolated from RWA that could play a key role in determining plant compatibility. When injected into susceptible wheat strains, this protein fraction induced the leaf-rolling symptom typical of RWA feeding. Injecting the protein fraction into RWA-resistant strains did not induce leaf rolling, but increased the expression of defensive peroxidases and catalases compared with the RWA-susceptible strains. While this study suggests that important proteins exist in the body or saliva of an aphid that act as elicitors, their characterization has not been completed.

The availability of the pea aphid (*Acyrthosiphon pisum*) genome has facilitated studies of aphid salivary gland genes (Carolan *et al.* 2009, 2011; Mutti *et al.* 2006, 2008; Ramsey *et al.* 2008). Based on orthology to salivary gland ESTs in the pea aphid, seventeen salivary gland gene transcripts have been examined in two USA biotypes of RWA and were found to be quite divergent and polymorphic (Cui *et al.* 2012). RWA biotypes 1 and 2 cause distinct symptoms and injury rates on different wheat cultivars (Haley *et al.* 2004; Burd *et al.* 2006). It remains unknown however if these variable transcripts affect protein function, plant immunity or host compatibility.

A recent study on the population genetics of RWAs has revealed the global invasion pathway of this species (see Chapter 3). In this study we examine salivay gland gene diversity in worldwide populations of RWAs to determine whether there may be a functional genetic basis for RWA invasiveness. This involved 1) screening seventeen salivary gland RNA transcripts in RWAs to determine the most polymorphic gene at the DNA level; 2) examining variation at this polymorphic locus within and among worldwide populations of RWA; and 3) analyzing the effect of selection by comparing salivary gland and neutral microsatellite genetic diversity and differentiation. This research will indicate whether selective pressures associated with an aphid saliva-host plant interaction have acted on native and introduced populations of RWA.

## 4.3 Material and Methods

*4.3.1 Sample preparation*

Russian wheat aphids were collected from twenty locations worldwide, including eight sites in western China. RWAs were sampled from wheat, rye, and barley crops in addition to native grasses (Table 4.1). At each location, up to 50 aphids were collected from plants located a minimum of 50 meters apart or in different fields. Only one parthenogenetic wingless aphid was collected per plant to minimize the chance of sampling aphids from the same colony. At two locations, extra aphids were sampled (fifteen aphids from each of three colonies) to ensure a sufficient high yield of RNA. Aphids were preserved in 100% ethanol at 4°C until DNA extraction and in RNAlater (Invitrogen) at -80°C until RNA extraction.

*4.3.2 RNA extraction, cDNA synthesis and amplification*

Three colonies from two sample locations (Qitai and Toli) were screened for salivary gland transcript variation. Total RNA was extracted from fifteen aphids following homogenization using a pestle and the RNAeasy Mini Kit (QIAGEN) according to the manufacturer's protocol. The cDNA integrated with aminoallyl-dUTP (Sigma) was prepared from total RNA by reverse transcription. Up to 3ug of total RNA was reverse-transcribed into cDNA following manufacturer’s directions (Invitrogen). The cDNA templates were stored at -20°C until salivary gland gene amplification.

Seventeen salivary gland transcripts were amplified using primers and PCR conditions reported in Cui *et al.* (2012). PCR products were resolved in a 1% agarose gel and then purified using a QIAquick gel extraction kit (Qiagen). The purified products were T-A cloned into pGEMT-Easy vector (Promega) and transfected into DH5α cells (Tiangen Biomedical, China). At least three clones of each transcript for each site were extracted using a TIANprep mini plasmid kit (Tiangen Biomedical, China) and sequenced on an Applied Biosystems 3730x1 DNA Analyzer.

*4.3.3 DNA extraction and amplification*

Total genomic DNA was extracted from single adult aphids using a salting-out method (Sunnucks & Hales 1996). Seventeen genes were amplified in a total volume of 50μl containing 10 nmol primers (Sangong Company, China), 0. 5 U Taq, 1× PCR Buffer, 0.3 mM each dNTP, 2mM $MgCl_2$ (TaKaRaTaq™, Takara Biomedical) and 50ng of aphid DNA. PCR cycling conditions followed the RT-PCR protocol of Cui *et al.* (2012), except that different annealing elongation times were used (Appendix S4.1). PCR products were purified using an ABgene Ultra PCR Clean-Up Kit (Thermo Scientific) kit following electrophoresis on a 1% agarose gel and then sequenced on an Applied Biosystems 3730x1 DNA Analyzer.

RWAs were screened for ten microsatellite loci, including nine newly developed loci (Zhang *et al.* 2012), and one cross-species locus Sa4Σ (Simon *et al.* 1999). 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 $MgCl_2$ (TaKaRaTaq™, Takara Biomedical) and 20ng of aphid DNA. PCR cycling conditions followed Zhang *et al.* (2012). Electrophoresis of the amplification products was conducted on an Applied Biosystems 3730x1 DNA Analyzer, 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.

*4.3.4 Data analysis*

Both DNA and cDNA sequences were aligned and edited using BioEdit v7.0.0 (Hall 1999) and MEGA v4.1 (Tamura *et al.* 2007). Degenerate codes were used when double peaks were encountered. The phase and frequency of alleles were calculated using DnaSP v5 (Librado & Rozas 2009). Genetic diversity estimates were calculated using Arlequin v3.5.1.2 (Excoffier *et al.* 2005) and included: observed heterozygosity (Ho), number of alleles (Na), and gene diversity (Hs). Genetic diversity estimates were compared between native and introduced populations using t-tests (http://www.graphpad.com/quickcalcs/ttest1.cfm).

Pairwise *F*st estimates were calculated from the microsatellite and peroxidase sequences using Arlequin v3.5.1.2 (Excoffier *et al.* 2005). We also calculated Tajima's *D* (Tajima 1989) and Fu and Li's F (Fu& Li 1993) implemented in DnaSP v5 (Librado & Rozas 2009) to infer deviations from neutrality and to detect demographic changes or selection (Fu & Li 1993).

**Table 4.1** RWA sample information.

| Site (Abbr.) | Coordinates | Altitude(m) | Host |
|---|---|---|---|
| Turkey (TK) | 38.57 23.72 | 70-1151 | Durum wheat, rye, oats, barley, volunteer wheat |
| Iran (IR) | 37.80 47.47 | 1638 | Grasses |
| Syria (SY) | 36.10 35.60 | 389-660 | Wheat |
| Ethiopia (ET) | 09.80 38.73 | 3082 | Barley |
| Kenya (KY) | 05.69 35.31 | 2154-3079 | Wheat |
| South Africa (SA) | -34.34 19.89 | 1577-1717 | Wheat, *Bromus*, oats, rye |
| Mexico (MX) | 19.53 -98.85 | 2273 | Spring wheat, barley |
| USA (US) | 37.38 -102.5 | 278-1214 | Wheat, barley |
| Chile (CH) | -32.83 -70.57 | 731-1000 | Barley |
| Argentina (AR) | -38.37 -60.27 | 83-765 | *Bromus*, wheat, barley |
| Hungary (HU) | 47.50 19.08 | 155 | Barley |
| Qapqia (QP) | 43.59 81.17 | 1178 | Spring wheat |
| Yumin (YM) | 46.11 82.75 | 1070 | Spring wheat |
| Haba (HB) | 48.15 86.42 | 601 | Spring wheat |
| Mori (MR) | 43.77 90.17 | 1234 | Spring wheat |
| Wuqia (WQ) | 39.71 75.17 | 2151 | Spring wheat |
| Cele (CL) | 36.29 81.25 | 2300 | Spring wheat |
| Tajikistan (TJ) | 38.57 68.71 | 803 | Unknown |

## 4.4 Results

*4.4.1 cDNA transcript sequences*

We screened seventeen salivary gland gene transcripts in three clones from each of two sample locations and found no variation within and among locations for four transcripts (juvenile hormone binding protein (JHBP), peroxidase, and two non-annotatable proteins). The other thirteen transcripts were polymorphic, however most of the variable sites were detected only once (single base pair mutation). None of the variants were unique to a sample location, i.e. none were found exclusively in all three clones.

*4.4.2 Salivary gland DNA sequences*

DNA could only be amplified from six salivary gland genes using the same primers. It should also be noted that compared with the RNA transcripts, these six genes all contained at least one intron. Of the six genes, two revealed no variation within and among sites (coated-vesicle membrane protein and laccase 1). Glucose dehydrogenase, probable ER retained protein, peroxidase and one non-annotatable protein were polymorphic. The peroxidase gene had the highest number of single nuclear polymorphic sites (SNPs). These twelve SNPs were identified and detected consistently among sample locations, however no SNPs were unique to a location (Table 4.2).

*4.4.3 Peroxidase gene diversity and differentiation*

We screened 134 RWAs from seventeen worldwide locations. The 468 bp fragment contained three exons and two introns, resulting in a total of 104 coded amino acids. Six polymorphic sites were detected in both the protein coding regions and noncoding regions (Figure 4.1). Five sites resulted in synonymous changes, while one non-synonymous change was identified in the middle exon with valine replaced by the less common isoleucyl amino acid. The nucleotide diversity of nonsynonymous and synonymous sites was 0.0005 and 0.0203, respectively.

**Table 4.2** Peroxidase gene SNPs among three individuals from each of two populations; Toli (TL) and Qitai (QT).

| SNP site (bp) | Population_individual | | | | | |
|---|---|---|---|---|---|---|
| | TL_1 | TL_2 | TL_3 | QT_1 | QT_2 | QT_3 |
| 21 | A/G | A/G | A/G | G | A/G | G |
| 36 | T | T | T | T/C | T | T/C |
| 39 | A/G | A/G | A/G | G | A/G | G |
| 54 | A/G | A/G | A/G | G | A/G | G |
| 71 | A/T | A/T | A/T | A | A/T | A |
| 96 | T/C | T/C | T/C | C | T/C | C |
| 121 | A/G | T/G | A/G | G | A/G | G |
| 127 | A/C | A/T | A/C | A/T | A/C | A/C |
| 278 | G | G | G/A | G | G/A | G |
| 307 | T/C | T/C | T/C | C | A/C | C |
| 329 | A/T | A/T | A/T | T | A/T | T |
| 386 | A/T | A/T | A/T | T | A/T | T |

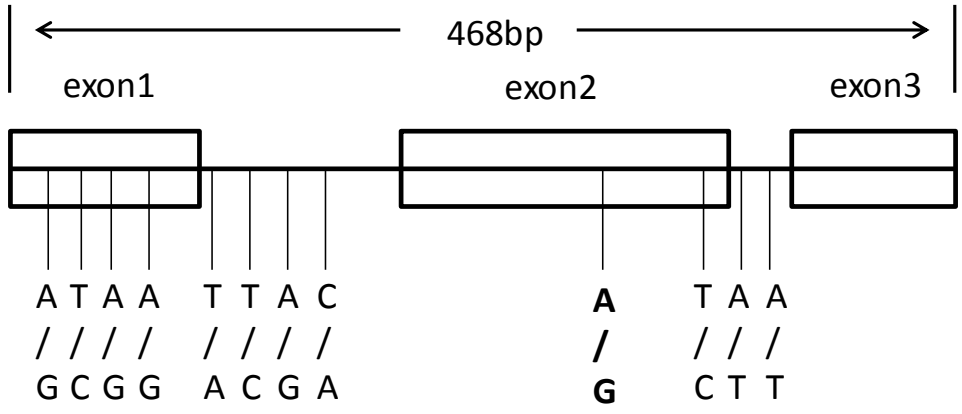



**Figure 4.1** Schematic representation of the peroxidase gene sequence, showing the fragment length, exon numbers, and SNP sites. The nucleotides in bold indicate the SNP that resulted in an amino acid change.

Seven peroxidase alleles were detected and overall allelic diversity was 0.408. One dominant allele had a frequency of 75% and was shared among locations, with the exception of Hungary (Table 4.3). This is likely to represent a sampling artifact as only one RWA was sampled from Hungary. Three alleles were more commonly detected at seven sample locations in East Asia where RWAs are native, while only one allele was dominant in Middle Eastern native populations (Table 4.3). Introduced populations were monomorphic. Peroxidase alleles clustered into two distinct groups, A and B (Table 4.3). The proportion of individuals with alleles representing each cluster varied considerably between East Asia and the Middle East. Cluster A and B were found in nearly equal proportions in East Asian populations (except Cele), while cluster A predominates in Middle Eastern populations.

Nucleotide diversity was high at most locations where RWAs are native. The highest nucleotide diversity was detected in Wuqia in eastern Asia (0. 012), compared with the lowest (0.004) in Turkey from the Middle East (Table 4.4). Throughout the invasive range no nucleotide diversity was found as all locations possessed only one dominant allele. RWAs from eastern Asia (four sites in China, one site in Tajikistan) exhibited similar allele frequencies and the allele containing the non-synonymous mutation was only discovered in China and Tajikistan. One site in China (Cele) was monomorphic for the dominant allele. In Syria and Iran where RWA is native, only the dominant allele was found.

We calculated pairwise estimates of genetic differentiation (*F*st) between native populations that had more than one peroxidase allele. Significant Fst values ($p<0.001$) were only found between Turkey and Yumin (*F*st=0.287), Turkey and Mori (*F*st=0.413), and Turkey and Tajikistan (*F*st=0.402).

*4.4.4 Comparison of microsatellite and peroxidase diversity*

A comparison of microsatellite and peroxidase diversity indices revealed that native populations had significantly higher numbers of alleles (peroxidase $p<0.05$; microsatellite $p<0.001$) and significantly lower homozygosity (peroxidase $p<0.05$; microsatellite $p<0.001$) than invasive populations (Table

4.5). There were no significant differences in microsatellite allele number ($p>0.05$) and homozygosity ($p>0.05$) between East Asian and Middle East populations, but both peroxidase allelic number ($p<0.05$) and homozygosity ($p<0.05$) were significantly different.

*4.4.5 Tests for departure from neutrality*

Departures from neutrality were investigated using Tajima's D, and Fu and Li's F for the seven native sites having variable peroxidase alleles (Table 4.4). The Chinese and Tajikistan sites all displayed significant and positive Tajima's D values indicating balancing selection favoring multiple alleles in far eastern Asian RWA. In Turkey a negative, but not significant, Tajima's D value indicated positive selection.

Finally, excess homozygosity may be indicative of selection. Limited peroxidase diversity was found in Iran and Syria, and a similar pattern was also identified in Turkey (Table 4.5). The Turkish samples also exhibited higher homozygosity than the eastern Asian sites. These results support our conclusion of purifying selection acting in native populations from the Middle East.

To determine whether specific peroxidase alleles may play a functional role in driving the departure from neutrality, we correlated pairwise Fst values calculated from the microsatellite and peroxidase data. No significant correlation was detected ($R= 0.37$, $N=21$, $p>0.05$).

**Table 4.3** Peroxidase allele frequencies for each location among native and invasive ranges. The grey highlights the most common allele.

| Cluster | Alleles | NATIVE-Eastern Asia | | | | | | | NATIVE-Middle East | | | | INVASIVE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | WQ | CL | ML | HB | YM | QP | TJ | TK | IR | SY | HU | KY | SA | Ams |
| | | N=10 | N=10 | N=8 | N=10 | N=10 | N=10 | N=10 | N=10 | N=10 | N=10 | N=1 | N=10 | N=10 | N=23 |
| A | ATAATTACGTAA | 0.45 | 1.00 | 0.38 | 0.50 | 0.55 | 0.55 | 0.35 | 0.90 | 1.00 | 1.00 | 0 | 1.00 | 1.00 | 1.00 |
| | ATAATTAAGTAA | 0.05 | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | GTGGACGAGCTT | 0.25 | 0 | 0.38 | 0.35 | 0.25 | 0.35 | 0.25 | 0.10 | 0 | 0 | 1.00 | 0 | 0 | 0 |
| | GTGGACGAACTT | 0.25 | 0 | 0.00 | 0.05 | 0.05 | 0.05 | 0.35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | GCGGACGCGCTT | 0 | 0 | 0.24 | 0.05 | 0.15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | GTGGACGAATAA | 0 | 0 | 0 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | GTGGACAAGCAA | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

WQ, Wuqia; CL, Cele; ML, Mori; HB, Haba; YM, Yumin; QP, Qapqia; TJ, Tajikistan; TK, Turkey; IR, Iran; SY, Syria; HU, Hungary; KY, Kenya; SA, South Africa; Ams, all American sites including US, Chile, Mexico, and Argentina. N, sample size.

**Table 4.4** Statistical tests for departure from neutral expectations.

| | Wuqia | Mori | Haba | Yumin | Qapqia | Tajikistan | Turkey |
|---|---|---|---|---|---|---|---|
| **S** | 11 | 11 | 12 | 12 | 11 | 11 | 10 |
| **k** | 5.653 | 5.400 | 5.532 | 5.500 | 5.263 | 5.505 | 1.895 |
| **Hd** | 0.705 | 0.700 | 0.653 | 0.642 | 0.600 | 0.726 | 0.189 |
| **Pi** | 0.0121 | 0.0115 | 0.0118 | 0.0118 | 0.0113 | 0.0118 | 0.00405 |
| **Tajima's D** | 2.919** | 2.376* | 2.281* | 2.247* | 2.474* | 2.751** | -1.146 |
| **Fu & Li's F** | 2.161** | 1.953* | 1.631* | 1.620* | 1.649* | 2.105** | 0.783 |

S, segregating sites; k, average number of nucleotide differences; Pi, nucleotide diversity; Hd, haplotype diversity.

**Table 4.5** Comparison between peroxidase and microsatellite variation among native and invasive populations.

| Area | Population | Peroxidase | | Microsatellite | |
|---|---|---|---|---|---|
| | | Na | Ho | Na | Ho |
| **Native** | Wuqia | 3 | 0 | 5.60 | 0.64 |
| | Cele | 1 | 1 | 3.33 | 0.80 |
| | Mori | 3 | 0.13 | 5.00 | 0.65 |
| | Haba | 5 | 0.30 | 6.80 | 0.64 |
| | Yumin | 4 | 0.30 | 6.50 | 0.70 |
| | Qapqia | 4 | 0.30 | 5.50 | 0.68 |
| | Tajikistan | 4 | 0.10 | 4.90 | 0.76 |
| | **Ave (eastern Asia)** | **3.43** | **0.30** | **5.38** | **0.70** |
| | Turkey | 2 | 0.8 | 5.67 | 0.77 |
| | Iran | 1 | 1 | 3.44 | 0.86 |
| | Syria | 1 | 1 | 2.25 | 1 |
| | **Ave (Middle East)** | **1.33** | **0.93** | **3.79** | **0.88** |
| **Introduced** | Kenya | 1 | 1 | 2 | 1 |
| | South Africa | 1 | 1 | 2.38 | 1 |
| | Mexico | 1 | 1 | 2 | 1 |
| | US | 1 | 1 | 2.13 | 1 |
| | Chile | 1 | 1 | 2 | 1 |
| | Argentina | 1 | 1 | 2 | 1 |
| | **Ave** | **1** | **1** | **2.09** | **1** |

Na, mean number of alleles; Ho, observed homozygosity.

## 4.5 Discussion

We sequenced seventeen RWA genes with strong homology to pea aphid salivary gland ESTs and found genetic variation at both the DNA and RNA level. Screening of worldwide RWA populations for the most variable salivary gland gene, peroxidase, indicated that this gene is under selection.

### *4.5.1 Salivary gland transcript variation*

Previous studies on RWA saliva have focused on characterizing the molecular basis of impact on resistant host cultivars. Salivary gland transcripts of RWA biotypes 1 and 2 have been compared and revealed high variation within and between biotypes (Cui *et al.* 2012). As the polymorphism of these transcripts was representative of allelic variation, not duplicated genes, Cui *et al.* (2012) concluded that the salivary gland transcripts were responding to positive selection pressure. We also found several salivary gland transcripts to be variable between the two Chinese sites examined and, like Cui *et al.* (2012), most of these were single base pair mutations found only in one individual. The rapid appearance and disappearance of low-abundance transcripts suggests that salivary gland genes should not be used as a biotype marker because of the inconsistency of low abundance transcripts. However, polymorphism in the mRNA transcripts of protein coding salivary gland genes could possibly provide an indication of whether a functional improvement takes place as a result of amino acid changes.

In this study we successfully amplified seventeen salivary gland transcripts, however using the same primers we could only amplify DNA for six genes. We found many introns scattered across these genes and we hypothesize that they may be a ubiquitous feature of salivary gland genes. The frequent shift in position of the intron and exon probably caused the failure of DNA amplification by the exon-priming intron-crossing primers. We also found more nucleotide variation at the conjunction part of the exon and intron. Half of the SNPs in peroxidase were detected within 15 bp of a conjunction. These SNPs may act as a signal mechanism for alternative splicing.

*4.5.2 Peroxidase gene diversity*

Analysis of peroxidase gene diversity and differentiation among worldwide RWA populations revealed that native populations had higher genetic diversity than populations sampled from the introduced range. Our study showed that the majority of native populations had more than one peroxidase allele (excluding Iran, Syria and Cele (from China)), while all locations where RWA is an introduced species were monomorphic.

If peroxidase is evolving according to neutral expectations it should exhibit a similar pattern of diversity and differentiation as the neutral microsatellite markers. However, when we compared peroxidase and microsatellite diversity, they exhibited different trends among populations. Native populations exhibited equivalent levels of microsatellite diversity, while peroxidase gene diversity was significantly different. The only exception in the native range was Cele (from China), which showed low genetic diversity and high homozygosity for both datasets. Previous research (see chapter 2) has identified that Cele presents a genetic signature of either being a recently founded population or one that has undergone a bottleneck. Low genetic diversity for both microsatellite and peroxidase markers for all introduced populations is indicative of these populations having experienced a founder effect and supports previous research (see Chapter 2). However it is possible that the dominant allele found in all populations provides an allele-dependent advantage that has aided the invasion process.

*4.5.3 Natural selection in the native range of RWAs*

Given the lack of peroxidase gene variation in introduced populations, the impact of natural selection on the peroxidase gene could only be examined in RWA from native populations. We detected two different signatures of selection in native populations of RWA; balancing selection in East Asia and purifying selection in the Middle East. Balancing selection has lead to elevated genetic diversity and maintenance of polymorphism linked to environmental adaptation in East Asia, while purifying selection in the Middle East has resulted in the predominance of one allele that is probably best adapted to local conditions. Peroxidase may function in oxidation and

detoxification of host defensive phytochemicals (Madhusudhan & Miles 1998; Miles & Peng 1989), but may also be involved in regulating the redox response associated with plant cellular defences. In plant pathogen effectors studies, the peroxidase is involved in reactive oxygen species (e.g., $H_2O_2$) generation, both locally and systemically. It plays a significant role to activate cell death and resistant gene induction during the defense response to pathogen invasion (Bindschedler *et al.* 2006; Choi *et al.* 2007). Even though we have not determined physiologically a peroxidase interaction with wheat, our finding that peroxidase gene diversity in native East Asian populations is maintained by balancing selection suggests that this gene could play an important functional role in mediating host plant interactions.

The fundamental connection between an aphid and its host plant lies in the interaction between the injection of aphid saliva and the plant response. Peroxidase activity and the response of RWAs to resistant/susceptible host plants were evaluated by Ni *et al.* (2000). RWA feeding elicited a nine-fold increase in peroxidase specific activity on susceptible “Morex” barley and a three-fold increase on resistant “Halt” wheat when compared with control plants. This finding suggests that RWA feeding probably resulted in oxidative stress in host plants. However, Cooper *et al.* (2010) did not detect peroxidase in the saliva of RWAs and a more recent proteomic analysis of secreted saliva from RWA biotypes also failed to identify peroxidase (Nicholson *et al.* 2012). These findings may indicate that potential phytotoxins, related to RWA biotype differentiation and virulence, were absent in the cultivars tested. It is possible that the peroxidase test failed to detect peroxidase in RWA because of the specific substrate used. For example, in the pea aphid glutathione peroxidase no longer uses glutathione as a substrate, but is instead active as a lipid peroxidase (Owain Edwards unpublished data).

In conclusion, we examined microsatellite and peroxidase genetic diversity in worldwide RWA populations and found that the peroxidase gene is under selection and may be linked to the successful establishment of RWAs on host plants.

## 4.6 Reference

Bindschedler LV, Dewdney J, Blee KA *et al.* (2006) Peroxidase-dependent apoplastic oxidative burst in Arabidopsis required for pathogen resistance. *Plant Journal*, **47**, 851-863.

Burd JD (2002) Physiological modification of the host feeding site by cereal aphids (Homoptera: Aphididae). *Journal of Economic Entomology*, **95**, 463-468.

Campbell BC, Dreyer DL (1985) Host-plant resistance of sorghum - differential hydrolysis of sorghum pectic substances by polysaccharases of greenbug biotypes (*Schizaphis graminum*, Homoptera, Aphididae). *Archives of Insect Biochemistry and Physiology*, **2**, 203-215.

Carolan JC, Fitzroy CIJ, Ashton PD *et al.* (2009) The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterised by mass spectrometry. *Proteomics*, **9**, 2457-2467.

Choi HW, Kim YJ, Lee SC *et al.* (2007) Hydrogen peroxide generation by the pepper extracellular peroxidase CaPO2 activates local and systemic cell death and defense response to bacterial pathogens. *Plant Physiology*, **145**, 890-904.

Cooper WR, Dillwith JW, Puterka GJ (2010) Salivary Proteins of Russian Wheat Aphid (Hemiptera: Aphididae). *Environmental Entomology*, **39**, 223-231.

Cui F, Smith CM, Reese J *et al.* (2012) Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. *Insect Science*, DOI: 10.1111/j.1744-7917.2011.01487.x

Dreyer DL, Campbell BC (1984) Association of the degree of methylation of intercellular pectin with plant-resistance to aphids and with induction of aphid biotypes. *Experientia*, **40**, 224-226.

Dutoit F (1987) Resistance in wheat (*Triticum Aestivum*) to *Diuraphis noxia* (Hemiptera, Aphididae). *Cereal Research Communications*, **15**, 175-179.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47-50.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.

Fu YX, Li WH (1993) Statistical Tests of Neutrality of Mutations. *Genetics*, **133**, 693-709.

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

Jimoh MA, Botha CEJ, Edwards OR (2011) Russian wheat aphid biotypes RWASA2 causes more vascular disruption than RWASA1 on resistant barley lines. *South African Journal of Botany*, **77**, 755-766.

Jones JDG, Dangl JL (2006) The plant immune system. *Nature*, **444**, 323-329.

Jyoti JL, Qureshi JA, Michaud JP, Martin TJ (2006) Virulence of two Russian wheat aphid biotypes to eight wheat cultivars at two temperatures. *Crop Science*, **46**, 774-780.

Kaloshian I, Walling LL (2005) Hemipterans as plant pathogens. *Annual Review of Phytopathology*, **43**, 491-521.

Kerns DL, Puterka GJ, Peters DC (1989) Intrinsic rate of increase for greenbug (Homoptera, Aphididae) biotypes E, F, G, and H on small grain and sorghum varieties. *Environmental Entomology*, **18**, 1074-1078.

Lapitan NLV, Li YC, Peng JH, Botha AM (2007) Fractionated extracts of Russian wheat aphid eliciting defense responses in wheat. *Journal of Economic Entomology*, **100**, 990-999.

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

Ma WB, Guttman DS (2008) Evolution of prokaryotic and eukaryotic virulence effectors. *Current Opinion in Plant Biology*, **11**, 412-419.

Madhusudhan VV, Miles PW (1998) Mobility of salivary components as a possible reason for differences in the responses of alfalfa to the spotted alfalfa aphid and pea aphid. *Entomologia Experimentalis et Applicata*, **86**, 25-39.

Miles PW (1989) Specific responses and damage caused by Aphidoidea. In Aphids: their Biology, Natural Enemies and Control. (ed. A. K. Minks and P. Harrewijn), 2C: 23-47

Miles PW (1999) Aphid saliva. *Biological Reviews of the Cambridge Philosophical Society*, **74**, 41-85.

Miles PW, Oertli JJ (1993) The significance of antioxidants in the aphid-plant interaction - the Redox hypothesis. *Entomologia Experimentalis et Applicata*, **67**, 275-283.

Miles PW, Peng Z (1989) Studies on the salivary physiology of plant bugs - detoxification of phytochemicals by the salivary peroxidase of aphids. *Journal of Insect Physiology*, **35**, 865-872.

Morrison WP, Peairs FB (1998) Response model concept and economic impact. In: A response model for an introduced pest – the Russian wheat aphid. Ed. by Quisenberry SS, Peairs FB, Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD, 1-11.

Mutti NS, Louis J, Pappan LK *et al.* (2008) A protein from the salivary glands of the pea aphid, Acyrthosiphon pisum, is essential in feeding on a host plant. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9965-9969.

Mutti NS, Park Y, Reese JC, Reeck GR (2006) RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. *Journal of Insect Science*, **6**, 1-7.

Ni XZ, Quisenberry SS, Pornkulwat S *et al.* (2000) Hydrolase and oxido-reductase activities in *Diuraphis noxia* and *Rhopalosiphum padi* (Hemiptera : Aphididae). *Annals of the Entomological Society of America*, **93**, 595-601.

Nicholson SJ, Hartson SD, Puterka GJ (2012) Proteomic analysis of secreted saliva from Russian wheat aphid (*Diuraphis noxia* Kurd.) biotypes that differ in virulence to wheat. *Journal of Proteomics*, **75**, 2252-2268.

Ramsey JS, Wilson ACC, de Vos M *et al.* (2007) Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics*, **8**, 432.

Ryan JD, Dorschner KW, Girma M *et al.* (1987) Feeding-behavior, fecundity, and honeydew production of 2 biotypes of greenbug (Homoptera, Aphididae) on resistant and susceptible wheat. *Environmental Entomology*, **16**, 757-763.

Saheed SA, Botha CEJ, Liu L, Jonsson L (2007) Comparison of structural damage caused by Russian wheat aphid (*Diuraphis noxia* Mordvilko) and Bird cherry-oat aphid (*Rhopalosiphum padi* L.) in a susceptible barley cultivar,

*Hordeum vulgare* L. cv Clipper. *Physiologia Plantarum,* **129**, 429-435.

Simon JC, Baumann S, Sunnucks P *et al.* (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, **8**, 531-545.

Smith CM, Belay T, Stauffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera : Aphididae) populations virulent to the Dn4 resistance gene. *Journal of Economic Entomology*, **97**, 1112-1117.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510-524.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Thompson GA, Goggin FL (2006) Transcriptomics and functional genomics of plant defense induction by phloem-feeding insects. *Journal of Experimental Botany*, **57**, 755-766.

Will T, Tjallingii WF, Thonnessen A, van Bel AJE (2007) Molecular sabotage of plant defense by aphid saliva. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 10536-10541.

Zhang B, Edwards OR, Kang L, Fuller SJ (2012) Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction? *Molecular Ecology*, **21**, 2130-2144.

## Appendices

**Appendix S4.1**. Amplification information of 17 salivary gland genes in RWA.

| Genes | Abbr. | Ta (℃) | cDNA | | DNA | |
|---|---|---|---|---|---|---|
| | | | Length (bp) | Te (s) | Length (bp) | Te (s) |
| Coated vesicle membrane protein | CVMP | 55 | 465 | 30 | 850 | 60 |
| AphidB1_C07_t7_050 | ABC | 58 | 204 | 20 | - | - |
| ID0AAH13AH01ZM2 | ID0 | 58 | 159 | 20 | 238 | 30 |
| Peroxidase | PA | 58 | 384 | 20 | 468 | 30 |
| Sucrase | SA | 58 | 874 | 60 | - | - |
| Dipeptidyl carboxypeptidase | DCPA | 60 | 965 | 60 | - | - |
| Cathepsin B | CB | 58 | 719 | 60 | - | - |
| Cathepsin L | CL | 56 | 433 | 30 | - | - |
| Endoprotease FURIN | EF | 58 | 273 | 20 | - | - |
| Glucose dehydrogenase | GDA | 60 | 850 | 60 | 790 | 60 |
| JHBP | JHBP | 58 | 228 | 20 | - | - |
| Trehalase | TRA | 60 | 456 | 30 | - | - |
| Peptidase M1 | PAM1 | 58 | 410 | 30 | - | - |
| Probable ER retained protein | PERP | 56 | 401 | 30 | 540 | 40 |
| Emp24 | EMP24 | 60 | 365 | 20 | - | - |
| C002 | C002 | 53 | 765 | 60 | - | - |
| Laccase 1 | LA1 | 55 | 367 | 20 | 410 | 30 |

Ta, annealing temperature; Te, elongation time.

# Chapter 5

# Mitochondrial genome of RWA

In this chapter the RWA mitochondrial genome is compared with other aphid mitochondrial genomes to investigate the possible causes of low mitochondrial DNA polymorphism that renders mitochondrial DNA markers relatively ineffective for aphid population genetics studies.

# Statement of Joint Authorship

B. ZHANG, C. MA, O.R. EDWARDS, S.J. FULLER,L. KANG. Large repetitive sequences outside the control region in aphid mitochondrial genomes.

**Bo Zhang***

Designed and developed the experimental protocol, completed the field and laboratory work, analyzed the data and wrote the manuscript.

**Chuan Ma***

Designed and developed the experimental protocol, analyzed the data and contributed to the structure and contents of the manuscript.

**Owain R. Edwards**

Co-supervisor of the study. Assisted in sourcing mitogenome data for other aphid species. Contributed to data interpretation and the manucript structure.

**Susan J. Fuller**

Principal supervisor of this study. Contributed to the manuscript revision.

**Le Kang**

Co-supervisor of this study. Designed the project, coordinated the field sampling and laboratory work and involved with manuscript revision.

*These authors contributed equally to the design, development, preparation and writing of the material presented in this paper.

# Large repetitive sequences outside the control region in aphid mitochondrial genomes

Bo ZHANG[1, 2, 3*], Chuan MA[1*], Owain EDWARDS[3, 4], Susan FULLER[2, 4], Le KANG[1※]

[1]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; [2]Science & Engineering Faculty, Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia; [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*, mitochondrial genome, tRNA-like pseudogenes, Sternorrhyncha, Aphidoidea, repeat region

**Running title**: Mitochondrial genome of *Diuraphis noxia*

*These authors contributed equally to this work.

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

## 5.1 Abstract

In this study, we sequenced and annotated the complete mitochondrial genome (mitogenome) of the Russian wheat aphid, *Diuraphis noxia* (Hemiptera: Aphidoidea). The 15,784 bp circular genome includes all 13 protein-coding genes, 2 ribosomal RNA genes, 23 transfer RNA (tRNA) genes, the control region, and the repeat region. All the genes were arranged in the same order as that of the putative ancestor of insects, except that we did find one extra transfer RNA gene (*trnM*). The mitogenome was characterised by having a high A+T content (84.76%) and a strong GC skew (-0.26), which is consistent with the four other complete aphid mitogenomes examined. The 22 tRNAs in the *D. noxia* mitogenome have a typical clover-leaf structure, except *trnS(AGN)*, which is consistent with that reported in other aphid species. We also discovered tRNA-like pseudogenes in the control region of other Aphidoidea species. The repeat region and the control region are highly divergent within the superfamily. We found that the repeat region consistently varied by two tandem repeats among species in the Aphidoidea, however the length and copy number were unique to each species. No evidence of any consensus sequences was found to predict the origin of the repeat region. Finally, we conclude that mtDNA markers from mitogenomes that possess high A+T% and low GC skew are not useful for studies of intraspecific divergence. In addition, the evolution of the mitogenome should beconsidered in conjunction with endosymbiont evolution as both genomes are associated with exogenous bacteria which may affect host evolution.

## 5.2 Introduction

Mitochondria are present in most eukaryotic cells, varying in number from hundreds to thousands (Simon *et al.* 2006). Energy production in the form of ATP is the major mitochondrial function, but mitochondria are also involved in a range of other chemical processes, such as transference signaling, cellular differentiation, cell death, as well as the control of the cell cycle and cell growth (McBride *et al.* 2006). The origin of mitochondria lies in the endosymbiotic association of oxidative bacteria and glycolytic proto-eukaryotic cells (Gray *et al.* 1999).

Compared with nuclear genomes, animal mitogenomes are characterized by several distinct features. Firstly, because the mitogenome experiences a high substitution rate, it has become an attractive source of DNA polymorphism to study population genetics and evolution (Ingman & Gyllensten 2006). Secondly, the lack of mitogenome recombination allows the tracing of a direct genetic line, enabling the study of relationships at species or higher taxonomic levels. Finally, the relatively stable gene content and multiple mitochondrial DNA (mtDNA) copies per cell facilitate mtDNA amplification. Thus, mitochondrial sequences have been the main tool in a large number of studies on the population genetics and phylogenetics of insects (Sheffield *et al.* 2008; Negrisolo *et al.* 2011). In addition, insect mitogenomes possess several interesting evolutionary features such as atypical stop codons, base compositional bias, codon usage bias, and gene rearrangement (Negrisolo *et al.* 2011; Sheffield *et al.* 2008; Thao *et al.* 2004). Mitochondrial gene order can provide an independent source of information for mitogenome evolution and phylogenetic inference (Boore 1999; Rokas & Holland 2000). Various mitochondrial gene rearrangements relative to the ancestral gene order have been detected in a variety of species (Boore 1999). Some of these gene arrangements are shared derived characters of an order, e.g., the arrangement of *trnM-trnI-trnQ* is present in all Lepidoptera mitogenomes examined to date. Different rearrangements are also present in Hemiptera mitogenomes, and they are confined only to the suborder Sternorrhyncha (Thao *et al.* 2004). Additional mitochondrial sequencing and comparative

analysis will provide new insights into molecular mechanisms underlying mitogenome evolution and, in particular, gene rearrangement.

The order Hemiptera, the largest and most diverse group of hemimetabolous insects, consists of three suborders: Sternorrhyncha, Auchenorrhyncha, and Heteroptera (Lee *et al.* 2009). Currently, full-length mitogenomes of Hemiptera are known for 46 species and are available in GenBank. Hemiptera mitogenomes have been studied for a variety of purposes, including divergence within suborder (Hua *et al.* 2009), gene rearrangements among families (Thao *et al.* 2004), and phylogenetic relationships within suborders (Li *et al.* 2011). Most of the available Hemiptera mitogenomes are from Heteroptera, while only 9 Sternorrhynchan mitogenomes have been studied (Thao *et al.* 2004). Meanwhile, only two mitogenomes of aphids, *Acyrthosiphon pisum* and *Schizaphis graminum*, have been determined for Sternorrhyncha. These aphid mitogenomes both have a large repeat region located between *trnE* and *trnF*, and both differ in repeat unit sequence and copy number. To our knowledge, this repeat region has never been reported in other insect mitogenomes. There has been no attempt to examine the diversity of this repeat region among aphid mitogenomes or to determine whether this structural feature is common in all aphid species. It is clear that mitogenome sequencing of more aphid species is required to address this interesting question. Particularly, analysis with a comparative mitogenomic and evolutionary perspective will shed light on the origin and diversification of the repeat region.

In this paper, we report the description and annotation of the complete mitochondrial genome of the Russian wheat aphid, *Diuraphis noxia*, and compare with other members of the Aphidoidea and more broadly within the order Hemiptera. We calculated the base composition bias within the order Hemiptera, and in particular, described the specific nucleotide features among species within the suborder Sternorrhyncha. To provide insight into mtDNA evolution, we compared the mitogenomes of representatives of each superfamily within this suborder, including *Bemisiata baci* (Aleyrodoidea), *Pachypsylla venusta* (Psylloidea) and *Diuraphis noxia* (Aphidoidea). Finally,

structural comparisons were made from an analysis of the complete mitogenomes of two species of Aphidoidea, *Acyrthosiphon pisum* and *Schizaphis graminum*. Some analyses were also based on unpublished partial genome data from *Aphis gossypii* and *Daktulosphaira vitifoliae*. A range of shared characteristics and unique features of aphids are presented and explain potentially the low variability and relative insensitivity of mitochondrial markers that typifies many studies of aphid genetics. We also discuss the possible effects of endosymbionts on the evolution of aphid mitogenomes, both of which are exogenous bacterial organisms.

## 5.3 Material and Methods

*5.3.1 Sample origin and DNA extraction*

Parthenogenetic female *D. noxia* were collected from Mori (43.779°N, 90.177°E, 1234 m elevation), in eastern Xinjiang Province, northwestern China. The samples were preserved in 100% ethanol and stored at 4°C until used for DNA extraction. Total genomic DNA was extracted from single adult aphids using the salting-out method (Sunnucks & Hales 1996). The final DNA template elution was 40ng/µl, and stored at -20°C until required.

*5.3.2 PCR amplification and sequencing of mtDNA*

PCR amplification of the *D. noxia* mitogenome was performed using universal primers and primers designed based on the *Acyrthosiphon pisum* (GenBank: NC011594) and *Schizaphis graminum* (GenBank: AY531391) mitogenomes (Simon *et al.* 2006). Fourteen overlapping fragments were amplified in total. The complete primer list is supplied in Appendix S6.1. The final 50µl PCR volume included 80ng of aphid DNA, 2 U ExTaq, 1×PCR Buffer, 0.3mM of each dNTP and 2mM $MgCl_2$ (TaKaRaTaq™, Takara Biomedical, Japan). The PCR cycling parameters were as follows: initial denaturation 3 min at 94°C; 30 cycles of 30 sec at 94°C, 30 sec at the specified annealing temperature (Appendix S6.1), 3min at 72°C; and a final elongation of 10 min at 72°C. The PCR products were visualized by electrophoresis in a 1.5% agarose gel and purified using anABgene Ultra PCR Clean-Up Kit (Thermo Scientific) before being sequenced on a 3730xl DNA Analyser (Applied Biosystems). The fragment containing the repeat region was inserted into pGEM-T Easy vector (Promega, UK) according to manufacturer's directions and multiple clones were sequenced.

*5.3.3 Gene annotation and analysis*

The mtDNA sequences were assembled using Lasergene software (DNAStar, Inc.) and BioEdit (Hall 1999). The complete mtDNA genome was analyzed as follows. Firstly, protein coding genes (PCGs) and their boundaries were investigated using the BLAST program available on the NCBI website. In our gene annotation of *D. noxia*, as well as re-annotations of the other three

aphids, gene boundaries were assigned to avoid gene overlaps, which are not favorable under natural selective pressures (Sheffield *et al.* 2010). Gene overlap was allowed between adjacent protein coding genes (PCGs), but not between PCGs and downstream transfer RNA (tRNA) genes. Specifically, when the full stop codon of a protein-coding gene overlapped with a tRNA gene, an incomplete stop codon was designated. The common termination codons could be generated by the post-transcriptional polyadenylation. Secondly, tRNA genes were identified using the tRNAscan-SE program (Schattner *et al.* 2005), and recognized manually, to find the appropriate anticodon and the typical cloverleaf secondary structure. Finally, the boundaries of the ribosomal RNA (rRNA) genes, *rrnL* and *rrnS*, were determined by comparison with orthologous genes of two other aphid mitogenomes. In addition, BioEdit (Hall 1999) and MEGA5 (Tamura *et al.* 2011) were used to explore nucleotide and protein pairwise divergence among the four aphid species, including the *Daktulosphaira vitifoliae* partial mitogenome (DQ021446) and the unpublished *Aphis gossypii* mitogenome.

*5.3.4 Genomic analysis*

Nucleotide composition was calculated using BioEdit (Hall 1999). ATskew [(A-T)/(A+T)] and GC skew [(G-C)/(G+C)] were used to measure base compositional differences. Codon usage was calculated using the Sequence Manipulation Suite program (http://www.bioinformatics.org/sms2/). The relative synonymous codon usage (RSCU) values were calculated using MEGA5 (Tamura *et al.* 2011). Repeat motifs were identified using the Tandem Repeats Finder program (Benson 1999). Consensus repeat sequences were recognized as a minimum match of 75% among strings. Overall genetic distances of thirteen PCGs were calculated using MEGA5 under the Kimura-2-Parameter model for nucleotide sequences and p-distance for amino acid sequences.

## 5.4 Results

*5.4.1 Genome composition*

The complete mitogenome of *D. noxia* is 15,784bp in size containing 13 PCGs, 23tRNA genes, 2rRNA genes, a control region, and a large repeat region (Figure 5.1, Table 5.1). These genes/regions are arranged in the same order as the most wide spread insect ancestral mitogenome (Boore 1999), except that the *D. noxia* mitogenome possessed one extra tRNA gene and a separate large non-coding repeat region located between *trnE* and *trnF*.

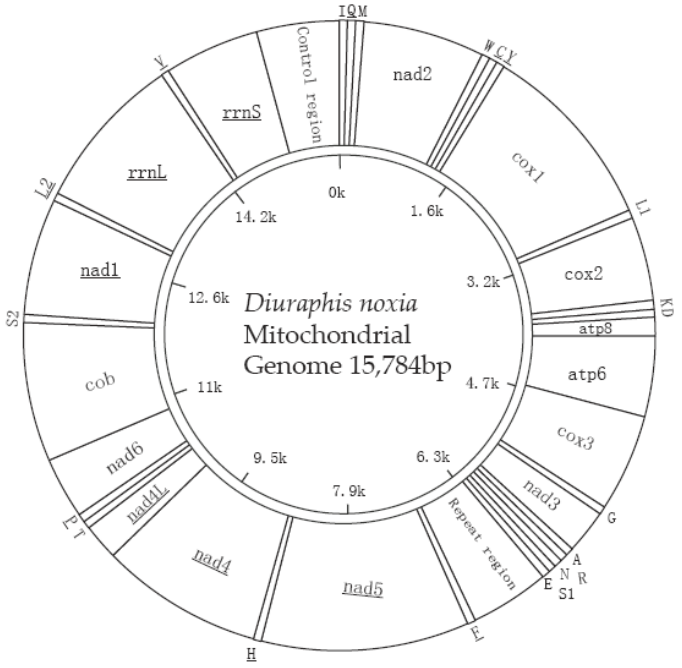



**Figure 5.1** Map of the *D. noxia* mitogenome. Genes are transcribed in a clockwise direction, except for those underlined. The transfer RNAs are labeled according to the IUPAC/IUB single-letter amino acid codes.

**Table 5.1** The organization and annotation of the *D. noxia* mitogenome.

| Feature | Strand | Position | Length (bp) | No. of amino acid | Initiation codon | Stop codon | Anticodon | AT% | intergenic nucleotide |
|---|---|---|---|---|---|---|---|---|---|
| *trnI* | J | 1-64 | 64 | | | | GAT | 78.1 | - |
| *trnQ* | N | 62-127 | 66 | | | | TTG | 83.4 | -3 |
| *trnM* | J | 133-199 | 67 | | | | CAT | 79.1 | 5 |
| *nad2* | J | 200-1175 | 976 | 325 | ATT | T-- | | 88.3 | 0 |
| *trnW* | J | 1176-1241 | 66 | | | | TCA | 87.9 | 0 |
| *trnC* | N | 1234-1298 | 65 | | | | GCA | 86.1 | -8 |
| *trnY* | N | 1302-1367 | 66 | | | | GTA | 81.8 | 3 |
| *cox1* | J | 1369-2899 | 1531 | 510 | ATA | T-- | | 76.9 | 1 |
| *trnL(UUR)* | J | 2900-2967 | 68 | | | | TAA | 80.8 | 0 |
| *cox2* | J | 2971-3642 | 672 | 223 | ATA | TAA | | 81.3 | 3 |
| *trnK* | J | 3645-3717 | 73 | | | | GTC | 75.3 | 2 |
| *trnD* | J | 3718-3779 | 62 | | | | CTT | 85.5 | 0 |
| *atp8* | J | 3789-3938 | 150 | 49 | ATA | TAA | | 92.7 | 9 |
| *atp6* | J | 3925-4572 | 648 | 215 | ATA | TAA | | 83.5 | -14 |
| *cox3* | J | 4572-5356 | 785 | 261 | ATG | TA- | | 83.7 | -1 |
| *trnG* | J | 5357-5420 | 64 | | | | TCC | 90.6 | 0 |
| *nad3* | J | 5421-5774 | 354 | 117 | ATT | TAA | | 87.6 | 0 |
| *trnA* | J | 5775-5838 | 64 | | | | TGC | 89.1 | -3 |
| *trnR* | J | 5838-5901 | 63 | | | | TCG | 85.7 | 0 |
| *trnN* | J | 5902-5967 | 66 | | | | GTT | 83.3 | 0 |
| *trnS (AGN)* | J | 5968-6028 | 61 | | | | GCT | 83.6 | 0 |
| *trnE* | J | 6032-6097 | 66 | | | | TTC | 95.4 | 3 |
| Repeat region | J | 6099-6737 | 639 | | | | | 90.7 | 1 |
| *trnF* | N | 6738-6803 | 66 | | | | GAA | 84.8 | 0 |
| *nad5* | N | 6804-8474 | 1671 | 556 | ATT | TAA | | 86.3 | 0 |
| *trnH* | N | 8525-8588 | 64 | | | | GTG | 89.1 | 0 |
| *nad4* | N | 8589-9897 | 1309 | 436 | ATA | T-- | | 85.7 | 0 |
| *nad4L* | N | 9906-10196 | 291 | 96 | ATA | TAA | | 89.4 | 8 |
| *trnT* | J | 10198-10259 | 62 | | | | TGT | 91.9 | 1 |
| *trnP* | N | 10261-10327 | 67 | | | | TGG | 86.6 | 1 |
| *nad6* | J | 10329-10820 | 492 | 163 | ATT | TAA | | 89.6 | 1 |
| *cob* | J | 10824-11942 | 1119 | 372 | ATG | TAA | | 81.1 | 3 |
| *trnS (UCN)* | J | 11947-12011 | 65 | | | | TGA | 90.8 | 4 |
| *nad1* | N | 12021-12956 | 936 | 311 | ATT | TAA | | 84.4 | 9 |
| *trnL (CUN)* | N | 12957-13021 | 65 | | | | TAG | 86.2 | 0 |
| *rrnL* | N | 13022-14280 | 1259 | | | | | 85.1 | 0 |
| *trnV* | N | 14281-14343 | 63 | | | | TAC | 87.3 | 0 |
| *rrnS* | N | 14355-15120 | 766 | | | | | 84.6 | 11 |
| A+T-rich region | J | 15121-15784 | 664 | | | | | 86.6 | 0 |
| *trnM* | N | 15443-15514 | 72 | | | | TAT | 90.2 | - |

The A+T% and G+C% values as well as the AT and GC skews were calculated for all available complete mitogenomes of Hemiptera species and are presented graphically in Figure 5.2. The AT content of Hemipteran mitogenomes ranged from 68.86% (*Neuroctenus parus*) to 86.34% (*Aleurodicus dugesii*). Notably, the three suborders of Hemiptera exhibited similar nucleotide skews (C skew and A skew) except for six Sternorrhynchan species, which had highly re-arranged gene orders, while the four other Sternorrhynchans clustered as a group.

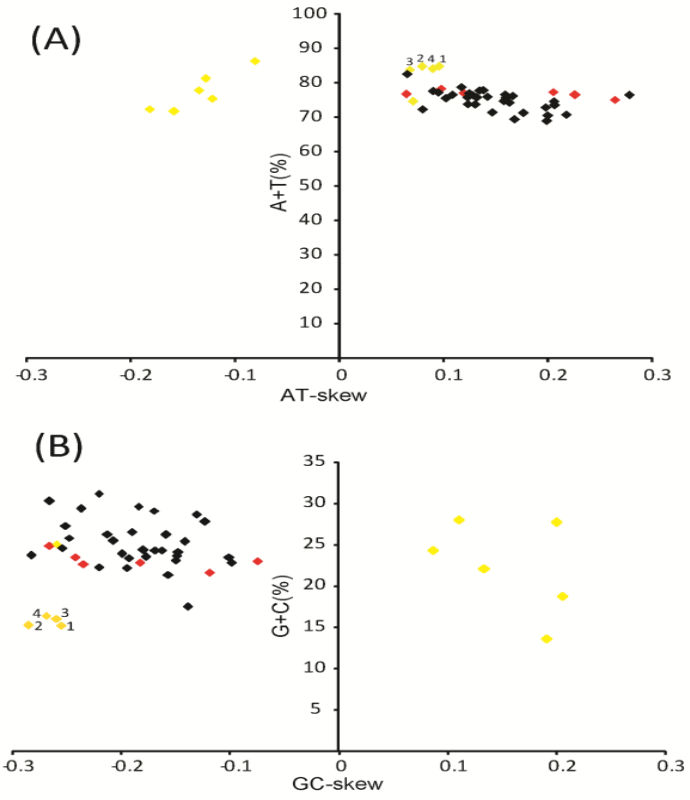



**Figure 5.2** Graphical summary of nucleotide content across 49 Hemiptera mtDNAs. (A) A+T% vs AT skew. (B) G+C% vs GC skew. The yellow: Sternorrhyncha; red: Auchenorrhyncha; black: Heteroptera. 1, 2, 3 and 4 refer to the four aphid mitogenomes, *D. noxia*, *A. pisum*, *S. graminum*, and *A. gossypii,* respectively.

The overall nucleotide composition of *D. noxia* was typically A+T biased (84.76%; the second highest amongst Hemiptera species), and was slightly A skewed (0.08) and strongly C skewed (-0.26) (Figure 5.2, Table 5.2).Similar nucleotide composition patterns were detected in the four aphid species (Figure 5.2, Table 5.2). The majority of second codon positions of PCGs were T, with an average AT skew of -0.40 in the three aphid species examined (Table 5.2). Furthermore, the third codon positions of PCGs were predominantly A+T and accounted for approximately 95% of the total A+T content in Aphidoidea mitogenomes. Finally, the A+T content as well as the A and C bias of the repeat region of the Aphidoidea species examined, greatly exceeded the whole genome indices. This was particularly evident in *D. noxia*, with the repeat region having an A+T content of 90.61%, an AT skew of 0.23 and a GC skew of -0.67. Similarly, the conventional A+T-rich control region in Aphidoidea showed a high A+T% compared with the genomic average, but less than that of the repeat region in *D. noxia*.

**Table 5.2** Nucleotide composition of the *D. noxia* (*D.n*), *A. pisum* (*A.p*) and *S. graminum* (*S.g*) mitogenomes.

| Feature | (A+T)% | | | AT skew | | | GC skew | | |
|---|---|---|---|---|---|---|---|---|---|
| | *D.n* | *A.p* | *S.g* | *D.n* | *A.p* | *S.g* | *D.n* | *A.p* | *S.g* |
| Whole genome | 84.76 | 84.72 | 83.94 | 0.08 | 0.10 | 0.07 | -0.26 | -0.28 | -0.26 |
| Protein-coding genes | 84.05 | 83.63 | 83.21 | -0.15 | -0.15 | -0.16 | -0.03 | -0.05 | -0.04 |
| 1st codon position | 80.99 | 80.11 | 79.88 | 0.00 | 0.02 | 0.01 | 0.15 | 0.14 | 0.14 |
| 2nd codon position | 75.78 | 75.65 | 75.6 | -0.39 | -0.39 | -0.40 | -0.12 | -0.13 | -0.13 |
| 3rd codon position | 95.38 | 95.13 | 94.15 | -0.09 | -0.09 | -0.11 | -0.32 | -0.39 | -0.26 |
| tRNAs | 85.58 | 86.02 | 85.56 | 0.04 | 0.02 | 0.03 | 0.18 | 0.20 | 0.18 |
| rRNAs | 84.94 | 84.6 | 84.89 | -0.08 | -0.08 | -0.06 | 0.33 | 0.36 | 0.34 |
| Repeat region | 90.61 | 88.91 | 84.91 | 0.23 | 0.27 | 0.16 | -0.67 | -0.57 | -0.52 |
| Control region | 86.6 | 89.17 | 86.82 | 0.02 | 0.04 | -0.07 | -0.26 | -0.38 | -0.19 |

### *5.4.2 Protein-coding genes*

The thirteen PCGs of the *D. noxia* mitogenome were distributed on both strands as previously reported in other insects. The gene boundaries and annotation are given in Table 1. Most PCGs were initiated with an ATN start codon in *D. noxia*, and terminated by a conventional TAA stop codon (Table

5.1 and 5.3). However, the genes including *nad2*, *cox1*, *cox3* and *nad4*, ended with a single T or TA residues. Incomplete termination codons were also observed in the other aphid species (Table 5.3).

**Table 5.3**. Comparison of aphid protein-coding genes. *D.n*, *D. noxia*; *A.p*, *A. pisum*; *S.g*, *S. graminum*; *D.v*, *D. vitifoliae*.

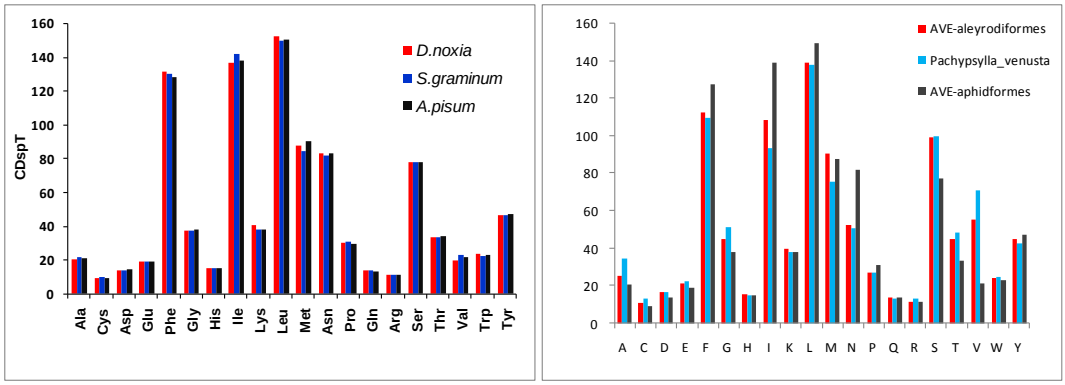
| Genes | Start codon | | | | Stop codon | | | | Protein length | | | | p-distance | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *D.n* | *A.p* | *S.g* | *D.v* | *D.n* | *A.p* | *S.g* | *D.v* | *D.n* | *A.p* | *S.g* | *D.v* | 3 aphids | 4 aphids |
| *nad2* | ATT | ATA | ATA | ATA | T-- | T-- | T-- | T-- | 325 | 325 | 325 | 327 | 0.107 | 0.15 |
| *cox1* | ATA | ATA | ATA | ATA | T-- | T-- | T-- | TAA | 510 | 510 | 510 | 510 | 0.082 | 0.095 |
| *cox2* | ATA | ATA | ATA | ATA | TAA | TAA | TAA | TAA | 223 | 223 | 223 | 223 | 0.073 | 0.105 |
| *atp8* | ATA | ATA | ATA | ATA | TAA | TAA | TAA | TAA | 49 | 49 | 55 | 56 | 0.179 | 0.223 |
| *atp6* | ATA | ATA | ATT | ATA | TAA | TAA | TAA | TAA | 215 | 215 | 217 | 216 | 0.103 | 0.148 |
| *cox3* | ATG | ATG | ATG | ATG | TA- | TA- | TA- | - | 261 | 261 | 261 | - | 0.104 | - |
| *nad3* | ATT | ATG | ATA | - | TAA | TA- | TAA | - | 117 | 117 | 117 | - | 0.102 | - |
| *nad5* | ATT | ATT | ATA | - | TAA | TAA | TAA | - | 556 | 556 | 556 | - | 0.069 | - |
| *nad4* | ATA | ATA | ATA | ATA | T-- | T-- | T-- | TAG | 436 | 436 | 436 | 436 | 0.069 | 0.105 |
| *nad4L* | ATA | ATA | ATA | ATA | TAA | TAA | TAA | TAA | 96 | 96 | 96 | 96 | 0.056 | 0.104 |
| *nad6* | ATT | ATT | ATT | ATT | TAA | TAA | TAA | TAA | 163 | 164 | 164 | 163 | 0.13 | 0.172 |
| *cob* | ATG | ATG | ATG | ATG | TAA | TAA | T-- | TAA | 372 | 371 | 371 | 370 | 0.081 | 0.097 |
| *nad1* | ATT | ATT | ATT | ATA | TAA | TAA | TAA | TAG | 311 | 311 | 311 | 312 | 0.059 | 0.095 |

The PCGs exhibited different levels of genetic divergence among the four Aphidoidea species. Nucleotide and amino acid sequences showed overall congruent profiles. The *cox1* gene, followed by *cob*, had the smallest distances at both the nucleotide and amino acid level (Figure 5.5, Table 5.3). This result is consistent with the prevailing mitogenome feature that the mutation rates of *cox1* and *cob* are slower than other mitochondrial genes in most metazoans (Castellana *et al.* 2011). In contrast, *atp8* presented the largest distances, indicating a fast mutation rate for this gene.

### *5.4.3 Codon usage*

A total of 3634 codons, excluding termination codons, were uncovered in *D. noxia*. Approximately equivalent codon numbers were found in *A. pisum* (3635) and *S. graminum* (3643). The codon families exhibited the same pattern in codons per thousand codons in Aphidoidea (Figure 5.3A and 5.4). The three most predominant codon families are Leu, Ile, and Phe. Because of the inconsistency of codon usage in Aleyrodoidea, the comparison of the three superfamilies in Sternorrhyncha was based on the average number of codons for each species. The most common amino acid was found to be Leu

(Figure 5.3B); however, discrepancies were identified among superfamilies in the top four common codons and the abundance of amino acids, such as Ile, Asn and Val.




(A) (B)

**Figure 5.3** Comparison of codon distribution of Sternorrhynchan mtDNAs. (A) comparison among three Aphidiformes species; (B) comparison among three superfamilies in Sternorrhynchan based on average data. Only one Psylliformes species represents this superfamily. CDspT, codons per thousand codons.

Relative Synonymous Codon Usage (RSCU) also reflected the nucleotide bias in the aphid mitogenomes. The four- and two-fold degenerate codon usage in *D. noxia* was A+T biased in the third codon position (Figure 5.4). The largest difference of RSCU among the representative species from the three superfamilies was found in the codon usage of Leu. The RSCU of Leu in *D. noxia* was more biased to TTA in Leu2, compared with the predominance of CTA/CTT of Leu1 in the other two superfamilies. In *D. noxia*, a noticeable bias from TTG to TTA accounted for 96.9% of codons coding for Leu2. The codons of Leu1 in *D. noxia* were biased to CTA and CTT while CTC and CTG codons were lacking. All missing codons were G+C rich. With respect to the number of codons lost in Sternorrhynchan, one codon was absent in *B. tabaci*, while eight codons were missing in *D. noxia*. The largest number (13) of missing codons was found in *A. pisum*.

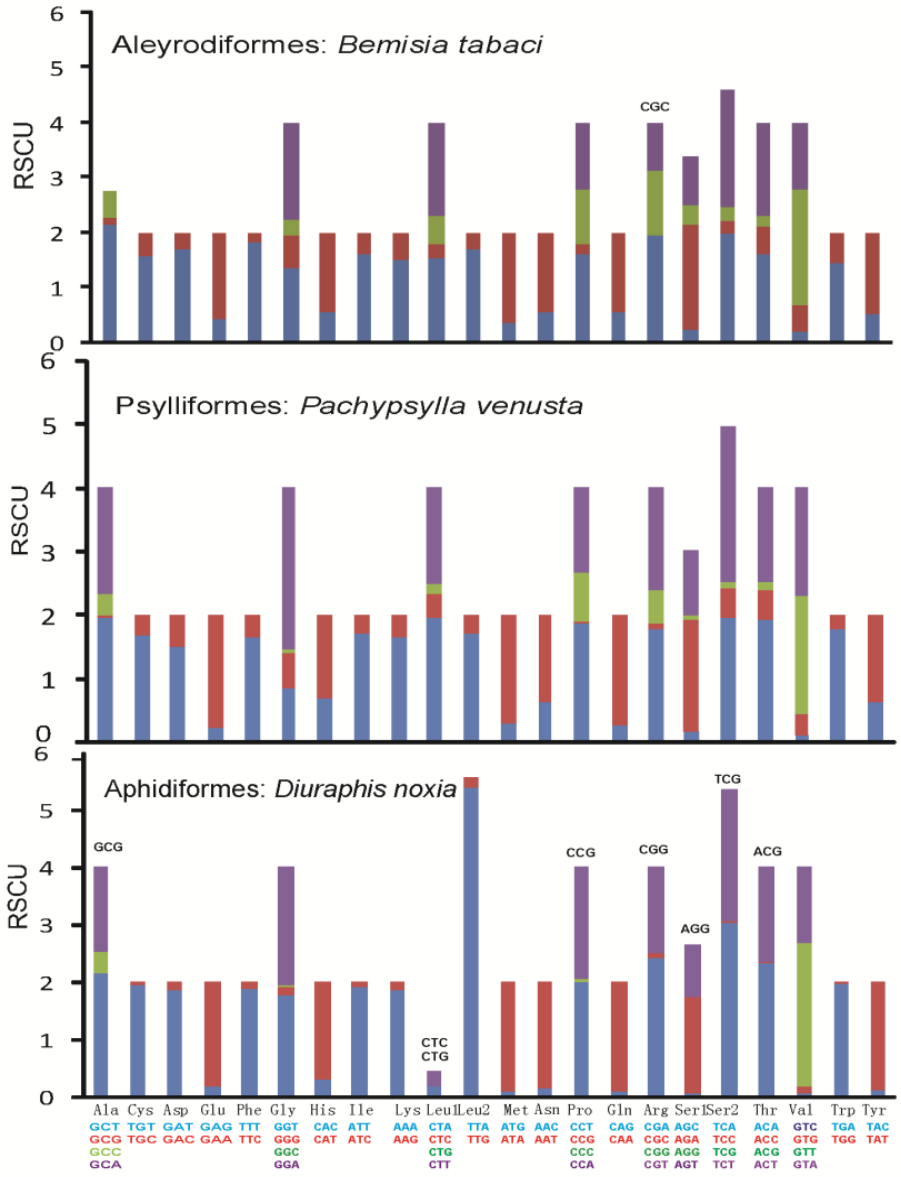



**Figure 5.4** Relative Synonymous Codon Usage (RSCU) of three representative Sternorrhynchan mtDNAs. Codon Families are provided on the X axis. The codons listed on the bar suggest the codons that are absent in the mitogenome.

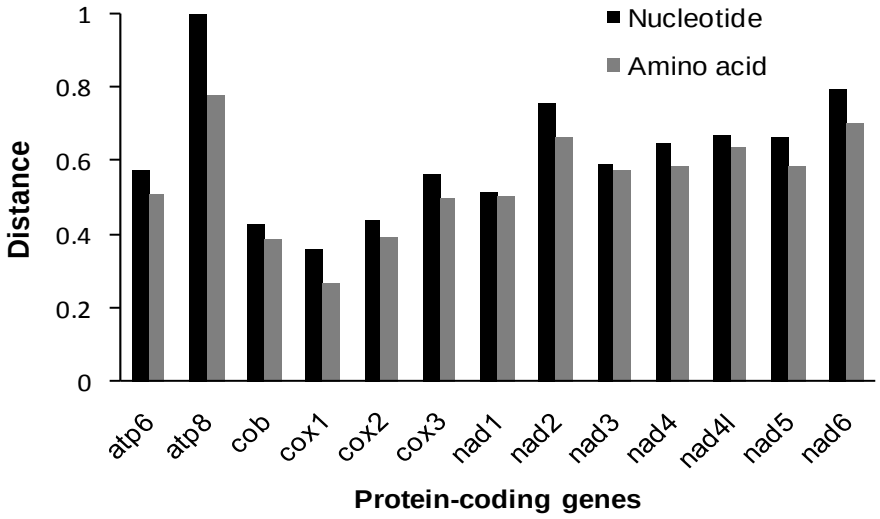



**Figure 5.5** The overall mean distance of the thirteen protein-coding genes among the ten Sternorrhynchan mtDNAs. Black bar: nucleotide distance; grey bar: amino acid distance.

*5.4.4 Transfer RNAs and Ribosomal RNAs*

We found 23 tRNA genes in the *D. noxia* mitogenome, 22 of which are typically found in arthropod mitogenomes. Interestingly, an extra *trnM* gene was found in the control region of *D. noxia*, while one *trnL(UAA)* copy existed in the control region of *S. graminum* and one *trnM*(UAU) in *A. pisum*. Furthermore, two *trnK* genes were discovered in the repeat region of *S. graminum*. Collectively, the three aphids possessed different numbers of tRNA genes. The tRNAs of *D. noxia* ranged in size from 61 bp (*trnS*(AGN)) to 73 bp (*trnK*). Fourteen of the 22 tRNA genes were from the J-strand. All the tRNAs could be folded into a classic clover-leaf secondary structure with the exception of *trnS*(AGN), of which the dihydrouridine (DHU) arm simply formed a loop instead of a stem. Similar DHU arm structures of *trnS*(AGN) were also found in *A. pisum* and *S. graminum*, as reported in many other insect mitogenomes (Negrisolo *et al.* 2011).

tRNA-like pseudogenes were also detected. We identified three *trnM*-like genes and one *trnL*-like gene from the control regions of *A. pisum* and *S. graminum*, respectively. The nucleotide sequences of these tRNA-like pseudogenes all contained an intron ranging from 18 bp to 31bp.

The large and small rRNA subunits (*rrnL* and *rrnS*) in *D. noxia* were located at *trnL*(CUN) to*trnV* and *trnV* to the control region, respectively. Both *rrnL* and *rrnS* are congruent with secondary structure models. The multiple alignment of *rrnL* from four aphid species contained 1092 conserved sites (86.74%), 167 variable sites (13.26%) and a p-distance of 0.072. Likewise, the *rrnS* from these four aphid species showed similar differentiation containing 665 conserved sites (86.81%) and 101 variable sites (13.19%), with a p-distance of 0.069.

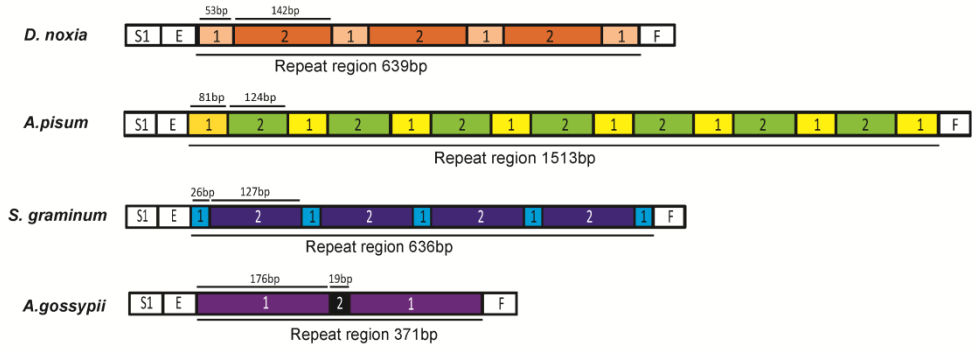
*5.4.5 Non-coding regions*

Two non-coding regions, the control region and the repeat region, both presented relatively lower nucleotide identity in sequence alignment amongst aphid mitogenomes. In the *D. noxia* mitogenome, the control region spanned 664 bp and was located downstream of *rrnS*. It was also A+T rich (86.6%) having a higher A+T content than the whole J-strand. The control region of the four aphid species shared 54.37% nucleotide identity. The investigation of four control regions revealed distinct structural patterns. Firstly, the length of the control region was highly variable, ranging from 662 bp in *S. graminum* to 1,006 bp in *A. pisum*. Secondly, the repeat motif was different based on 75% minimum identity cutoff. For example, the 133-bp consensus motif repeated three times in *A. pisum*, while the 12-bp motif was found two times in *D. vitifoliae*. The AT strings were also not consistent. Finally, *D. vitifoliae* exhibited fewer conserved motifs and elements than the other three aphid species. This point provided molecular evidence for a more distant relationship between *D. vitifoliae* and the other three aphid species, concordant with the taxonomy that *D. vitifoliae* belongs to Phylloxeridae, while the others are from Aphididae. In addition, a conserved sequence block (CSB) was discovered in the four aphid mitogenomes (Figure 5.6B). Indeed, they also shared 78.2% nucleotide identity in the common GC rich motif. The

control region exhibited considerable gaps and low identity reflecting its fast-evolution in aphid mitogenomes.

The repeat region was exclusively discovered in the four aphid species examined as a separate, large non-coding region located consistently between *trnE* and *trnF*. The structure of this region consisted of two tandem consensus sequences that were repeated several times, followed by a partial repeat unit. The repeat region (639 bp) in *D. noxia* consisted of three two-tandem repeat units followed by a partial repeat unit (Figure 5.6A). For the other three aphid species, the number of repeats varied (between 1 and 7 times) and was followed by a partial repeat unit. None of these repeat units were conserved among aphid species, and they had no similarity to any recorded sequences in GenBank.

(A)




(B)

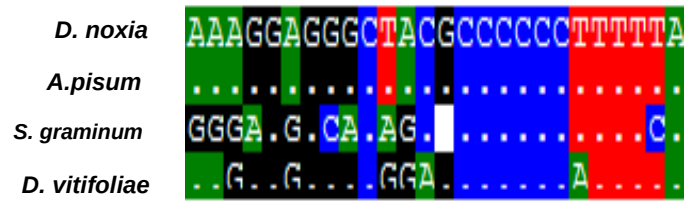



**Figure 5.6** The structure organization of the repeat regions flanked by *trnE* (E) and *trnF* (F) (A) and conserved sequence blocks in the control region of aphids (B).

## 5.5 Discussion

In this study we sequenced the complete mitogenome of the Russian wheat aphid (*D. noxia*) and compared with the mitogenomes of four other aphid species. We report three previously undescribed characteristics that are shared among species within the Aphidoidea. The three main findings are: 1) high A+T content and highly negative GC skew; 2) the presence of a uniquely structured repeat region; and 3) the occurrence of tRNA-like pseudogenes.

### *5.5.1 Extreme nucleotide composition*

High A+T content and highly negative GC skew was found in *D. noxia*, 84.76% and -0.26 and the same pattern was also detected in the other four aphid mitogenomes examined. Similarly, very high (A+T) content (over 84%) has also been found in some Hymenopteran and Dipteran mitogenomes, some of which also exhibited a strong negative GC skew, e.g. lower than -0.25 in *Apis mellifera*, *Bombus ignitus*, and *Melipona bicolour* (McMahon *et al.* 2009; Negrisolo *et al.* 2011). Unlike the aphids, however, it is not a common feature at the order or family level for these insects.

In our study, nucleotide composition differences were also accompanied by significantly different codon usage patterns. (A+T) rather than (G+C) dominated codons predominated, with many GC-rich codons absent in the aphid species studied. The prevailing theories of neutral mutation indicate that GC content strongly determines codon bias, as well as genome-wide mutational pressure and selection (Hershberg & Petrov 2008). In *Drosophila* species it has been reported that G will most likely mutate to A, and has a mutation rate seven times higher than the overall mutation rate of the mitogenome (Haag-Liautard *et al.* 2008). The low GC content found in aphids may therefore provide a potential explanation for the low mutation rate commonly observed in aphid mtDNA studies. Furthermore, the fact that divergence estimates among aphid mitogenomes were lower than that found in other insects in the same order (e.g. white flies) also indicates that aphid mtDNA experiences a relatively lower mutation rate (Thao *et al.* 2004).

Population genetic studies have revealed that aphids generally have low mtDNA diversity (Peccoud *et al.* 2009; Shufran *et al.* 2007; Zhang *et al.* 2012) and phylogenetic studies based on barcoding sequences have also shown limited genetic divergence among Aphidinae species (Lee *et al.* 2011; Wang *et al.* 2011). Similarly, the mtDNA sequences of several Hymenopteran species that have an A+T-rich and G-deficit mitogenome also exhibit relatively low variation at the population level (Franck *et al.* 1998; Franck *et al.* 2001; Hufbauer *et al.* 2004; Shao *et al.* 2004; Widmer *et al.* 1998). Therefore, we conclude that mtDNA is not the marker of choice for population genetic studies in species that have a mitogenome with high A+T content and negative GC skew.

*5.5.2 tRNA-like pseudogenes*

We identified several tRNA-like pseudogenes in the control region of aphids, a phenomenon that has been reported previously in lepidopteran species (Kim *et al.* 2009; Kim *et al.* 2011).The possible explanation for the presence of tRNA-like pseudogenes has been suggested to be the failure to remove the tRNA primer sequence from the nascent DNA strand during mtDNA replication (Cantatore *et al.* 1987; Kim *et al.* 2009; Kim *et al.* 2011).

*5.5.3 The repeat region*

One of the most unusual features of the *D. noxia* mitogenome is the presence of a large non-coding repeat region located between *trnE* and *trnF*. The location and repeat mode of the repeat region are conserved among all available aphid mitogenomes, but the repeat unit sequences and copy numbers are different. It has been suggested that the occurrence of tandem repeat units results from slipped-strand mispairing during mtDNA replication (Moritz & Brown 1987).We found that the repeat unit sequences of all the aphid mitogenomes that we studied could be folded into secondary structures, which may promote replication slippage by stabilizing the slipped strand or blocking the polymerase (Savolainen *et al.* 2000). Similar large repetitive sequences have also been reported in many other insect mitogenomes, but they are generally located within the A+T-rich region. Therefore, the distinctive repeat region represents an aphid-specific feature, indicating that

the repeat region likely originated in the most recent common ancestor of aphids because of the consistent position. It is probable that there was only one ancestral repeat unit in the aphid ancestor. The ancestral repeat sequence could possibly have originated from an extraneous fragment, most likely from an endosymbiont or the nuclear genome of the host. Under relaxed selective constraints, the ancestral repeat unit sequence would have evolved at a fast rate and ultimately would be quite divergent from the original sequence. After aphid speciation, the nucleotide mutations may have continued independently in different aphid species and slipped-strand mispairing may have occurred to different extents among aphids. We hypothesize that these processes may have jointly resulted in the various repeat regions that we observed in the aphid mitogenomes.

The high levels of sequence identity (over 99%) among repeat units in each aphid mitogenome might be a consequence of either recent duplication events or concerted evolution. Under the scenario of recent duplication, the time for mutation fixation in tandem repeats is not sufficient, leading to the observation of high sequence similarity between these repeats. On the other hand, a growing body of evidence supports concerted evolution of duplicated mtDNA sequences (Sammler *et al.* 2011; Tatarenkov & Avise 2007). Such concerted evolution would result in homogenization of tandem repeats, although the underlying mechanism for concerted evolution in mtDNA is not fully understood (Sammler *et al.* 2011; Tatarenkov & Avise 2007). Concerted evolution may therefore be an alternative explanation for the high level of identity found between repeat sequences within aphid species. Clearly, more research is needed in the future if we are to understand the origin and evolution of the unique repeat region found in aphids.

Natural selective pressures on mtDNA tend to minimize non-functional gene length, eliminate the redundant region and shrink genome size (Schneider & Ebert 2004). Mitochondrial evolution has traditionally been viewed as favoring genome size reduction (Boore 1999; McKnight & Shaffer 1997; Rand 1993). From an evolutionary perspective, it makes sense that the non-functional repeat region would be eliminated over time in a highly reduced

and efficient mitogenome. Nevertheless, the presence of multiple tandem units within the repeat region in aphid mitogenomes indicates that the selective force has not been as efficient as expected in eliminating these sequences. Therefore, we predict that the copy number of repeat unit will continue to increase overtime.

*5.5.4 Genomic implications*

A tight relationship between the mitochondrion and certain endosymbionts has been reported in invertebrates (Werren *et al.* 2008). Endosymbionts can contribute to the biological formation of species and can induce indirect selection on insect mtDNA (Gueguen *et al.* 2010). *Wolbachia* affect host biology in many ways including parthenogenesis (Stouthamer *et al.* 1999) and interestingly, it has been found that the parthenogenesis of at least 40 species of Hymenoptera is influenced by the presence of *Wolbachia* (Stouthamer *et al.* 1999). However, no evidence currently exists to suggest that aphid parthenogenesis is triggered by endosymbionts. When *Wolbachia* acts as a driving factor influencing Hymenopteran reproduction, it seems to also affect other cytoplasmic factors, such as mitochondrial variation (Stouthamer *et al.* 1999; Turelli *et al.* 1992). Hurst & Jiggins (2005) reported that the occurrence of *Wolbachia* in native fire ants (*Solenopsis invicta*) altered the distribution of mtDNA variation without affecting nuclear DNA.

The endosymbionts in Aphidinae (obligate *Buchnera* and several other facultative symbionts) have been studied extensively (Moran *et al.* 2008; Moran *et al.* 2009). As mitochondria reside in all types of host cells including germ line cells, they could potentially hitchhike along with the vertical transmitted endosymbionts to the next generation, under the same selective pressures. If endosymbionts do lead to selective sweeps and potentially drive species formation, it may explain why aphid mitogenomes have limited variation and low levels of intraspecific differentiation. However, little is known about the interaction between mitochondria and endosymbionts in aphid species. Future studies on aphids that target this association will provide us with a better understanding of the mutual evolution of hosts, mitochondria and endosymbionts.

## 5.6 References

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573-580.

Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research*, **27**, 1767-1780.

Cantatore P, Gadaleta MN, Roberti M *et al.* (1987) Duplication and remolding of transfer-RNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature*, **329**, 853-855.

Castellana S, Vicario S, Saccone C (2011) Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biology and Evolution*, **3**, 1067-1079.

Franck P, Garnery L, Loiseau A *et al.* (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420-430.

Franck P, Garnery L, Solignac M *et al.* (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): New insights from microsatellite and mitochondrial data. *Evolution*, **52**, 1119-1134.

Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science*, **283**, 1476-1481.

Gueguen G, Vavre F, Gnankine O *et al.* (2010) Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology*, **19**, 4365-4378.

Haag-Liautard C, Coffey N, Houle D *et al.* (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *Plos Biology*, **6**, 1706-1714.

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

Hershberg R, Petrov DA (2008) Selection on Codon Bias. *Annual Review of Genetics*, **42**, 287-299.

Hua JM, Li M, Dong PZ *et al.* (2009) Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. *BMC Evolutionary Biology*, **9**, 134.

Hufbauer RA, Bogdanowicz SM, Harrison RG (2004) The population genetics of a biological control introduction: mitochondrial DNA and microsatellie variation in native and introduced populations of *Aphidus ervi*, a parisitoid wasp. *Molecular Ecology*, **13**, 337-348.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **272**, 1525-1534.

Ingman M, Gyllensten U (2006) mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Research*, **34**, D749-751.

Kim MI, Baek JY, Kim MJ *et al.* (2009) Complete nucleotide sequence and organization of the mitogenome of the red-spotted apollo butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) and comparison with other lepidopteran insects. *Molecules and Cells*, **28**, 347-363.

Kim MJ, Jeong HC, Kim SR *et al.* (2011) Complete mitochondrial genome of the nerippe fritillary butterfly, *Argynnis nerippe* (Lepidoptera: Nymphalidae). *Mitochondrial DNA*, **22**, 86-88.

Lee W, Kang J, Jung C *et al.* (2009) Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. *Molecules and Cells*, **28**, 155-165.

Lee W, Kim H, Lim J *et al.* (2011) Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: updating the global data set. *Molecular Ecology Resources*, **11**, 32-37.

Li H, Gao J, Liu H *et al.* (2011) The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). *International Journal of Biological Sciences*, **7**, 792-804.

McBride HM, Neuspiel M, Wasiak S (2006) Mitochondria: More than just a powerhouse. *Current Biology*, **16**, R551-R560.

McKnight ML, Shaffer HB (1997) Large, rapidly evolving intergenic spacers in the mitochondrial DNA of the salamander family Ambystomatidae (Amphibia: Caudata). *Molecular Biology and Evolution*, **14**, 1167-1176.

McMahon DP, Hayward A, Kathirithamby J (2009) The mitochondrial genome of the 'twisted-wing parasite' *Mengenilla australiensis* (Insecta, Strepsiptera): a comparative study. *BMC Genomics*, **10**, 603.

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165-190.

Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, **323**, 379-382.

Moritz C, Brown WM (1987) Tandem duplications in animal mitochondrial DNAs - variation in incidence and gene content among lizards. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 7183-7187.

Negrisolo E, Babbucci M, Patarnello T (2011) The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. *BMC Genomics*, **12**, 221.

Peccoud J, Ollivier A, Plantegenest M *et al.* (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7495-7500.

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. *Journal of Molecular Evolution*, **37**, 281-295.

Rokas A, Holland PWH (2000) Rave genomic changes as a tool for phylogenetics. *Trends in Ecology & Evolution*, **15**, 454-459.

Sammler S, Bleidorn C, Tiedemann R (2011) Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics*, **12**, 35.

Savolainen P, Arvestad L, Lundeberg J (2000) mtDNA tandem repeats in domestic dogs and wolves: Mutation mechanism studied by analysis of the sequence of imperfect repeats. *Molecular Biology and Evolution*, **17**, 474-488.

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**, W686-689.

Schneider A, Ebert D (2004) Covariation of mitochondrial genome size with gene lengths: Evidence for gene length reduction during mitochondrial evolution. *Journal of Molecular Evolution*, **59**, 90-96.

Shao ZY, Mao HX, Fu WJ *et al.* (2004) Genetic structure of Asian populations of *Bombus ignitus* (Hymenoptera : Apidae). *Journal of Heredity*, **95**, 46-52.

Sheffield NC, Hiatt KD, Valentine MC *et al.* (2010) Mitochondrial genomics in Orthoptera using MOSAS. *Mitochondrial DNA*, **21**, 87-104.

Sheffield NC, Song H, Cameron L *et al.* (2008) A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution*, **25**, 2499-2509.

Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single introduction into the United States. *Journal of the Kansas Entomological Society*, **80**, 319-326.

Simon C, Buckley TR, Frati F *et al.* (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **37**, 545-579.

Stouthamer R, Breeuwer JAJ, Hurst GDD (1999) *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annual Review of Microbiology*, **53**, 71-102.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution*, **13**, 510-524.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

Tatarenkov A, Avise JC (2007) Rapid concerted evolution in animal mitochondrial DNA. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 1795-1798.

Thao ML, Baumann L, Baumann P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evolutionary Biology*, **4**, 25.

Turelli M, Hoffmann AA, Mckechnie SW (1992) Dynamics of cytoplasmic incompatibility and mtdna variation in natural Drosophila simulans populations. *Genetics*, **132**, 713-723.

Wang JF, Jiang LY, Qiao GX (2011) Use of a mitochondrial COI sequence to identify species of the subtribe Aphidina (Hemiptera, Aphididae). *Zookeys*, **122**, 1-17.

Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741-751.

Widmer A, Schmid-Hempel P, Estoup A *et al.* (1998) Population genetic structure and colonization history of *Bombus terrestris* s.l. (Hymenoptera : Apidae) from the Canary Islands and Madeira. *Heredity*, **81**, 563-572.

Zhang B, Edwards OR, Kang L, Fuller SJ (2012) Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction? *Molecular Ecology*, **21**, 2130-2144.

# Chapter 6

# General discussion and conclusions

## 6.1 Research key findings

This PhD study has examined the population genetics of the Russian wheat aphid (RWA, *Diuraphis noxia*), one of the world's most invasive agricultural pests, throughout its native and introduced global range.

This research has four main outcomes. Firstly, an analysis of the population genetic structure revealed that RWAs in China represent a native range expansion rather than a recent introduction. Secondly, significant insights were gained into the source and pathway of global invasion and the potential existence of a wheat-adapted clone that has colonized major wheat growing countries worldwide except for Australia. Thirdly, screening of the salivary gland gene, peroxidase, indicated that this gene is under selection and may be linked to the successful establishment of RWAs on host plants. Finally, aphid mitochondrial genomes had similar nucleotide composition with high AT content that indicated low variation and a slow mutation rate among Aphidoidea. In total, this PhD research has improved our understanding of the evolutionary biology and human-assisted movement of Russian wheat aphids and more generally, has furthered our knowledge of the scientific discipline of invasion genetics.

### *6.1.1 Characterizing the population genetics of RWA in its native range*

In the current study the highest mtDNA diversity was found in Chinese RWA populations compared with other native populations from the Middle East and Tajikistan. Rare haplotypes were found that were unique to Chinese RWA, and showed recent population expansion in last ten thousand years (Chapter 2, Figure 2.2). Considering that mtDNA markers have generally revealed low variability in aphid population genetic studies (Lozier *et al.* 2007; Peccoud *et al.* 2009), the number of haplotyes found in this study provides strong evidence that western China forms part of the native range of RWAs. This was further supported by the microsatellite data that revealed high clonal (MLG) diversity in Chinese RWA. In total, these results refute the hypothesis that the RWA is an exotic species present in China since 1975 (Zhang *et al.* 1999a).

The center of origin of RWAs has been proposed to lie in southern Russia and central-western Asia (Stary 1999; Smith *et al.* 2004; Liu *et al.* 2010). In this PhD study, RWAs were collected from Turkey, Iran, Syria and Tajikistan and these sites exhibited similar microsatellite diversity to China, but limited mtDNA diversity. Only two mtDNA haplotypes were detected from Turkey, Iran, Syria and Tajikistan (N=40 combined), while twelve were found in China (N=60). The higher nucleotide and haplotype diversity in Chinese RWA suggests that this species has had a longer history in the far east of the native range in Asia.

Furthermore, the endosymbiont markers also indicated that China is the center of RWA origin. Chinese haplotypes were located in the center of the native branch of the network (Chapter 3, Figure 3.3), and one haplotype was found in 34 individuals from China. It is possible that some locations such as Russia, Kazakhstan, or Uzbekistan that were not sampled in this study may also have high genetic diversity and be ancestral in the endosymbiont network, however, we hypothesize that western China, together with Kazakhstan may be the native origin of RWAs. This speculation is based on geographic knowledge and genetic data obtained in this study (Chapter 2). RWAs from northern Xinjiang in western China had higher genetic diversity than southern Xinjiang. No barriers to migration were detected among northern China populations located along the thousand kilometer long Tianshan Mountain Range, but migration across the Tianshan Mountain Range from north to south in western China was limited. However, there is no geographic isolation between north-west China and Kazakhstan as Yili Valley is a west-facing valley in China bordered by the Tianshan Mountain range and opening to Kazakhstan. The earliest date of damage by RWAs in Kazakhstan was recorded in 1910 (Zhang *et al.* 1999b).

Finally, an examination of the peroxidase salivary gland gene, a functional protein coding gene critical for aphid survival and feeding, also revealed that Chinese RWA have higher genetic diversity than other central Asian native populations. Five of the six Chinese populations exhibited unique peroxidase

alleles that were only found in East Asia. In summary, all the markers used in this study indicate that China forms part of the center of origin of the native range of RWAs.

If western China is the center of origin, an interesting question is raised; if RWAs spread westward towards Eurasia, why did they fail to spread eastward to invade other wheat-growing provinces in eastern China? RWAs have not been found in Gansu province (Du 2000), which adjoins eastern Xinjiang. The most easterly site in Xinjiang where RWAs have been detected (in very low numbers) in the past is Qincheng, located near the border of Gansu province (Du 2000; Zhang *et al.* 1999a). Despite extensive searching during our study, RWA were not detected in Qincheng, or anywhere east of Mori in Xinjiang. The Climex model however, predicts with high probability that many areas of China provide suitable habitat for RWA, particularly other wheat growing districts in north-eastern China (Liang *et al.* 1999). It is possible that a geographic barrier may exist along the eastern boundary of Xinjiang region, such as the Gobi desert where cereal crops and alternative perennial hosts rarely occur, or that RWAs have a small effective population size in the east that limits their potential invasive capacity. It is also possible that the natural dispersal capacity of RWA is not very strong.

### *6.1.2 Comparing genetic diversity within and among native and introduced populations of RWA*

This study identified strong genetic structure in native and introduced populations worldwide. High genetic diversity was found in the native range indicating long-term historical evolution compared with negligible genetic diversity in introduced populations as a consequence of founder effects.

Results suggest that RWAs have existed in northern China for a long period of time, while in southern China (Cele) low genetic diversity was found indicating more recent range expansion/introduction. All genetic markers including microsatellite, endosymbiont and peroxidase genes, consistently presented low genetic diversity and high homozygosity in southern China. These results suggest that the population in Cele was probably founded

recently by very few RWAs – probably colonizing from the neighboring site, Pishan (Figure 2.3B). Wheat seeds planted in Cele originated from a wheat distribution center located in northern Xinjiang, however no genetic signature of northern populations was detected in Cele (Figure 2.3B). Therefore it is unlikely that RWA eggs/adults were transported to Cele on seed stock or by humans. The most likely hypothesis is that Cele represents a natural range expansion by RWAs to south east Xinjiang, where RWA populations occur in small patches (or oases) discontinuously located along the edge of the Taklamakam Desert. Strong geographical isolation is likely to mean that limited dispersal occurs, resulting in strong founder effects in newly colonized populations in the south.

The expansion of RWAs in western China appears to have been facilitated by agricultural activities associated with the human domestication of wheat. Furthermore, it can be hypothesized that the widespread planting of domesticated wheat may have changed the population structure of RWAs across their entire native range by selecting for wheat-adapted genotypes. Exclusively parthenogenetic reproduction during the wheat growing season would facilitate the fixation of a single wheat-adapted maternal lineage (a "superclone"), as has been observed in other aphid species (Abbot 2011; Harrison & Mondor 2011; Vorburger 2006). Under this hypothesis, all existing RWAs in Xinjiang and elsewhere in its native range would be descendents from this original wheat-adapted genotype – the dominant mtDNA haplotype 1 found in this study. Additional samples from throughout the native distribution of RWA should be analyzed to further test this hypothesis.

*6.1.3 Determining the pathway of global invasion of RWAs*

Examination of the historical pathways of invasion of a species can provide critical information relating to the biological, ecological and anthropogenic factors that enable a species to successfully invade a new environment (Ghabooli *et al.* 2011; Pyšek *et al.* 2011). With increasing commercial global transportation, there is increased potential for plants and animals to be introduced into new ecosystems (Conn 2012; Foucaud *et al.* 2010). The introduction of RWAs into new areas is likely to have resulted from both

human-aided (contaminated materials) and natural (windborne) means. Introduction of RWAs to South Africa, Chile and Argentina probably resulted from human movement, potentially through the transportation of contaminated wheat material. The precise method of introduction is unknown, although the possibility exists that overwintering eggs may have been accidently transported, as eggs are difficult to detect and could survive extended journeys (Plant Health Australia, Pest Review Fact Sheet, 2010). However, the dispersal of RWAs from Mexico to the USA was most likely by wind current (Stoetzel 1987) and results from this PhD study also show that RWAs have dispersed naturally, probably via wind from south-west to south-east China.

The current study has revealed that two independent invasion pathways from the Middle East to Africa have occurred and has also identified that a single invasion event led to the colonization of the New World. All invasive populations could be traced back to several clones from the Middle East (from Turkey and Syria). Major grain research centers are located in Turkey and Syria and it is possible that human movement and/or the distribution of seedling material and cultivars from these centers led to the spread of RWA from the Middle East to the New World. Although there is no evidence to suggest that grain research centers were responsible directly for the two invasions identified in this PhD study (from Syria to Kenya and from Turkey to South Africa and the Americas), it is important to highlight that the frequent transportation of wheat seeds and seedlings among regional institutions poses a significant risk potentially facilitating the spread of this damaging pest.

### *6.1.4 Examining the role of natural selection in global RWA invasion*

The current study examined microsatellite and peroxidase genetic diversity in worldwide RWA populations and found that the peroxidase gene is under selection and may be linked to the successful establishment of RWAs on host plants. Balancing selection acting on peroxidase in native East Asian populations may have affected the invasion and establishment potential of RWAs in this region. The existence of a dominant peroxidase allele present

in Middle Eastern populations indicates that strong purifying selection has acted on standing molecular variation resulting in the selection of one well-adapted allele. Introduced populations were also monomorphic, however it is impossible to conclude whether this is the result of demographic founder effects or directional selection providing an allele-dependent advantage. Further studies investigating peroxidase enzyme activity for specific alleles and injury rates on wheat are required.

*6.1.5 Investigating the low mitogenome variation of RWAs*

Comparative results suggest that divergence estimates among aphid mitogenomes were lower than that found in other insects in the same order (e.g. white flies), indicating that aphid mtDNA experiences a relatively lower mutation rate (Thao *et al.* 2004). Aphid mitogenomes were found to share common features of high A+T content and negative GC skew. These features may provide a potential explanation for the low mutation rate commonly observed in aphid mtDNA studies.

## 6.2 Research implications for biosecurity

Significant insights have been gained into the source and pathway of global RWA invasion and the potential existence of a wheat-adapted genotype that has colonized major wheat growing countries worldwide except for Australia. This research has major biosecurity implications for Australia's grain industry.

The RWA is a major pest of grain producing countries worldwide, except Australia, where it has been given the highest biosecurity risk rating and pre-emptive management strategies have been developed. This PhD research has shown that despite human movement, grain transportation and economic trade, RWAs have not successfully invaded or established in other areas of China. Natural (flight and wind assisted) dispersal appears to be the main mechanism responsible for the expansion of RWA throughout China. Given the geographic isolation of Australia from other wheat producing nations, natural dispersal via wind currents seems unlikely, however air surveillance may be a useful precaution. Provided strict quarantine measures are in place,

there is a low probability that RWAs will invade new areas through grain importation channels. Historically however, insufficient attention has been given to ensuring effective quarantine examination of grain imports and the transfer of plant cultivars from wheat breeding centers, and contaminated material has been the likely source of most of the invasions of RWAs throughout the New World.

Finally, global management of RWA relies on strategies that provide immediate control or damage relief (eg. pesticide use) and long-term solutions based on wheat breeding programs to identify resistant cultivars. Wheat varieties grown in China appear to be tolerant to RWA, and the damage inflicted by RWAs to grain yield is relatively low (0 to 10.56% yield loss; Zhang *et al.* 1999a). To date, no work on resistance genes in these tolerant cultivars has been undertaken and this is an important avenue for future research.

## 6.3 Research gaps

While this PhD study is the most comprehensive population genetics study undertaken on RWA to date, some questions still remain unanswered. Firstly, this study revealed the unexpected result that RWA populations in Ethiopia had very high microsatellite genetic diversity and most likely form part of the native range. However, both the endosymbiont and salivary gland genes failed to amplify despite exhaustive efforts. This was the only sample worldwide for which this problem was encountered. Why? Mitochondrial DNA was easily amplified and revealed that the samples were correctly identified as RWA. It is presumed that strong isolation and independent diversification in Ethiopia may have resulted in such a level of divergence that primers no longer matched or that the lineages underwent replacement by rapidly evolving genes (i.e. pseudogenes of endosymbiont). Such isolation might also explain why Ethiopian RWAs have not (apparently) expanded into neighboring countries.

Secondly, little insight into the genetic structure of European RWA has been gained. Unfortunately, only one French and one Hungarian RWA (individual) was obtained from Europe. Some populations in Europe such as Ukraine are believed to be part of the native range of RWAs (Stary 1999; Zhang *et al.* 1999b), however it is unclear in other countries such as Hungary and Czech whether they are introduced or native (Stary 1999; Zhang *et al.* 1999b). What is the border of the native range in Europe? In this PhD study, the only sample from Hungary suggested that it had a similar genetic structure to native populations, while the French sample presented a signature of being an introduced population (Chapter 3, Figures 3.1, 3.2 and 3.3). However, no conclusions can be formed on the pattern of RWA distribution in Europe based on two individuals, nor can the boundary between native and introduced populations be discerned. If neighboring native and introduced populations can be identified, it would be interesting to study the admixture between populations and where they are sympatric, compare relative competitive ability and performance on different hosts.

## 6.4 Research contribution to the field of invasion genetics

### *6.4.1 Evolutionary mechanisms of invasion*

Successful biological invasion can be achieved by two mutually exclusive strategies, by having a general purpose genotype or through rapid adaptation (Parker *et al.* 2003). Invasive species with a general purpose genotype can colonize a wide range of environments because of phenotypic or developmental plasticity (Baker 1974). Several factors point to this second strategy as contributing to RWA invasion success. RWA, like all aphids, have a propensity for both phenotypic and developmental plasticity. We have also shown that invasive populations of RWA are characterized by a single mitochondrial haplotype, a similar nuclear multilocus genotype, as well as a shared endosymbiont haplotype. Although we did not examine the physiological tolerance and plasticity of RWA in invasive populations, this limited genetic variation amongst invasive populations is consistent with the idea of a general purpose invasive RWA genotype. Asexual reproduction is also thought to facilitate successful invasion, proliferation and spread by

eliminating the need to find mates at low densities. Indeed, invasive populations of RWA have all been asexual.

In less plastic species, rapid local adaptation is often essential for the establishment and spread of invasive species in novel habitats, often leading to the formation of regional ecotypes (Clements & Ditommaso 2011; Sakai *et al.* 2001; Sexton *et al.* 2002). Rapid adaptation during invasions most often occurs in an outcrossing breeding species with high genetic diversity, multiple independent introductions, and frequent gene flow among populations (Parker *et al.* 2003). Many adaptations have been reported in invasion genetics studies, such as the evolutionary changes of flight morphology in *Pararge argeria* (Hill *et al.* 1999), the selection of flowering time at different latitudes in *Solidago altissima* (Weber & Schmid 1998), and alternative modes of reproduction in introduced aquatic *Butomus umbellatus* in North America (Eckert *et al.* 2003). However, incipient adaptations in response to invasions are difficult to detect in terms of underlying selection on dispersal capacity and physiological tolerance to immediate environmental stress (Lee 2002).

In Chapter 4, evidence has been provided that rapid adaptation has occurred in both native and invasive RWA populations by directional selection of a salivary peroxidase gene in native populations. Rapid evolution in salivary gland genes has also been reported in the pea aphid by gene duplication and diversifying selection (Carolan *et al.* 2010). Despite limited genetic variability in invasive RWA populations, resistance-breaking biotypes have evolved in both South Africa and the USA (Haley *et al.* 2004; Tolmay *et al.* 2007). Aphid salivary gland proteins are thought to be essential in regulating host plant interactions, including resistance responses (Walling 2008). Hence, rapid adaptation in salivary gland genes may be responsible for these instances of biotype evolution.

Studies of aphid invasion genetics provide an opportunity to examine the two mechanisms of invasion (“general purpose genotype” vs “rapid adaptation”) and to reconcile the potential occurrence of both mechanisms acting at the

same time because aphids reproduce asexually and undergo rapid local adaptation in response to environmental change.

*6.4.2 Evolution and local adaptation*

The evolution of locally adapted genotypes requires that sufficient genetic diversity and a large candidate gene pool exists (Sakai *et al.* 2001). In the current study, native populations of RWA were found to possess high levels of genetic variation as a consequence of having a holycyclic life cycle and high levels of gene flow within each local area. This is expected to result in RWAs having a high capacity for range expansion in native areas. However high rates of gene flow could also constrain adaptation to local conditions as gene flow may limit invasive species range expansion (Holt & Keitt 2005; Sakai *et al.* 2001). RWAs in northern Xinjiang (China) had the highest genetic diversity in the far eastern Asian range and a high rate of gene flow was found among populations (see Chapter 2). However, limited long distance migration was detected. These results suggest that gene flow from the center of the distribution range may prevent adaptation of the peripheral population, and then prevent more distant range expansion (as has been proposed by Sexton *et al.* 2009). Thus, gene flow not only enhances the homogenizing effect that stabilizes effective population size (Motro & Thomson 1982), but also may prevent the potential spread of this species from its initial range to new areas. Comparative invasion biology studies examining the level of genetic exchange among populations are required if we want to predict possible future expansions of invasive species.

Significant local adaptation in native RWA populations arises as a consequence of a host plant-aphid interaction and selection acting on salivary gland genes (Chapter 4). Strong evidence for natural selection acting on the peroxidase salivary gland gene in native populations in the Middle East and East Asia was found and peroxidase gene diversity was highly structured in these two different regions. As introduced populations were monomorphic for peroxidase, no local adaptation was detected and it was impossible to disentangle demographic founder effects from possible

directional selection effects. Definitive conclusions cannot be formed as to whether peroxidase plays a functional role in enhancing RWA invasiveness.

Invasive species may expand their distributional ranges via local adaptation, and the rapid response to natural selection resulting from high levels of genetic variance within populations could help to increase the rate of outcrossing as well as the number of founders that newly colonize an area. The capacity for asexual reproduction paradoxically influences the invasion genetics of aphids. Predominantly anholycyclic parthenogenesis in invasive ranges decreases genetic diversity but increases rapid population expansion and enables rapid local adaptation to environmental stress (Simon *et al.* 1999). Furthermore, population bottleneck events occur during colonization and genetic drift will have an especially strong effect in less genetically diverse introduced populations compared with native populations.

Two important evolutionary findings have arisen from this study on the global invasion of RWAs. Firstly, strong evidence of local adaptation in both native and invasive populations was found i.e., directional selection acting on salivary gland peroxidase in native populations and new biotypes emerging in invasive populations. Secondly, human aided dispersal has spread RWA into a range of new exotic environments, potentially accelerating the evolution of new local adaptations. In the current study, very little genetic variation was found among invasive populations, however only three decades have passed since this pest has invaded the New World. Evolutionary changes will become apparent as RWA populations respond and adapt to new and dynamic environments (Lee 2002; Sakai *et al.* 2001). Understanding the evolutionary processes that influence biological invasions is critical to predicting the long-term ecological success of an invasive species and the long-term risks of broad scale establishment of an invasive species (Parker *et al.* 2003).

*6.4.3 Genetic diversity and invasion success*

Populations of invasive species generally exhibit low genetic diversity and the genetic consequences of this can be broadly categorized into two groups;

low genetic diversity does not limit population growth and persistence and does not lower the probability of invasion success or conversely, low genetic diversity erodes as a result of inbreeding, slowing evolutionary responses to selection and constraining a species ability to adapt to the new environment. The fact that new virulent biotypes of RWAs have emerged in introduced populations in the USA and South Africa in the last decade provides evidence that RWAs are capable of rapid evolution and local adaptation (Basky *et al.* 2002). Results from this PhD study indicate that both USA and South African RWA populations possess very little genetic diversity, and in South Africa it appears that genetic variation may have eroded over the last decade. Three multilocus genotypes were found in 2000, while the 2011 sample was monomorphic. Therefore, limited genetic diversity does not appear to constrain RWA invasion success. Similar results have been reported for other invasive species (Sakai *et al.* 2001; Suarez *et al.* 2008).

*6.4.4 Genetic markers for studying recent invasions*

Finally, molecular genetic markers have been used widely in invasive species studies to provide information on temporal and spatial genetic patterns. Neutral markers are generally targeted and in most cases, they have provided an adequate level of resolution and have enabled the advancement of our understanding of invasion genetics. However, for some taxonomic groups (such as aphids) and in the case of very recent invasion events traditional markers have proven to be inadequate. In particular, previous studies that have examined the invasion history of RWA have failed to identify sufficient genetic variation to discriminate among introduced populations (Shufran *et al.* 2007; Shufran & Payton 2009). In this PhD study, fast evolving *Buchnera* genes were used to examine RWA global invasion during the last thirty years, that is, population genetic changes over an extremely short time frame. Microsatellite markers failed to provide such a level of resolution. The fast mutation rate of evolution of RWA endosymbionts (Moran *et al.* 2009) enabled a high level of sensitivity for analysis of recent global invasion by this damaging pest species. Results confirm that a tight relationship between the bacterial symbiont and insect host exists, and rapid evolution of the endosymbiont at the population level of the species. The use

of endosymbiont markers in aphid invasion genetics has proven to be a novel and valuable approach. Furthermore, it is becoming clear that endosymbionts are much more common in insects than previously thought (Kitushi 2009; Oliver & Moran 2009). As such, endosymbiont sequence analysis should be applied to population genetic studies in insects more broadly.

## 6.4 References

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

Baker HG (1974) The evolution of weeds. *Annual Review of Ecology and Systematics*, **5**, 1-24.

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.

Carolan JC, Caragea D, Reardon KT *et al.* (2011) Aphids as plant pathogens: Predicted effector molecules and their evolution in the salivary secretome of the pea aphid (*Acyrthosiphon pisum*). *Journal of Proteome Research,* **10**, 1505-1518.

Clements DR, Ditommaso A (2011) Climate change and weed adaptation: can evolution of invasive plants lead to greater range expansion than forecasted? *Weed Research*, **51**, 227-240.

Conn JS (2012) Pathways of invasive plant spread to Alaska: III. contaminants in crop and grass seed. *Invasive Plant Science and Management*, **5**, 270-281.

Du BR (2000) Investigation of the distribution and suitable area of Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) in Xinjiang. *Xinjiang Agricultural Science, Z1*, 82-85.

Eckert CGK, Lui K, Bronson P *et al.* (2003) Population genetic consequences of extreme variation in sexual and clonal reproduction in an aquatic plant. *Molecular Ecology*, **12**, 331-344.

Foucaud J, Orivel J, Loiseau A *et al.* (2010) Worldwide invasion by the little fire ant: routes of introduction and eco-evolutionary pathways. *Evolutionary Applications*, **3**, 363-374.

Ghabooli S, Shiganova TA, Zhan A *et al.* (2011) Multiple introductions and invasion pathways for the invasive ctenophore *Mnemiopsis leidyi* in Eurasia. *Biological Invasions*, **13**, 679-690.

Haley SD, Peairs FB, Walker CB *et al.* (2004) Occurrence of a new Russian wheat aphid biotype in Colorado. *Crop Science*, **44**, 1589-1592.

Harrison JS, Mondor EB (2011) Evidence for an invasive aphid “superclone”: Extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States. *PLoS One*, **6**, e17524.

Hill JK, Thomas CD, Blakeley DS (1999) Evolution of flight morphology in a butterfly that has recently expanded its geographic range. *Oecologia*, **121**, 165-170.

Holt RD, Keitt TH (2005) Species’ borders: a unifying theme in ecology. *Oikos*, **108**, 3-6.

Kikuchi Y (2009) Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments*, **24**, 195-204.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386-391.

Liang HB, Zhang RZ, Zhang GX (1999) Prediction of suitable areas for Russian wheat aphid in China. *Acta Entomologica Sinica*, **42S**, 55-61.

Lozier JD, Roderick GK, Mills NJ (2007) Genetic evidence from mitochondrial, nuclear and endosymbiont markers for the evolution of host plant associated

species in the aphid genus *Hyalopterus* (Hemiptera: Aphididae). *Evolution*, **61**, 1353-1367.

Moran NA, McLaughlinHJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, **323**, 379-382.

Motro U, Thomson G (1982) On heterozygosity and the effective size of populations subject to size changes. *Evolution*, **36**, 1059-1066.

Oliver KM, Moran NA (2009) Defensive symbionts in aphids and other insects. In: White J, Torres M, editors. Defensive Mutualism in Microbial Symbiosis. 1st ed. CRC Press, Taylor & Francis Group. pp. 129–148.

Parker IM, Rodriguez J, Loik ME (2003) An evolutionary approach to understanding the biology of invasions: local adaptation and general-purpose genotypes in the weed *Verbascum thapsus*. *Conservation Biology*, **17**, 59-72.

Peccoud J, Ollivier A, Plantegenest M *et al.* (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences USA,* **106**, 7495-7500.

Pyšek P, Jarošík V, Pergl J (2011) Alien Plants Introduced by Different Pathways Differ in Invasion Success: Unintentional Introductions as a Threat to Natural Areas. *PLoS ONE*, **6**, e24890.

Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305-332.

Sexton JP, McKay JK, Sala A (2002) Plasticity and genetic diversity may allow saltcedar to invade cold climates in North America. *Ecological Applications*, **12**, 1652-1660.

Sexton JP, Mclntyre PJ, Angert AL, Rice KJ (2009) Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415-436.

Shufran KA, Kirkman LR, Puterka GJ (2007) Absence of mitochondrial DNA sequence variation in Russian wheat aphid (Hemiptera: Aphididae) populations consistent with a single introduction into the United States. *Journal of the Kansas Entomological Society,* **80**, 319-326.

Shufran KA, Payton TL (2009) Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology,* **102**, 440-445.

Simon JC, Baumann S, Sunnucks P *et al.* (1999) Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, **8**, 531-545.

Smith CM, Belay T, Stuffer C *et al.* (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) populations virulent to the *Dn4* resistance gene. *Journal of Economic Entomology*, **97**, 1112-1117.

Stary P (1999) Distribution and ecology of the Russian wheat aphid, *Diuraphis noxia* (Kurdj.), expanded to Central Europe (Hemiptera: Aphididae). *Journal of Pest Science*, **72**, 25-30.

Stoetzel MB (1987) Information on and identification of *Diuraphis noxia* (Homoptera: Aphididae) and other aphid species colonizing leaves of wheat and barley in the United States. *Journal of Economic Entomology*, **80**, 696-704.

Suarez AV, Holway DA, Tsutsui ND (2008) Genetics and behavior of a colonizing species: The invasive Argentine ant. *American Naturalist*, **172**, S72-84.

Thao ML, Baumann L, Baumann P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evolutionary Biology*, **4**, 25.

Tolmay V, Lindeque RC, Prinsloo GJ (2007) Preliminary evidence of a resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae) in South Africa. *African Entomology*, **15**, 228-230.

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

Walling LL (2008) Avoiding effective defenses: Strategies employed by phloem-feeding insects. *Plant Physiology*, **146**, 859-866.

Weber E, Schmid B (1998) Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. *American Journal of Botany,* **85**, 1110-1121.

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

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