This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2

Feng Cui, C. Michael Smith, John Reese, Owain Edwards, Gerald Reeck

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Cui, F., Smith, C. M., Reese, J., Edwards, O., & Reeck, G. (2012). Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. Retrieved from http://krex.ksu.edu

Published Version Information

Citation: Cui, F., Smith, C. M., Reese, J., Edwards, O., & Reeck, G. (2012). Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2. Insect Science, 19(4), 429-440.

Copyright: © 2012 Institute of Zoology, Chinese Academy of Sciences

Digital Object Identifier (DOI): doi:10.1111/j.1744-7917.2011.01487.x

Publisher's Link: http://onlinelibrary.wiley.com/doi/10.1111/j.1744-7917.2011.01487.x/pdf

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at http://krex.ksu.edu

Polymorphisms in salivary-gland transcripts of Russian wheat aphid biotypes 1 and 2

Feng Cui[1,2,5*], C. Michael Smith[3], John Reese[3], Owain Edwards[4,5], Gerald Reeck[1]

1 Department of Biochemistry, Kansas State University, Manhattan, Kansas, USA

2 State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

3 Department of Entomology, Kansas State University, Manhattan, Kansas, USA

4 CSIRO Entomology, Canberra, Australia

5 Cooperative Research Centre for Plant Biosecurity, Canberra, Australia

* Corresponding author: Feng Cui, State Key Laboratory of Integrated Management of Pest Insects & Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China. Tel: 86-10-64807218. Fax: 86-10-64807099. Email: cuif@ioz.ac.cn.

Short title: Gene polymorphism of RWA biotypes

Abstract

The Russian wheat aphid (RWA), *Diuraphis noxia* (Mordvilko) (Homoptera: Aphididae), is a major pest of small grains. As with plant-feeding aphids in general, the interaction between RWA and host plants is governed, on the insect side, by proteins and enzymes of saliva. In this work, we examined sequence variations in transcripts encoding proteins and enzymes of RWA salivary glands. We conducted RT-PCR in RWA biotypes 1 and 2 using primers derived from pea aphid orthologs, and cloned regions of 17 putative salivary gland transcripts. For 4 of the transcripts, we observed no difference in sequences between the two biotypes. For the other 13 transcripts, for example, the transcripts encoding sucrase, trehalase and protein C002, large amount of variations, both within each biotype and between the two biotypes, were observed. Usually the two biotypes shared only one variant, which was typically the most common variant in both biotypes. Most of the transcripts had more non-synonymous than synonymous codon changes among their variants. Our results offer possible molecular markers for distinguishing the two biotypes and insights into their evolution.

Key words biotype, molecular marker, polymorphism, Russian wheat aphid, salivary gland, transcript

1 # Introduction

2 The Russian wheat aphid (RWA), *Diuraphis noxia* (Mordvilko) (Homoptera:
3 Aphididae), is a major phloem-feeding pest. It mainly damages wheat, *Triticum*
4 *aestivum* L., and barley, *Hordeum vulgare* L. throughout the major small grain
5 production areas of the world except for Australia (Stoetzel, 1987). Until 1935 this
6 aphid was found only in the Ukraine, central Asia and western Asia. In the 1940s, its
7 range expanded to include Europe, in 1950-60s Africa, in the 1980s North and South
8 America (Zhang *et al*., 1999). The first detection of this aphid in the United States was
9 near Muleshoe, Texas, in 1986 (Peairs, 1987). Since then the RWA has spread across
10 17 states, following a northerly and westerly direction, largely lacking eastward
11 movement. Initial phylogeographic analysis using random amplified polymorphic
12 DNA and allozyme markers demonstrated that RWA populations in the USA, Canada,
13 and Mexico were most closely related to those of South Africa, France, and Turkey
14 (Puterka *et al*., 1993). More recent phylogeographic analyses of several global
15 populations using amplified restriction fragment polymorphism markers demonstrate
16 two major global clades: one from the Middle East-Africa and one from Europe (Liu
17 *et al*., 2010).

18 RWA biotypes have been designated based on the damage resulting from aphid
19 feeding on wheat cultivars containing resistance genes *Dn1* to *Dn9*. Using this system,
20 5 biotypes, RWA1-RWA5, have been identified in the United States (Puterka *et al*.,
21 2007). In studies of samples collected in 2005 from 98 fields of wheat or barley in
22 Oklahoma, Texas, New Mexico, Colorado, Kansas, Nebraska and Wyoming, only

1 RWA1 and RWA2 were found, and RWA2 was the predominant biotype (Puterka *et*
2 *al*., 2007). RWA1 is virulent only to wheat carrying resistance genes *Dn1*, *Dn8* and
3 *Dn9*. RWA2 is virulent to wheat containing any of the *Dn* genes other than *Dn7*
4 (Haley *et al*., 2004; Puterka *et al*., 2007; Qureshi *et al*., 2005). Liu *et al*. (2010)
5 demonstrated that RWA1 and RWA2 are contained in the Middle East-African clade,
6 and RWA3, RWA4, and RWA5 are part of the European clade. RWA biotypes also
7 occur in Africa, Asia, Europe, and South America (Basky, 2003; Dolatti *et al*., 2005;
8 Malinga *et al*., 2007; Smith *et al*., 2004; Tolmay *et al*., 2007).

9 Saliva, injected into plant tissue, is the point of contact between aphid and plant.
10 Proteins and enzymes in saliva are believed to play several roles in allowing
11 continued feeding by an aphid on the phloem (Miles, 1999; Tjallingii, 2006). In broad
12 terms, some, and possibly all, of the proteins of aphid saliva can be thought of as
13 "effectors," a term introduced to designate proteins secreted by plant pathogens for
14 the purpose of establishing "colonization" of the plant by the pathogen (Hogenhout *et*
15 *al*., 2009). Indeed this suggestion has been made specifically in the case of the RWA
16 (Boyko *et al*., 2006; Lapitan *et al*., 2007; Cooper *et al*., 2010), and RWA secretes
17 protein effectors that differ among biotypes (Van Zyl, 2007). However, for the most
18 part, we know little about the functions of individual components of aphid saliva.
19 Recent descriptive and functional studies on aphid salivary gland genes and proteins
20 support effector functions for many aphid salivary proteins (Carolan *et al*., 2009, 2011;
21 Bos *et al*., 2010). Results from the pea aphid also suggest that genes transcribed in
22 salivary glands are evolving faster than their orthologs in other insects (Carolan *et al*.,

1 2011).

2 Several studies have looked for genetic differences between RWA1 and RWA2.
3 Shufran *et al.* (2007) found little or no difference in the sequence of a 332-base-pair
4 region of the mitochondrial gene encoding cytochrome oxidase I in RWA samples
5 collected from 1986 through 2006, a time period in which RWA2 arose. In a
6 subsequent study, Shufran and Payton (2009) reported little or no variation in the
7 same gene-sequence or in several simple-sequence repeat loci between RWA1 and
8 RWA2. On the other hand, Weng *et al*. (2007) reported that 19 of 57 simple-sequence
9 repeat markers examined differed between RWA1 and RWA2.

10 In this study, we looked for polymorphism in several RWA salivary-gland
11 transcripts that were orthologs of pea aphid salivary-gland transcripts. Our results
12 pointed to considerable polymorphism, both within RWA biotypes 1 and 2 and
13 between these biotypes. We interpreted the variants as allelic variation, and the
14 differences that we observed readily allowed distinction between RWA1 and RWA2
15 at the molecular genetic level. The existence of extensive polymorphism in genes
16 encoding proteins and enzymes of salivary glands (some of which were likely
17 components of saliva) suggested that adaptation to new cultivars may be achieved,
18 and achieved rapidly, by shifts in the frequencies of alleles of one or more such genes.

19 Materials and methods

20 *Insects*

21 Populations of *D. noxia* biotype 1 (RWA1) collected from wheat fields near Hays,
22 KS, in 2002, and biotype 2 (RWA2) individuals collected from wheat fields near

1 Briggsdale, CO (via the USDA-ARS Plant Science Research Laboratory at Stillwater,
2 OK), in 2003 were cultured continuously in separate locations in the greenhouse on
3 susceptible 'Jagger' wheat plants at Kansas State University before use in the
4 experiments. The identity of each biotype was verified in diagnostic plant differential
5 greenhouse assays at Stillwater, OK, and Manhattan, KS.

6 *mRNA extraction and cDNA synthesis*

7 Fifty milligrams of RWA (at least 50 aphids), including adults and nymphs, was
8 homogenized with a polypropylene pestle in 1 ml of TRIZOL regent (Invitrogen,
9 Carlsbad, CA, USA). Total RNA was extracted according to the protocol supplied
10 with the reagent. Twenty microgram of total RNA was treated with TURBO
11 DNA-free kit (Ambion, Austin, TX, USA) to remove genomic DNA contamination.
12 Up to 5 ug of DNA-free total RNA was reverse-transcribed into cDNA with
13 SuperScript™ III first strand synthesis system for RT-PCR (Invitrogen, Carlsbad, CA,
14 USA).

15 *RT-PCR and cloning of regions of transcripts*

16 We selected 17 transcripts for this work, based on sequences obtained from
17 salivary-gland cDNA libraries of the pea aphid, *Acyrthosiphon pisum* (Carolan *et al*.,
18 2011) and predicted to possibly encode secreted proteins (that is, proteins of saliva)
19 based on the existence of hydrophobic regions (possible secretion signals) at the
20 N-termini of the encoded polypeptides. In the case of the laccase-1 transcript, an EST
21 has not been reported in a salivary-gland EST library, but studies in one of our
22 laboratories indicates that the protein occurs in salivary glands (Liang, 2006). Regions

1 within the RWA orthologs of these transcripts were amplified from RWA1 and RWA2

2 cDNAs and sequenced. Primers for PCR-amplification were based on sequences of

3 the pea aphid ESTs. Sequences of primer-pairs and predicted PCR product sizes for

4 the 17 transcripts are detailed in Table 1. PCRs (25 μL) were comprised of 0.5 μM of

5 each primer, 1 μL of template cDNA and 12.5 μL of PCR master mix that consisted of

6 50 units/ml *Taq* DNA polymerase, 400 μM dNTP and 3 mM $MgCl_2$ (Promega,

7 Madison, WI, USA). Initial denaturation of the template cDNA was at 94 °C for 2 min,

8 and was followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s to

9 1 min depending on the size of products, and a final step for 5 min at 72 °C. The PCR

10 products were separated in 1% agarose gel and then purified from the gel with

11 QIAquick gel extraction kit (Qiagen, Valencia, CA, USA). The purified products were

12 T-A cloned into pCR2.1-TOPO vector and transfected into TOP10 or TOP10F' cells.

13 At least 4 plasmids for each transcript of each biotype were extracted with QIAprep

14 spin miniprep kit (Qiagen, Valencia, CA, USA) and sequenced in the DNA

15 Sequencing and Genotyping Facility at Kansas State University, using an Applied

16 Biosystems 3730 DNA Analyzer. We obtained 750-900 bases of sequence information

17 on each run. The sequencing was found to be error-free for 600 bases. The clones

18 longer than 600 bases were always sequenced in two directions. Sequences for each

19 transcript were aligned with the online ClustalW2 sever at EBI

20 (www.ebi.ac.uk/Tools/clustalw2/index.html) and formatted with BioEdit software.

21 *Phylogenetic analysis of RWA and pea aphid salivary genes*

22 The transcript variants identified in RWA for each salivary gland gene were analyzed

along with the predicted paralogs of these genes from the pea aphid genome (International Aphid Genomics Consortium, 2010). The maximum parsimony trees were conducted with Mega4 (Tamura *et al.*, 2007) using complete deletion of gaps in the amino acid sequence alignments. Bootstrap analysis (1000 replicates) was applied to evaluate the internal support of the tree topology.

Results

Sequence analysis of the cloned transcripts

We amplified and cloned regions of 17 transcripts from RWA biotypes 1 and 2. These regions were interpreted to be RWA orthologs of 17 pea-aphid transcripts obtained as ESTs from salivary-gland cDNA libraries, having sequence identity at the nucleotide level as high as 95% between the RWA and pea aphid sequences (Table 2). We worked with genes that encoded proteins with a secretory signal peptide in the N-terminal hydrophobic regions. In several cases there was strong evidence for secretion of the encoded protein. Protein C002, for instance, was known to be transferred to plant tissue during pea aphid infestation of a host plant (Mutti *et al.*, 2008), and there was evidence for secretion of dipeptidyl carboxypeptidase, glucose dehydrogenase, and peptidase M1 in aphid saliva in one or more of several recent proteomics studies (Carolan *et al.*, 2009, 2011; Harmel *et al.*, 2008). Among the 17 transcripts two (AphidB1_C07_t7_050 and ID0AAH13AH01ZM2) cannot be annotated, i.e. not similar in sequence to previously studied genes or proteins.

Polymorphism between and within RWA biotypes

For four transcripts (those encoding a coated-vesicle membrane protein, a

1 peroxidase and the two non-annotatable proteins) we detected no polymorphism
2 within or between biotypes (Table 3). For the other 13 transcripts, we observed
3 polymorphism both within and between the biotypes, at the nucleotide level and the
4 predicted amino sequences (Table 3). Most of the variations were detected only once
5 within individual transcripts (Figure 1 - Figure 3 for representative examples; also see
6 Figure S1 - Figure S11) but some variants were found more than once in one or both
7 of biotypes, as in the case for cathepsin B (Figure S3). Usually the two biotypes only
8 shared one variant and this was usually the predominant variant in both of the two
9 biotypes.
10 To illustrate the nature of the observed polymorphisms, we presented our findings
11 on three transcripts here. All other sequences were shown in Supplemental Material.
12 The C002 transcript (Figure 1) encoded a protein required in the pea aphid for feeding
13 on a host plant (Mutti *et al*., 2006, 2008). For this transcript, 9 variants at the amino
14 acid level in the two biotypes were found, with 4 variants found only in RWA1 and
15 another 4 found only in RWA2. The sucrase transcript showed the highest
16 polymorphism among the 17 transcripts we examined (Figure 2). Nine amino acid
17 variants in RWA1 and 7 amino acid variants in RWA2 were found. A 68-nucleotide
18 deletion resulted in shorter ORF in RWA2 variant. The most complicated
19 polymorphism was observed in the trehalase transcript (Figure 3). There were 8
20 variants at the amino acid level in the two biotypes taken together. In addition to
21 single amino acid substitutions at several positions, 4 different polypeptide lengths
22 were encoded. In RWA1 there were three polypeptide lengths. These lengths (in the

1 region we cloned) included: 141 residues; 151 residues (a mutation from TAA to CAA
2 elongated the polypeptide) and 161 residues (due to a 96-nucleotide insertion). In
3 RWA2, two lengths of this polypeptide region were observed, one of 141 residues and
4 the other of 67 residues (a 2-nucleotide deletion shortened and shifted the open
5 reading frame).

6 *Non-synonymous and synonymous mutations*

7 Within the 17 cloned regions in our study, most of the transcripts had more
8 non-synonymous (N) than synonymous (S) codon changes among their variants, with
9 the ratio of N/S mutations larger than 1, even infinite. Only transcripts encoding
10 peptidase M1 and Emp 24 had fewer or equal non-synonymous mutations compared
11 to synonymous mutations. Overall, approximately 2/3 of codon changes were
12 non-synonymous and the ratio of N/S mutations across all 17 transcripts was 1.9
13 (Table 4).

14 *Phylogenetic analysis*

15 A phylogenetic analysis was conducted as a means of discriminating between
16 allelic and paralog variation in the transcripts identified for each prospective salivary
17 gland locus. All the transcripts were compared to the gene sequences of all paralogs in
18 the pea aphid genome (International Aphid Genomics Consortium, 2010) except the
19 four transcripts without polymorphism. Only one paralog was found in the pea aphid
20 genome for transcripts C002, cathepsin L, probable ER retained protein and Emp24.
21 For the other 9 transcripts, all the variants detected clustered as a single group related
22 to the same pea aphid paralog (or group of related paralogs) (Figure 4 for

representative examples; also see Figure S12, Figure S13).

Discussion

We have observed unexpectedly high variation in transcript sequences in the Russian wheat aphid corresponding to ESTs observed in salivary-gland cDNA libraries from the pea aphid. Of 17 putative salivary gland orthologs, which were successfully amplified using primers based on pea aphid genes, 13 exhibited nucleotide sequence variation both within a biotype (RWA1 or RWA2) and between the two biotypes. There was a common pattern at all polymorphic loci: the predominant variant was observed in both biotypes, while a number of less frequently observed variants was found in each biotype, almost always restricted to one or the other biotype.

We believe the detected polymorphism represents allelic variation rather than genetic variation arising from transcription at duplicate loci, even though some of the 17 genes have paralogs in pea aphid genome. When subjected to phylogenetic analysis using maximum parsimony, all the transcript variants for each of the 9 polymorphic genes clustered as a single group, and the single cluster never grouped with more than one gene copy (paralog) from the pea aphid. If the transcripts were generated from more than one gene copy, we would expect the transcripts to cluster into separate groups, and in some cases we would expect these clusters to be associated with multiple gene copies in the pea aphid genome as well. The pattern of transcript diversity that we observed is also consistent with allelic diversity. Where polymorphism was observed, there was always a predominant transcript shared by the

1 two biotypes and several less abundant transcript variants usually were unique to one
2 biotype. If the transcript variation was generated from different gene copies
3 transcribed at different levels, we would certainly expect some of the lower expressed
4 transcripts to be shared between the biotypes - especially considering that biotype 2 is
5 almost certainly derived from biotype 1.

6 Comparison of variants for a given cloned transcript revealed, in many cases,
7 more non-synonymous mutations than synonymous mutations. This may indicate that
8 the corresponding genes have been under positive selection pressure, which has been
9 observed for predicted salivary proteins in the pea aphid (Carolan *et al.*, 2011). The
10 bias towards non-synonymous mutations was observed for transcripts of proteins and
11 enzymes predicted to be present in the saliva (*e.g.* protein C002, dipeptidyl
12 carboxypeptidase), as well as transcripts that encode proteins (*e.g.* sucrase,
13 lipid-anchored protein) that are expressed in the salivary gland but are probably not
14 secreted in the saliva. Eukaryotic pathogen effector evolution is often driven by
15 diversifying (positive) selection together with the maintenance of multiple gene
16 copies or alleles (Ma & Guttmann, 2008), which would explain strong positive
17 selection in genes encoding the protein components of secreted saliva (Carolan *et al.*,
18 2011) but not salivary gland proteins that are not secreted into plants.

19 Perhaps the abundance of non-synonymous mutations results from novel
20 transcripts being generated quickly then disappearing just as quickly from the
21 population. Interestingly, the overall ratio of non-synonymous to synonymous
22 mutations across all 17 transcripts is approximately 2:1, which is precisely the ratio at

which these mutations should be generated (non-synonymous mutations in codon positions 1 and 2, synonymous mutations at codon position 3). The rapid appearance and disappearance of low-abundance transcripts would also explain why no low-abundance alleles are shared between the two biotypes. Another possible explanation of the apparent high abundance of non-synonymous mutation is that asexual aphids may lose a lot of genetic diversities in colonies and allele frequencies could be greatly skewed because of drift considering the two strains have been raised in the lab since 2002 and 2003 respectively.

Previous research has struggled to identify genetic variation between these biotypes that could be used as a molecular diagnostic (Shufran *et al*., 2007; Shufran & Payton, 2009). In this study, we identified in salivary gland transcripts a high level of sequence variations among less abundant alleles that could potentially be used to distinguish the biotypes. However, these variants would only be effective as a diagnostic if they were maintained over several generations. On the other hand, if these less abundant alleles arise and disappear quickly in each biotype, their effectiveness as markers would be lost. Future research will focus on how the frequency of these less abundant alleles changes spatially and temporally within each biotype.

Acknowledgments

This work was supported by the CRC for National Plant Biosecurity (Australia), the Grains Research and Development Corporation (Australia), and the Kansas Agricultural Experiment Station (publication number XX-YYYY). We thank Dr.

1 Neal Dittmer for discussions and Matthew Heerman for the same and also for help in
2 depositing sequences.

3 References

4 Basky, Z. (2003) Biotypic and pest status differences between Hungarian and South
5 African populations of Russian wheat aphid, *Diuraphis noxia* (Kurdjumov)
6 (Homoptera: Aphididae). *Pest Management Science*, 59, 1152-1158.

7 Bos, J.I.B., Prince, D., Pitino, M., Maffei, M.E., Win, J. and Hogenhout, S.A. (2010)
8 A functional genomics approach identifies candidate effectors from the aphid
9 species *Myzus persicae* (green peach aphid). *PLoS Genet*, 6, e1001216.

10 Boyko, E.V., Smith, C.M., Thara, V.K., Bruno, J.M., Deng, Y., Starkey, S.R. and
11 Klaahsen, D.L. (2006) Molecular basis of plant gene expression during aphid
12 invasion: wheat Pto- and Pti-like sequences are involved in interactions between
13 wheat and Russian wheat aphid (Homoptera: Aphididae). *Journal of Economic*
14 *Entomology*, 99, 1430-1445.

15 Carolan, J.C., Fitzroy, C.I.J., Ashton, P.D., Douglas, A.E. and Wilkinson, T.L. (2009)
16 The secreted salivary proteome of the pea aphid *Acyrthosiphon pisum* characterised
17 by mass spectrometry. *Proteomics*, 9, 2457-2467.

18 Carolan, J.C., Caragea, D., Reardon, K.T., Mutti, N.S., Dittmer, N., Pappan, K., Cui,
19 F., Castaneto, M., Poulain, J., Dossat, C., Tagu, D., Reese, J.C., Reeck, G.R.,
20 Wilkinson, T.L. and Edwards, O.R. (2011) Predicted effector molecules in the
21 salivary secretome of the pea aphid (*Acyrthosiphon pisum*)-A dual
22 transcriptomic-proteomic approach. *Journal of Proteome Research*, 10, 1505-1518.

1 Cooper, W.R., Dillwith, J.W. and Puterka, G.J. (2010) Salivary proteins of Russian
2 wheat aphids (Hemiptera: Aphididae). *Environmental Entomology*, 39, 223-231.
3 Dolatti, L., Ghareyazie, B., Moharramipour, S. and Noori-Daloii, M.R. (2005)
4 Evidence for regional diversity and host adaptation in Iranian populations of the
5 Russian wheat aphid. *Entomologia Experimentalis et Applicata*, 114, 171-180.
6 Haley, S.D., Peairs, F.B., Walker, C.B., Rudolph, J.B. and Randolph, T.L. (2004)
7 Occurrence of a new Russian wheat aphid biotype. *Crop Science*, 44, 1589-1592.
8 Harmel, N., Létocart, E., Cherqui, A., Giordanengo, P., Mazzucchelli, G.,
9 Guillonneau, F., De Pauw, E., Haubruge, E. and Francis, F. (2008) Identification of
10 aphid salivary proteins: a proteomic investigation of *Myzus persicae*. *Insect*
11 *Molecular Biology*, 17, 165-174.
12 Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R. and Kamoun, S. (2009)
13 Emerging concepts in effector biology of plant-associated organisms. *Molecular*
14 *Plant-Microbe Interactions*, 22, 115-122.
15 International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid
16 *Acyrthosiphon pisum*. *PLoS Biology*, 8, e1000313.
17 Lapitan, N.L.V., Li, Y., Peng, J. and Botha, A. (2007) Fractionated extracts of
18 Russian wheat aphid eliciting defense responses in wheat. *Journal of Economic*
19 *Entomology*, 100, 990-999.
20 Liang, C.X. (2006) Laccase-1 in the pea aphid, *Acyrthosiphon pisum* (Harris). MSc
21 thesis, Kansas State University, Manhattan, Kansas, USA.
22 Liu, X., Marshall, J.L., Stary, P., Edwards, O., Puterka, G., Dolatti, L., El Bouhssini,

1 M., Malinga, J., Lage, J. and Smith, C.M. (2010) Global phylogenetics of an
2 invasive aphid species: Evidence for multiple invasions into North America.
3 *Journal of Economic Entomology*, 103, 958-965.
4 Ma, W. and Guttman, D.S. (2008) Evolution of prokaryotic and eukaryotic virulence
5 effectors. *Current Opinion in Plant Biology*, 11, 412-419.
6 Malinga, J.N., Kinuya, M.G., Kamau, A.W., Wanjama, J.K., Awalla, J.O. and Pathak,
7 R.S. (2007) Biotypic and genetic variation within tropical populations of Russian
8 wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae) in Kenya.
9 *Journal of Entomology*, 4, 350-361.
10 Miles, P.W. (1999) Aphid saliva. *Biological Reviews*, 74, 41-85.
11 Mutti, N.S., Park, Y., Reese, J.C. and Reeck, G.R. (2006) RNAi knockdown of a
12 salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*.
13 *Journal of Insect Science*, 6, 38.
14 Mutti, N.S., Louis, J., Pappan, L.K., Pappan, K., Begum, K., Chen, M., Park, Y.,
15 Dittmer, N., Marshall, J., Reese, J.C. and Reeck, G. (2008) A protein from the
16 salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a
17 host plant. *Proceedings of the National Academy of Sciences of the United States of*
18 *America*, 105, 9965-9969.
19 Peairs, F.B. (1987) Geographic and seasonal distribution including trapping and
20 observation. *Proceeding of the first Russian wheat aphid conference*, pp. 7-14.
21 September 23, Guymon, Oklahoma, USA.
22 Puterka, G.J., Black IV, W.C., Steiner, W.M. and Burton, R.L. (1993) Genetic

variation and phylogenetic relationships among worldwide collections of the Russian wheat aphid, *Diuraphis noxia* (Mordvilko), inferred from allozyme and RAPD-PCR markers. *Heredity*, 70, 604-618.

Puterka, G.J., Burd, J.D., Porter, D., Shufran, K., Baker, C., Bowling, B. and Patrick, C. (2007) Distribution and diversity of Russian wheat aphid (Hemiptera: Aphididae) biotypes in North America. *Journal of Economic Entomology*, 100, 1679-1684.

Qureshi, J.A., Jyoti, J.L. and Michaud, J.P. (2005) Differential colonization of wheat cultivars by two biotypes of Russian wheat aphid (Homoptera: Aphididae). *Insect Science*, 12, 341-349.

Shufran, K.A., Kirkman, L.R. and Puterka, G.J. (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, K.A. and Payton, T.L. (2009) Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal of Economic Entomology*, 102, 440-445.

Smith, C.M., Belay, T., Stauffer, C., Stary, P., Kubeckova, I. and Starkey, S. (2004) Identification of Russian wheat aphid (Homoptera: Aphididae) biotypes virulent to the *Dn4* resistance gene. *Journal of Economic Entomology*, 97, 1112-1117.

Stoetzel, M.B. (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.

1 Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular
2 evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology*
3 *and Evolution*, 24, 1596-1599.
4 Tjallingii, W.F. (2006) Salivary secretions by aphids interacting with proteins of
5 phloem wound responses. *Journal of Experimental Botany*, 57, 739-745.
6 Tolmay, V.L., Lindeque, R.C. and Prinsloo, G.L. (2007) Preliminary evidence of a
7 resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia*
8 (Kurdjumov) (Homoptera: Aphididae), in South Africa. *African Entomology*, 15,
9 228-230.
10 Van Zyl, R.A. (2007) Elucidation of possible virulence factors present in Russian
11 wheat aphid (*Diuraphis noxia*) biotypes' saliva. MSc thesis, University of Pretoria,
12 South Africa.
13 Weng, Y., Azhaguvel, P., Michels Jr, G.J. and Rudd, J.C. (2007) Cross-species
14 transferability of microsatellite markers from six aphid (Hemiptera: Aphididae)
15 species and their use for evaluating biotypic diversity in two cereal aphids. *Insect*
16 *Molecular Biology*, 16, 613-622.
17 Zhang, R., Liang, H. and Zhang, G. (1999) Research advances of Russian wheat aphid
18 (Homoptera: Aphididae): distribution, economical impact and biological characters.
19 *Acta Entomologica Sinica*, 42 Suppl, 130-140.
20

21 Appendices

22 **Figure S1.** Nucleotide and amino acid sequences of the transcripts without

1 polymorphism in the two biotypes of RWA. A. coated vesicle membrane protein. B.
2 AphidB1_C07_t7_050. C. ID0AAH13AH01ZM2. D. peroxidase.
3 **Figure S2.** Polymorphism of dipeptidyl carboxypeptidase (shorted as DC) in the two
4 biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3.
5 RWA2 has variants 1, 4-6. B. Alignment of nucleotide sequences of various clones
6 from each biotype. I: RWA1. II: RWA2.
7 **Figure S3.** Polymorphism of cathepsin B in the two biotypes of RWA. A. Alignment
8 of amino acid variants. RWA1 has variants 1-4. RWA2 has variants 1, 5-8. B.
9 Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II:
10 RWA2.
11 **Figure S4.** Polymorphism of cathepsin L in the two biotypes of RWA. A. Alignment
12 of amino acid variants. RWA1 has variant 1. RWA2 has variants 1 and 2. B.
13 Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II:
14 RWA2.
15 **Figure S5.** Polymorphism of endoprotease FURIN (shorted as Furin) in the two
16 biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3.
17 RWA2 has variants 1 and 4. B. Alignment of nucleotide sequences of various clones
18 from each biotype. I: RWA1. II: RWA2.
19 **Figure S6.** Polymorphism of glucose dehydrogenase (shorted as GD) in the two
20 biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-4.
21 RWA2 has variants 1, 2 and 5. B. Alignment of nucleotide sequences of various clones
22 from each biotype. I: RWA1. II: RWA2.

1 **Figure S7.** Polymorphism of JHBP in the two biotypes of RWA. A. Alignment of
2 amino acid variants. RWA1 has variants 1 and 2. RWA2 has variants 1 and 3. B.
3 Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II:
4 RWA2.
5 **Figure S8.** Polymorphism of peptidase M1 in the two biotypes of RWA. A. Alignment
6 of amino acid variants. RWA1 has variant 1. RWA2 has variants 1-3. B. Alignment of
7 nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.
8 **Figure S9.** Polymorphism of probable ER retained protein (named as C037) in the
9 two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3.
10 RWA2 has variants 1 and 4. B. Alignment of nucleotide sequences of various clones
11 from each biotype. I: RWA1. II: RWA2.
12 **Figure S10.** Polymorphism of Emp24 in the two biotypes of RWA. A. Alignment of
13 amino acid variants. RWA1 has variants 1-3. RWA2 has variant 1. B. Alignment of
14 nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.
15 **Figure S11.** Polymorphism of laccase1 in the two biotypes of RWA. A. Alignment of
16 amino acid variants. RWA1 has variants 1-3. RWA2 has variants 1 and 4. B.
17 Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II:
18 RWA2.
19 **Figure S12.** Maximum parsimony analysis on dipeptidyl carboxypeptidase (shorted
20 as DC) (A), cathepsin B (B), endoprotease FURIN (shorted as Furin) (C) and glucose
21 dehydrogenase (shorted as GD) (D) amino acid sequences of RWA and pea aphid. The
22 GenBank IDs of pea aphid homologs in NCBI are in bold. 1000 replicates were

performed to create the consensus trees. Only the bootstrap percentage values > 50% are shown at nodes.

Figure S13. Maximum parsimony analysis on JHBP (A), peptidase M1 (B) and laccase1 (C) amino acid sequences of RWA and pea aphid. The GenBank IDs of pea aphid homologs in NCBI are in bold. 1000 replicates were performed to create the consensus trees. Only the bootstrap percentage values > 50% are shown at nodes.

Figure Legends

Figure 1. Polymorphism of C002 in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-5. RWA2 has variants 1, 6-9. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

Figure 2. Polymorphism of sucrase in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-9. RWA2 has variants 1, 10-15. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

Figure 3. Polymorphism of trehalase in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-4. RWA2 has variants 1, 5-8. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

Figure 4. Maximum parsimony analysis on amino acid sequences of sucrase (A) and trehalase (B) from RWA and pea aphid. The GenBank IDs of pea aphid homologs in NCBI are in bold. 1000 replicates were performed to create the consensus trees. Only the bootstrap percentage values > 50% are shown at nodes.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17 **Table 1** Information of primers for cloning the 17 transcripts and PCR product lengths

Transcripts	Primers used for RT-PCR	Length of PCR product (bp)
Coated vesicle membrane protein	Forward: ACGCTCACGCTGAAGAATGT	465
	Reverse: CAAAGAATGACCACATCACAA	
AphidB1_C07_t7_050	Forward: CCGATAAGCTCTCGACTGGA	204
	Reverse: TTATTCGCCACGGTATGTGA	
ID0AAH13AH01ZM2	Forward: CAAAGACTATCCCGCTTCAAA	159
	Reverse: GACCGCTCAATGGCAGTATT	
Peroxidase	Forward: CATTGATTGGTAACGTTGATGG	384
	Reverse: CAGCAATAACACAACTTCCAGT	
Sucrase	Forward: CGCCTCCGAGTAATTGGTTA	874, 806
	Reverse: AGAGGAAGCCACAACGAAGA	

Dipeptidyl carboxypeptidase	Forward: AAGTGGCCGAAGAGTTTTTCA	965
	Reverse: AGGTTGACTTGTTTCACCCTT	
Cathepsin B	Forward: CGAACAAGCGTATTTCTTGGA	719
	Reverse: TCTACACCCCAACCGATCAA	
Cathepsin L	Forward: TGTGGATCATGCTGGTCATT	433
	Reverse: CACCATGGTCCAGTTCTGTG	
Endoprotease FURIN	Forward: CTACGTTGAGCACGTGCAAT	273
	Reverse: AATATCAATCGCCACTTTTTCA	
Glucose dehydrogenase	Forward: CCTCAGGTATTGGGCCTAAA	850
	Reverse: GCCCTTTCAGCGACCATGAT	
JHBP	Forward: GGTGAATACTGGGGTGAATA	228
	Reverse: CATTTCCTTGAGCAGTTCTTG	
Trehalase	Forward: GGCTGTAGCTAGTTCAGTGTTG	455, 551, 453
	Reverse: GTTTCCGGTGCATAGGCGAATAG	
Peptidase M1	Forward: GCTATCGACTGGACGTGACA	410
	Reverse: CTCGTCAAAACAGGGGAAAG	
Probable ER retained protein	Forward: CCGGAACCGATTTATCTGAA	401
	Reverse: GACACGCCAATGAGTTGAAA	
Emp24	Forward: GGGATGGCTGGTTTTGCCGT	365
	Reverse: CTGTTCACATATGTTTTGTTATC	
C002	Forward: GATAGCGATAATTTACAACAT	765
	Reverse: TATATCACTAGTCTGTATGGAC	
Laccase 1	Forward: GTNGARGARATHGARCARATGGA	367
	Reverse: ACRAANGGCCACCAYTTNCC	

1

2

3

4

5

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17 **Table 2** Percent identities of the 17 transcripts between RWA and pea aphid

Transcripts	Identity		GenBank ID	
	Nucleotide	Predicted protein	Pea aphid	RWA
Coated vesicle membrane protein	95%	100%	HS092218	HQ709426
AphidB1_C07_t7_050	90%	98%	HS092964	HQ709427
ID0AAH13AH01ZM2	97%	94%	DV748473	HQ709428
Peroxidase	89%	86%	DV751002	HQ709429
Sucrase	92%	93%	JG732090	HQ709431
Dipeptidyl carboxypeptidase	93%	95%	HS094983	HQ709433
Cathepsin B	80%	70%	JG732091	HQ709434
Cathepsin L	91%	96%	HS094667	HQ709435
Endoprotease FURIN	94%	97%	DV750164	HQ709436
Glucose dehydrogenase	90%	90%	HS092125	HQ709437

JHBP	92%	93%	HS095047	HQ709438
Trehalase	90%	89%	JG732092	HQ709432
Peptidase M1	86%	87%	HS095577	HQ709439
Probable ER retained protein	93%	97%	HS092530	HQ709440
Emp24	95%	99%	HS092977	HQ709441
C002	74%	51%	HS092532	HQ709430
Laccase 1	93%	93%	CN757762	HQ709442

1
2
3
4
5
6
7
8
9
10

Table 3 Polymorphism of the 17 transcripts in RWA biotypes 1 and 2

Transcripts	Number of amino acid residues	Number of nucleotide variants		Number of amino acid variants		Number of shared amino acid variant	Ratio* of shared amino acid variant	
		RWA1	RWA2	RWA1	RWA2		RWA1	RWA2
Coated vesicle membrane protein	154	1	1	1	1	1	5/5	5/5
AphidB1_C07_t7_050	67	1	1	1	1	1	5/5	5/5
ID0AAH13AH01ZM2	52	1	1	1	1	1	4/4	5/5
Peroxidase	127	1	1	1	1	1	5/5	5/5
Sucrase	290	10	8	9	7	1	2/10	4/10
Dipeptidyl carboxypeptidase	321	4	6	3	4	1	4/6	2/6
Cathepsin B	239	4	7	4	5	1	1/4	8/12
Cathepsin L	144	1	2	1	2	1	5/5	4/5
Endoprotease FURIN	90	3	4	3	2	1	3/5	4/5

Glucose dehydrogenase	282	4	4	4	3	2	2/5, 1/5	1/5, 3/5
JHBP	76	2	2	2	2	1	4/5	4/5
Trehalase	141	7	5	4	5	1	7/10	4/8
Peptidase M1	136	3	3	1	3	1	5/5	3/5
Probable ER retained protein	133	4	2	3	2	1	3/5	4/5
Emp24	121	4	2	3	1	1	3/5	5/5
C002	238	6	8	5	5	1	2/8	6/10
Laccase1	122	3	2	3	2	1	3/5	4/5

1 * The ratio of shared amino acid variant is defined as the clone number of shared
2 variant divided by the total number of sequenced clones.
3
4
5
6
7
8
9
10 **Table 4** Synonymous and nonsynonymous mutations in the 13 transcripts

Transcripts	Total mutations	Nonsynonymous (N)	Synonymous (S)	N/S
Sucrase	23	16	7	2.3
Dipeptidyl carboxypeptidase	10	6	4	1.5
Cathepsin B	37	25	12	2.1
Cathepsin L	1	1	0	∞
Endoprotease FURIN	5	3	2	1.5
Glucose dehydrogenase	8	5	3	1.7
JHBP	3	2	1	2.0
Trehalase	10	7	3	2.3
Peptidase M1	5	2	3	0.7

Probable ER retained protein	4	3	1	3.0
Emp24	4	2	2	1.0
C002	24	14	10	1.4
Laccase1	4	4	0	∞
Total	138	90	48	1.9