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:


Published Version Information


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


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 Cui¹,²,⁵*, C. Michael Smith³, John Reese³, Owain Edwards⁴,⁵, Gerald Reeck¹

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
Introduction

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

RWA biotypes have been designated based on the damage resulting from aphid feeding on wheat cultivars containing resistance genes *Dn1* to *Dn9*. Using this system, 5 biotypes, RWA1-RWA5, have been identified in the United States (Puterka *et al*., 2007). In studies of samples collected in 2005 from 98 fields of wheat or barley in Oklahoma, Texas, New Mexico, Colorado, Kansas, Nebraska and Wyoming, only
RWA1 and RWA2 were found, and RWA2 was the predominant biotype (Puterka et al., 2007). RWA1 is virulent only to wheat carrying resistance genes Dn1, Dn8 and Dn9. RWA2 is virulent to wheat containing any of the Dn genes other than Dn7 (Haley et al., 2004; Puterka et al., 2007; Qureshi et al., 2005). Liu et al. (2010) demonstrated that RWA1 and RWA2 are contained in the Middle East-African clade, and RWA3, RWA4, and RWA5 are part of the European clade. RWA biotypes also occur in Africa, Asia, Europe, and South America (Basky, 2003; Dolatti et al., 2005; Malinga et al., 2007; Smith et al., 2004; Tolmay et al., 2007).

Saliva, injected into plant tissue, is the point of contact between aphid and plant. Proteins and enzymes in saliva are believed to play several roles in allowing continued feeding by an aphid on the phloem (Miles, 1999; Tjallingii, 2006). In broad terms, some, and possibly all, of the proteins of aphid saliva can be thought of as “effectors,” a term introduced to designate proteins secreted by plant pathogens for the purpose of establishing “colonization” of the plant by the pathogen (Hogenhout et al., 2009). Indeed this suggestion has been made specifically in the case of the RWA (Boyko et al., 2006; Lapitan et al., 2007; Cooper et al., 2010), and RWA secretes protein effectors that differ among biotypes (Van Zyl, 2007). However, for the most part, we know little about the functions of individual components of aphid saliva. Recent descriptive and functional studies on aphid salivary gland genes and proteins support effector functions for many aphid salivary proteins (Carolan et al., 2009, 2011; Bos et al., 2010). Results from the pea aphid also suggest that genes transcribed in salivary glands are evolving faster than their orthologs in other insects (Carolan et al.,
Several studies have looked for genetic differences between RWA1 and RWA2. Shufran et al. (2007) found little or no difference in the sequence of a 332-base-pair region of the mitochondrial gene encoding cytochrome oxidase I in RWA samples collected from 1986 through 2006, a time period in which RWA2 arose. In a subsequent study, Shufran and Payton (2009) reported little or no variation in the same gene-sequence or in several simple-sequence repeat loci between RWA1 and RWA2. On the other hand, Weng et al. (2007) reported that 19 of 57 simple-sequence repeat markers examined differed between RWA1 and RWA2.

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

**Materials and methods**

*Insects*

Populations of *D. noxia* biotype 1 (RWA1) collected from wheat fields near Hays, KS, in 2002, and biotype 2 (RWA2) individuals collected from wheat fields near
Briggsdale, CO (via the USDA-ARS Plant Science Research Laboratory at Stillwater, OK), in 2003 were cultured continuously in separate locations in the greenhouse on susceptible ‘Jagger’ wheat plants at Kansas State University before use in the experiments. The identity of each biotype was verified in diagnostic plant differential greenhouse assays at Stillwater, OK, and Manhattan, KS.

mRNA extraction and cDNA synthesis

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

RT-PCR and cloning of regions of transcripts

We selected 17 transcripts for this work, based on sequences obtained from salivary-gland cDNA libraries of the pea aphid, *Acyrthosiphon pisum* (Carolan et al., 2011) and predicted to possibly encode secreted proteins (that is, proteins of saliva) based on the existence of hydrophobic regions (possible secretion signals) at the N-termini of the encoded polypeptides. In the case of the laccase-1 transcript, an EST has not been reported in a salivary-gland EST library, but studies in one of our laboratories indicates that the protein occurs in salivary glands (Liang, 2006). Regions
within the RWA orthologs of these transcripts were amplified from RWA1 and RWA2 cDNAs and sequenced. Primers for PCR-amplification were based on sequences of the pea aphid ESTs. Sequences of primer-pairs and predicted PCR product sizes for the 17 transcripts are detailed in Table 1. PCRs (25 μL) were comprised of 0.5 μM of each primer, 1 μL of template cDNA and 12.5 μL of PCR master mix that consisted of 50 units/ml Taq DNA polymerase, 400 μM dNTP and 3 mM MgCl₂ (Promega, Madison, WI, USA). Initial denaturation of the template cDNA was at 94 °C for 2 min, and was followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s to 1 min depending on the size of products, and a final step for 5 min at 72 °C. The PCR products were separated in 1% agarose gel and then purified from the gel with QIAquick gel extraction kit (Qiagen, Valencia, CA, USA). The purified products were T-A cloned into pCR2.1-TOPO vector and transfected into TOP10 or TOP10F’ cells. At least 4 plasmids for each transcript of each biotype were extracted with QIAprep spin miniprep kit (Qiagen, Valencia, CA, USA) and sequenced in the DNA Sequencing and Genotyping Facility at Kansas State University, using an Applied Biosystems 3730 DNA Analyzer. We obtained 750-900 bases of sequence information on each run. The sequencing was found to be error-free for 600 bases. The clones longer than 600 bases were always sequenced in two directions. Sequences for each transcript were aligned with the online ClustalW2 sever at EBI (www.ebi.ac.uk/Tools/clustalw2/index.html) and formatted with BioEdit software.

Phylogenetic analysis of RWA and pea aphid salivary genes

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
peroxidase and the two non-annotatable proteins) we detected no polymorphism within or between biotypes (Table 3). For the other 13 transcripts, we observed polymorphism both within and between the biotypes, at the nucleotide level and the predicted amino sequences (Table 3). Most of the variations were detected only once within individual transcripts (Figure 1 - Figure 3 for representative examples; also see Figure S1 - Figure S11) but some variants were found more than once in one or both of biotypes, as in the case for cathepsin B (Figure S3). Usually the two biotypes only shared one variant and this was usually the predominant variant in both of the two biotypes.

To illustrate the nature of the observed polymorphisms, we presented our findings on three transcripts here. All other sequences were shown in Supplemental Material. The C002 transcript (Figure 1) encoded a protein required in the pea aphid for feeding on a host plant (Mutti et al., 2006, 2008). For this transcript, 9 variants at the amino acid level in the two biotypes were found, with 4 variants found only in RWA1 and another 4 found only in RWA2. The sucrase transcript showed the highest polymorphism among the 17 transcripts we examined (Figure 2). Nine amino acid variants in RWA1 and 7 amino acid variants in RWA2 were found. A 68-nucleotide deletion resulted in shorter ORF in RWA2 variant. The most complicated polymorphism was observed in the trehalase transcript (Figure 3). There were 8 variants at the amino acid level in the two biotypes taken together. In addition to single amino acid substitutions at several positions, 4 different polypeptide lengths were encoded. In RWA1 there were three polypeptide lengths. These lengths (in the
region we cloned) included: 141 residues; 151 residues (a mutation from TAA to CAA elongated the polypeptide) and 161 residues (due to a 96-nucleotide insertion). In RWA2, two lengths of this polypeptide region were observed, one of 141 residues and the other of 67 residues (a 2-nucleotide deletion shortened and shifted the open reading frame).

Non-synonymous and synonymous mutations

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

Phylogenetic analysis

A phylogenetic analysis was conducted as a means of discriminating between allelic and paralog variation in the transcripts identified for each prospective salivary gland locus. All the transcripts were compared to the gene sequences of all paralogs in the pea aphid genome (International Aphid Genomics Consortium, 2010) except the four transcripts without polymorphism. Only one paralog was found in the pea aphid genome for transcripts C002, cathepsin L, probable ER retained protein and Emp24. For the other 9 transcripts, all the variants detected clustered as a single group related 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
two biotypes and several less abundant transcript variants usually were unique to one
biotype. If the transcript variation was generated from different gene copies
transcribed at different levels, we would certainly expect some of the lower expressed
transcripts to be shared between the biotypes - especially considering that biotype 2 is
almost certainly derived from biotype 1.

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

Perhaps the abundance of non-synonymous mutations results from novel
transcripts being generated quickly then disappearing just as quickly from the
population. Interestingly, the overall ratio of non-synonymous to synonymous
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.
Neal Dittmer for discussions and Matthew Heerman for the same and also for help in depositing sequences.

References


Liu, X., Marshall, J.L., Stary, P., Edwards, O., Puterka, G., Dolatti, L., El Bouhssini,


variation and phylogenetic relationships among worldwide collections of the
Russian wheat aphid, *Diuraphis noxia* (Mordvilko), inferred from allozyme and

Puterka, G.J., Burd, J.D., Porter, D., Shufran, K., Baker, C., Bowling, B. and Patrick,

cultivars by two biotypes of Russian wheat aphid (Homoptera: Aphididae). *Insect
Science, 12*, 341-349.

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.

Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States. *Journal
of Economic Entomology, 102*, 440-445.

Identification of Russian wheat aphid (Homoptera: Aphididae) biotypes virulent to

Stoetzel, M.B. (1987) Information on and identification of *Diuraphis noxia*
(Homoptera: Aphididae) and other aphid species colonizing leaves of wheat and


**Appendices**

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

**Figure S2.** Polymorphism of dipeptidyl carboxypeptidase (shorted as DC) in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3. RWA2 has variants 1, 4-6. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S3.** Polymorphism of cathepsin B 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 S4.** Polymorphism of cathepsin L in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variant 1. RWA2 has variants 1 and 2. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S5.** Polymorphism of endoprotease FURIN (shorted as Furin) in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3. RWA2 has variants 1 and 4. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S6.** Polymorphism of glucose dehydrogenase (shorted as GD) in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-4. RWA2 has variants 1, 2 and 5. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.
**Figure S7.** Polymorphism of JHBp in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1 and 2. RWA2 has variants 1 and 3. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S8.** Polymorphism of peptidase M1 in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variant 1. RWA2 has variants 1-3. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S9.** Polymorphism of probable ER retained protein (named as C037) in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3. RWA2 has variants 1 and 4. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S10.** Polymorphism of Emp24 in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3. RWA2 has variant 1. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S11.** Polymorphism of laccase1 in the two biotypes of RWA. A. Alignment of amino acid variants. RWA1 has variants 1-3. RWA2 has variants 1 and 4. B. Alignment of nucleotide sequences of various clones from each biotype. I: RWA1. II: RWA2.

**Figure S12.** Maximum parsimony analysis on dipeptidyl carboxypeptidase (shorted as DC) (A), cathepsin B (B), endoprotease FURIN (shorted as Furin) (C) and glucose dehydrogenase (shorted as GD) (D) 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 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.
Table 1 Information of primers for cloning the 17 transcripts and PCR product lengths

<table>
<thead>
<tr>
<th>Transcripts</th>
<th>Primers used for RT-PCR</th>
<th>Length of PCR product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coated vesicle membrane protein</td>
<td>Forward: ACGCTCACGCTGAAGAATGT</td>
<td>465</td>
</tr>
<tr>
<td></td>
<td>Reverse: CAAAGAATGACCACCATCACA A</td>
<td></td>
</tr>
<tr>
<td>AphidB1_C07_t7_050</td>
<td>Forward: CCGATAAGCTCTCGACTGGA</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>Reverse: TTATTCGCCACGGTATGTGA</td>
<td></td>
</tr>
<tr>
<td>ID0AAH13AH01ZM2</td>
<td>Forward: CAAAGACTATCCCGCTTCAAA</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>Reverse: GACCGCTCAATGGCAGTATT</td>
<td></td>
</tr>
<tr>
<td>Peroxidase</td>
<td>Forward: CATTGATTTGTAACGTTGATGG</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>Reverse: CAGCAATAACACAACCTCCAGT</td>
<td></td>
</tr>
<tr>
<td>Sucrase</td>
<td>Forward: CGCCTCGAGTAATTGGTTA</td>
<td>874, 806</td>
</tr>
<tr>
<td></td>
<td>Reverse: AGAGGAAGCCACAAACGAAGA</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Forward Sequence</td>
<td>Reverse Sequence</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Dipeptidyl carboxypeptidase</td>
<td>AAGTGGCCGAAGAGTTTTTCA</td>
<td>AGGTTGACTTGTTCACCCTT</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>CGAACAAAGCGTATTTCTTGGGA</td>
<td>TCTACACCCCAAACGATCAA</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>TGTGGATCATGCTGGTCAATT</td>
<td>CACCATGGTCAGTTCTTGTG</td>
</tr>
<tr>
<td>Endoprotease FURIN</td>
<td>CTACGGTGACACGGTCAAT</td>
<td>AATATCAATCGCCACTTTTTCA</td>
</tr>
<tr>
<td>Glucose dehydrogenase</td>
<td>CCTCAGGTATTGGGCCTAAA</td>
<td>GCCCTTTTCAGGACCAGCATGAT</td>
</tr>
<tr>
<td>JHBP</td>
<td>GGTGAATCTGGGGGTGAATA</td>
<td>CATTTCCTTGAGCAGTTCTTGT</td>
</tr>
<tr>
<td>Trehalase</td>
<td>GGCTGTAGCTAGTTCAGTGTG</td>
<td>GTTTCCGGTGATAGGCGAATAG</td>
</tr>
<tr>
<td>Peptidase M1</td>
<td>GCTATCGACTGGACGTGACA</td>
<td>CTCGTCAAAACACGGGAAAG</td>
</tr>
<tr>
<td>Probable ER retained protein</td>
<td>CCGGAACCAGTTTATCTGAA</td>
<td>GACACGCCAATGAGTTTAA</td>
</tr>
<tr>
<td>Emp24</td>
<td>GGGATGGCTGGTTTGGCCGT</td>
<td>CTGTTACATATGTITTTGTTATC</td>
</tr>
<tr>
<td>C002</td>
<td>GATACCGATAATTTACACAT</td>
<td>TATATCAGTGCTGTATGAC</td>
</tr>
<tr>
<td>Laccase 1</td>
<td>GTNGARGARATHGARCARATGGA</td>
<td>ACRAANGGCCACCAYTTNCC</td>
</tr>
</tbody>
</table>
Table 2  Percent identities of the 17 transcripts between RWA and pea aphid

<table>
<thead>
<tr>
<th>Transcripts</th>
<th>Identity</th>
<th>GenBank ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nucleotide</td>
<td>Predicted protein</td>
</tr>
<tr>
<td>Coated vesicle membrane protein</td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td>AphidB1_C07_t7_050</td>
<td>90%</td>
<td>98%</td>
</tr>
<tr>
<td>ID0AAH13AH01ZM2</td>
<td>97%</td>
<td>94%</td>
</tr>
<tr>
<td>Peroxidase</td>
<td>89%</td>
<td>86%</td>
</tr>
<tr>
<td>Sucrase</td>
<td>92%</td>
<td>93%</td>
</tr>
<tr>
<td>Dipeptidyl carboxypeptidase</td>
<td>93%</td>
<td>95%</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>80%</td>
<td>70%</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>91%</td>
<td>96%</td>
</tr>
<tr>
<td>Endoprotease FURIN</td>
<td>94%</td>
<td>97%</td>
</tr>
<tr>
<td>Glucose dehydrogenase</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Transcripts</td>
<td>Number of amino acid residues</td>
<td>Number of nucleotide variants</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Coated vesicle membrane protein</td>
<td>154</td>
<td>1</td>
</tr>
<tr>
<td>AphidB1_C07_t7_050</td>
<td>67</td>
<td>1</td>
</tr>
<tr>
<td>ID0AAH13AH01ZM2</td>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>Peroxidase</td>
<td>127</td>
<td>1</td>
</tr>
<tr>
<td>Sucrase</td>
<td>290</td>
<td>10</td>
</tr>
<tr>
<td>Dipeptidyl carboxypeptidase</td>
<td>321</td>
<td>4</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>239</td>
<td>4</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>144</td>
<td>1</td>
</tr>
<tr>
<td>Endoprotease FURIN</td>
<td>90</td>
<td>3</td>
</tr>
</tbody>
</table>

*Table 3 Polymorphism of the 17 transcripts in RWA biotypes 1 and 2*
<table>
<thead>
<tr>
<th>Transcripts</th>
<th>Total mutations</th>
<th>Nonsynonymous (N)</th>
<th>Synonymous (S)</th>
<th>N/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucrase</td>
<td>23</td>
<td>16</td>
<td>7</td>
<td>2.3</td>
</tr>
<tr>
<td>Dipeptidyl carboxypeptidase</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>37</td>
<td>25</td>
<td>12</td>
<td>2.1</td>
</tr>
<tr>
<td>Cathepsin L</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>∞</td>
</tr>
<tr>
<td>Endoprotease FURIN</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td>Glucose dehydrogenase</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>JHBP</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Trehalase</td>
<td>10</td>
<td>7</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>Peptidase M1</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

* The ratio of shared amino acid variant is defined as the clone number of shared variant divided by the total number of sequenced clones.

**Table 4** Synonymous and nonsynonymous mutations in the 13 transcripts
<table>
<thead>
<tr>
<th>Protein Type</th>
<th>4</th>
<th>3</th>
<th>1</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probable ER retained protein</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>Emp24</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>C002</td>
<td>24</td>
<td>14</td>
<td>10</td>
<td>1.4</td>
</tr>
<tr>
<td>Laccase1</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>∞</td>
</tr>
<tr>
<td>Total</td>
<td>138</td>
<td>90</td>
<td>48</td>
<td>1.9</td>
</tr>
</tbody>
</table>